Probing the Characteristics of Potential Energy Surfaces

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Bastian Schäfer
aus Deutschland

Basel, 2015

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Stefan Goedecker

Prof. Dr. Axel Groß

Basel, den 15. September 2015

Prof. Dr. Jörg Schibler
Dekan

Für Gabriele und Jürgen

Acknowledgements

First and foremost I want to thank my supervisor Prof. Dr. Stefan Goedecker for the opportunity to work and study in the field of computational physics. More importantly, I want to thank him for all the countless inspiring, encouraging and motivating discussions. I also thank Prof. Dr. Axel Groß for accepting to co-referee this thesis.
Furthermore, I want to thank Dr. Luigi Genovese for always helping out with problems concerning the BigDFT code and Dr. Alirza Ghasemi for his support with the AMBER code. I would like to thank Prof. Dr. Lai-Sheng Wang for the constructive discussion we had during the APS March meeting in Denver and Dr. Rhitankar Pal for the endless conversations we had on the interpretation of the photoelectron spectra. I also like to thank Prof. Dr. Anjali L. Kshirsagar, Deepashri Saraf and Prof. Dr. Bhalchandra Pujari for their hospitality and the interesting discussions during my stay in Pune. Furthermore, I thank Martin Jacquot from the SciCORE team for always promptly fixing problems with the computer clusters, even during weekends, and the secretaries of the Physics Department, Barbara Kammermann and Astrid Kalt for their kind support.
Last but not least I want to thank all the present and former fellow students and postdocs that are and were members of Stefan Goedecker's group during my PhD time for always creating a pleasant and enjoyable atmosphere for my work.

Abstract

The potential energy surface of multi-atomic systems encodes important aspects such as thermodynamic and dynamic properties or the equilibrium geometries. Collections of low-energy minima and the reaction pathways that connect the minima with each other can be key elements in the study of potential energy surfaces and their properties. The extension of the minima hopping (MH) global optimization method to the minima hopping guided path search method (MHGPS) forms the heart of this thesis. MHGPS is a MH based approach for finding complex reaction pathways that connect the local minima in an efficient, automatized and unbiased fashion. Also, in this context, novel stabilized quasi-Newton local optimizers for the computation of minima and saddle points are developed. These optimizers are designed for robustness to the noisy forces delivered by density functional codes. Using benchmarks, the MHGPS method as well as the stabilized quasi-Newton optimizers are found to compare favorably with existing algorithms. Using the MHGPS method, novel results are presented for previously extensively studied Lennard-Jones clusters. Besides that, an ab-initio structure prediction study using the MH global optimization method is presented for the neutral and anionic gold clusters with 26 atoms. Finally, computationally efficient methods for a qualitative characterization of potential energy surfaces are discussed. In this context, MHGPS is applied at the density functional level of theory to the potential energy surface of Si_{20}.

Contents

Introduction

Important aspects such as thermodynamic and dynamic properties or equilibrium conformations of multi-atomic systems like clusters, molecules or surfaces, are encoded in their potential energy surface (PES).[1–4] In the past years, computer aided global optimization or, more accurately, the automatized sampling of low-energy equilibrium geometries, has become a hot topic in the materials science community. To this end, powerful global optimization methods such as several genetic algorithms[5–9], basin hopping[10–12], the activation relaxation technique[13–17] and minima hopping (MH)[9,18–20] have been developed during the last three decades. However, when restricted to only the geometries and relative potential energies of the minima, the finite temperature and dynamic behavior of a system is not accessible and important questions of significant physical and chemical interest must remain unanswered: Is the identified potential energy global minimum actually observable in an experiment, or are there reasons to believe that other minima are entropically favored? Are there multiple kinetically stable states? What are the detailed atomistic mechanisms of a certain process? The key to answer these and similar questions lies in the detailed knowledge of the characteristics of a PES. Beyond the mere knowledge of the global minimum and additional local minima, the reaction pathways that connect these minima with each other are key elements needed to answer the aforementioned questions.

In principle, reactive processes are described by dynamical trajectories as can be obtained from molecular dynamics (MD) simulations. Even at ab-initio level of theory pure MD simulations can be used to study the reaction dynamics.[21–25] However, pure MD simulations are limited by the femto-second time scale of the fastest atomic motions. This is particullarily problematic for the purpose of observing reaction pathways that constitute rare events in MD simulations. Driven by this problem, new dynamical approaches like metadynamics, temperature accelerated dynamics or transition path sampling were developed in the past.[26–29] However, these methods can be challenging to use in practice. Despite their improvements over MD, the sampling of dynamic reaction pathways is still very demanding computationally and, for more complex PESs, can be beyond computational feasibility.[30] Furthermore, some of these improved dynamic methods rely on the definition of a reaction coordinate or on an order parameter. In particular, chosing proper reaction coordinates is non-trivial and the

outcome of free energy calculations and atomistic details of chemical reactions depends on their definition.[31,32]

Most notably the work of Wales and co-workers demonstrated that a detailed collection of energetically low-lying minima, transition states and the information which transition state is connected to which minima can form a basic element needed for answering the above posed questions.[1–4,32–40] In the past, several eigenvector following approaches, including an extension of the above mentioned activation relaxation technique, or discrete analogues of the transition path sampling method were exploited for the purpose of sampling these stationary points.[2,3,15–17,35,39,41] Even though many difficulties inherent to the before mentioned dynamical methods can be circumenvented within this coarse grained perception of a PES, the thorough and systematic sampling of all relevant minima, transition states and their connectivity remains a very challenging task. Even today, (semi-)empirical descriptions of the PES are frequently required for this purpose.[42–45] At more sophisticated levels of theory, like density functional theory (DFT), the study of reactive processes within this approach is often restricted to the computation of a very limited number of transitions, starting at one or a pair of carefully hand-selected initial atomic conformations.[46–48] For these purposes, variants of the nudged elastic band method or the (improved) dimer method are frequently used.[49–60] However, this non-automatized and highly selective approach based on human intuition and previously gathered experience can be suboptimal. Unforeseen phenomena might be missed, and the computational probing of reactive processes can be bounded by human-time limitations instead of available computer time. With the aim to alleviate these restrictions, minima hopping guided path search (MHGPS) was developed in the course of this thesis. MHGPS is a MH based method for the efficient, automatized and unbiased sampling of complex reaction pathways, which are defined by a sequence of minima and all the transition states connecting them. In the following, a chronological summary of the research conducted for this thesis is given.

The work for this thesis started with the exploration of the PES of neutral and anionic Au_{26} clusters using the MH global optimization method at the DFT level of theory.[61] The computational results suggested that multiple isomers should be observable experimentally. In close collaboration with two research groups in the United States of America, this prediction was confirmed for Au_{26}^-. Moreover, three specific, computationally predicted, geometries could reasonably explain the experimentally measured photoelectron spectra. Additionally, to obtain a rough estimate on the kinetics of the anionic system, a database consisting of minima, transition states and the information, which transition state is connected to which minima was computed at the DFT level of theory. The obtained data allowed to predict Au_{26}^- to exhibit structural fluxionality.

At the time of the study of these gold clusters no method at the DFT level of theory was available to the author for the sampling of complex reaction pathways in a completely automatized, efficient and unbiased fashion and, at the same time, not being prone to getting stuck in some part of a PES. For this reason, the database had to be created in a manual approach. The

experience gained by this manual sampling made clear that the present computer resources can be sufficient for generating databases of useful sizes at ab-initio level. Driven by this experience and the knowledge of the usefulness of such a method to the communities of computational physics, chemistry and biology, the above mentioned MHGPS was developed. Using Lennard-Jones benchmark systems, it was found that, compared to other methods, MHGPS is superior in finding lowest-barrier reaction pathways on complex PESs.[62] In a first application, novel results could be produced for the 75-atom and 102-atom Lennard-Jones systems, despite the fact that these systems already had been studied extensively before.[2,39,63]

Motivated by these results, the coupling of the MHGPS code to the BigDFT package was begun. However, it quickly became clear that the transition state optimization method that was used within the MHGPS code for the study of the Lennard Jones clusters was not efficient enough for simulations at the DFT level of theory. The demand for highly efficient optimizers, that are robust with respect to the noisy forces delivered by DFT codes, resulted in the development of a technique that allows the extraction of significant curvature information from noisy PESs.[64] This technique was used to construct both a stabilized quasi-Newton minimization method and a stabilized quasi-Newton saddle finding approach. With the help of benchmarks, both the minimizer and the saddle finding approach were demonstrated to be superior to existing methods.

With the novel stabilized quasi-newton optimizers at hand, it was possible to finish the coupling of the MHGPS method to the BigDFT code. Both the stabilized quasi-Newton optimizers and the MHGPS method are distributed along with the BigDFT suite under the GNU General Public License and can be downloaded free of charge from the BigDFT website.[65]

Besides the work for the MHGPS method, the author of this thesis also contributed to the development of distance-energy plots, a method that allows the efficient discrimination between glass-like and structure seeker PESs. Based on an empirical approach, it also was found that data from MH runs can be post-processed to obtain a first impression on the qualitative character of a PES. In practice this is useful, as it allows deciding if a certain system might be interesting enough for probing the characteristics of its PES more rigorously by means of the MHGPS code.

The thesis is structured as follows. Probing the characteristics of the PES is the central topic being studied, therefore, in Chap. 1 an introductory overview of the most important concepts related to the PES is given. After discussing relevant related work in the field of local optimizers, the development of the novel stabilized quasi-Newton minimizer and saddle search technique is described in Chap. 2. The MHGPS method is presented in Chap. 3, after having introduced relevant work in the field of reaction pathway search. The study of the gold clusters is detailed in Chap. 4. Finally, Chap. 5 is devoted to the discussion of the efficient post-processing of MH data for the purpose of obtaining a first impression on the characteristics of a PES.

Most of this thesis has been published in peer reviewed journals. The quasi-Newton optimizers and the MHGPS method have been separately published in the Journal of Chemical

Physics.[62,64] The study on the gold clusters has been published in ACS Nano,[61] whereas the distance-energy plots have been published in Physical Review Letters.[66]

1 The Potential Energy Surface

The potential energy of an N-atomic system is given by a real valued function

$$E(\mathbf{R}_1, \ldots, \mathbf{R}_N) : \mathbb{R}^{3N} \mapsto \mathbb{R}. \tag{1.1}$$

Here the vectors $\{\mathbf{R}_i\}_{i=1,\ldots,N}$ represent the positions of the atomic nuclei. The energy $E(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ can be looked at as a surface embedded in a $3N+1$ dimensional space and, therefore, it is commonly denoted as potential energy surface (PES) or potential energy landscape.[3,67] The concept of a potential energy describing the interaction of atomic nuclei is ultimately based upon the Born Oppenheimer approximation for which a brief review is given in Sec. 1.2.1. Detailed knowledge of the PES topology allows the prediction of the equilibrium conformations, thermodynamic and dynamic properties of multiatomic systems, like for example clusters, molecules or bulk.[3,68–70] For that reason, structure prediction and the investigation of chemical reactions are fundamentally based on the study of the PES. As a consequence, the development of novel methods that allow efficient exploration of the PES is a vivid research area.

1.1 Features, Properties and Important Details

A model PES that only depends on two conformational coordinates is visualized in Fig. 1.1. As is apparent from this figure, the PES can be thought of as a mountain landscape. The minima and mountain passes of this energy landscape are landmark points with notable physical significance. The forces $\mathbf{F} \in \mathbb{R}^{3N}$ on all the atoms are given by the negative gradient of the PES $\mathbf{F} = -\nabla E$. The most interesting sites of PESs are stationary points, that is, points at which the forces vanish. The local stability at such points is determined by the eigenvalues of the Hessian matrix. At a local minimum, the Hessian matrix is positive definite, which is equivalent to the condition that the curvatures into all directions are positive. This means, at a local minimum, the energy rises for small displacements into arbitrary directions and, therefore, (meta) stable structures can straightforwardly be identified with local minima. The lowest minimum of all local minima, also denoted as the global minimum, is considered to be the ground state

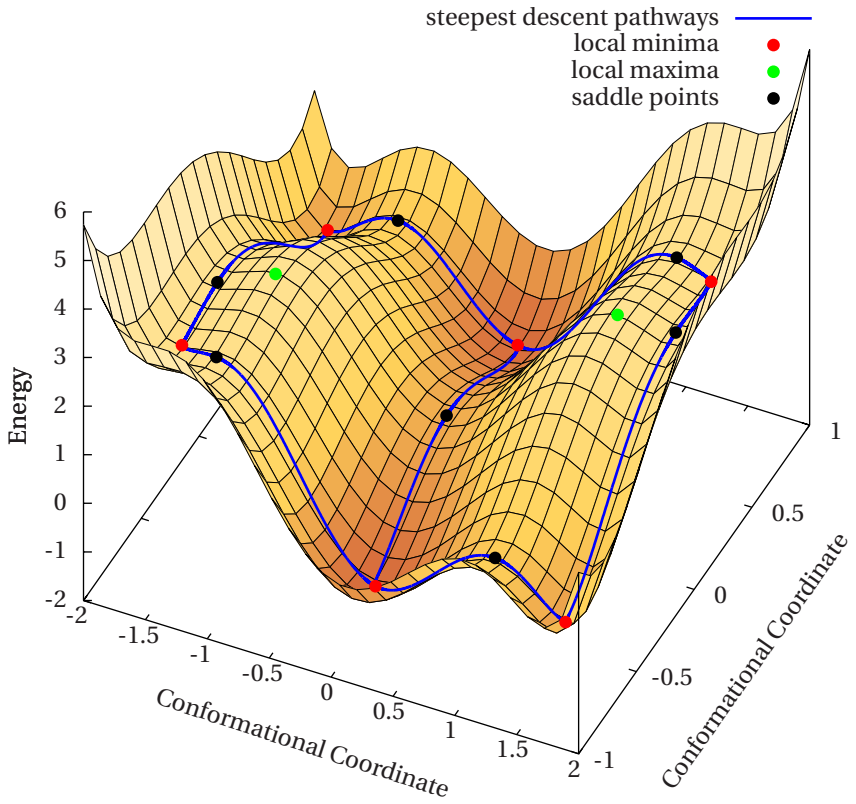

Figure 1.1: A model energy surface (six-hump camel back function[71]) depending on two conformational coordinates. Highlighted are landmark features of the energy surface that have important physical significance. Local minima correspond to (meta)stable states, first order saddle points can be identified with transition states of chemical reactions. Steepest descent pathways leading away from transitions states (blue lines) correspond to the reaction pathways.

structure of a system at vanishing temperatures. This assumption is originated in Anfinsen's thermodynamic hypothesis,[68] which, in a nutshell, states that the conformation of a protein is given by the structure that minimizes the free energy. Indeed, in particular in the research related to proteins and other biomolecules, often a free energy surface is used. However, for other multiatomic systems it is difficult, and in many cases virtually impossible, to define an unambiguous free energy surface, because a suitable definition of collective coordinates frequently is unclear. In the hope that the global free energy minimum is also a low energy minimum of the potential energy, one, therefore, commonly restores to the study of the PES.

For each minimum on a PES a catchment basin (CB) can be defined as the set of points from which steepest descent pathways (see Sec. 2.1.1) converge to this minimum.[72] The transition states are located on the borders that separate the different CBs. By introducing a threshold energy, the catchment basins can be grouped into sets of basins whose minima are mutually accessible without ever exceeding the threshold energy. These mutually accessible sets are

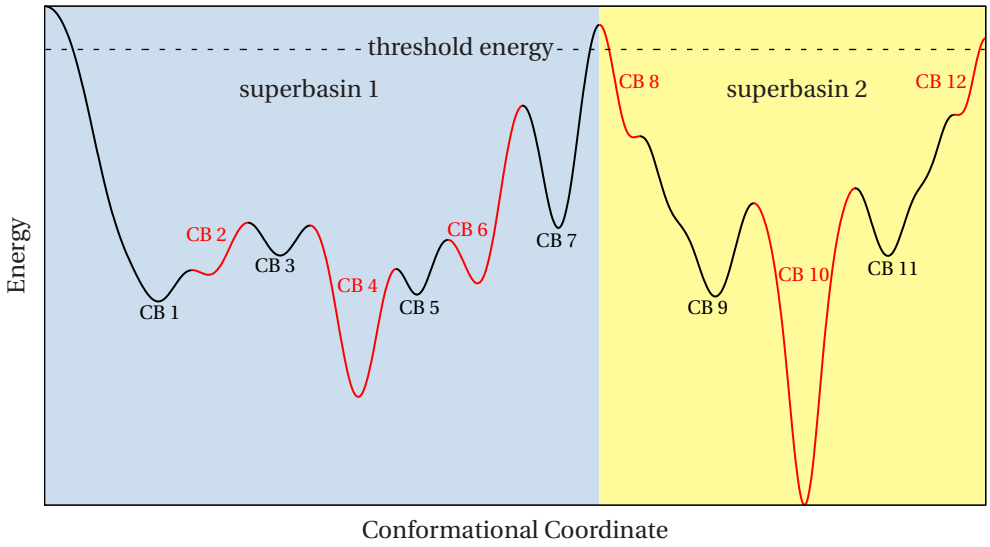

Figure 1.2: A one dimensional model energy surface visualizing the concept of catchment basins (CB) and superbasins. The catchment basins are labeled with their respective number and they are colored alternatingly in black and red. The two different superbasins for the given threshold energy are highlighted by a blue and yellow background, respectively. The threshold energy itself is given by the dashed horizontal line.

denoted as superbasins.[73] Furthermore, a superbasin is denoted as a funnel, if the lowest minimum of the superbasin can be reached by never exceeding a barrier that is significantly larger than the average energy differences of the minima in this superbasin.[3,18] A visualization of these concepts is given in Fig. 1.2. In fact, the idea of partitioning a PES into mutually accessible regions for a given set of different threshold energies can be used to visualize PESs of arbitrary dimensions. This is the basis of the disconnectivity graphs introduced by Becker and Karplus (see Sec. 1.4).[73]

Finding the global minimum of a PES is a formidable task. One reason can be found in the fact, that even for moderately sized atomic systems, the number of local minima is enormously large. Although there is no strict rule that would tell the exact number of local minima for a given system, it is possible to give an estimate for this number. A simple argument was given by Stillinger and Weber.[70] Here, this argument is reiterated in the formulation of Doye and Wales.[74] One considers an atomic system that consists of m subsystems. Each subsystem has N atoms. The system is assumed to be large enough, such that each subsystem, independently from all the other subsystems, can be located in a local minimum. Then, the number of minima n_{min} must fulfill the equation

$$n_{\mathrm{min}}(mN) = n_{\mathrm{min}}(N)^m, \tag{1.2}$$

which, for some system-depended constant α, is solved by

$$n_{\min}(N) = \exp(\alpha N). \tag{1.3}$$

However, not only a large number of local minima can render global minimum searches to be difficult, also the topology of the PES is an important factor. Wales et al. demonstrated in a series of publications,[33–35] that PESs with several energetically low lying regions separated by high potential energy barriers can pose severe problems to global optimization methods. Probably the most prominent example for such a case is LJ_{38}, a system consisting of 38 particles interacting via the Lennard-Jones potential.[35] The same is true for the PESs of glass like systems, that have no well defined lowest local minima but posses a multitude of energetically very similar minima that are separated by a large variety of energetically very different barriers.[36] Binary Lennard-Jones systems of certain sizes are prime examples for such glass-like systems.[4,66]

The trajectory of a chemical reaction that interconverts two local minima can be described as a minimum energy path (MEP). A MEP is a path on the PES for which the gradients at all its points are locally parallel to the path itself. At its energetically highest points, the MEP will pass through stationary points, the mountain passes, at which the Hessian has n negative eigenvalues. Such stationary points are named saddle points of index n. Murrel and Laidler argued[75] that if two minima are connected by a saddle point of index greater than one, then there must exist a lower energy path that involves only saddle points of index one. Their argument, commonly known as Murrel-Laidler-Theorem, can be understood if one realizes that an index n saddle can be regarded as a maximum in the subspace spanned by the Hessian eigenvectors belonging to the n negative eigenvalues. In such a case, any displacement in this subspace will lower the energy (the curvature in this subspace is negative in any direction) and, therefore, it is possible to surmount the index n saddle by a lower energy path. However, their argument implies the assumption of the existence of a Taylor expansion of the PES in terms of the Hessian eigenvectors at the index n saddle.[75,76] Indeed, Wales and Berry showed that there exist pathological cases in which the Murrel-Laidler-Theorem is not applicable, because, for example, a second derivative of such a Taylor expansion is not well defined. In such cases the highest energy point on the lowest energy path connecting two minima is not necessarily an index one saddle.[76] Nevertheless, these cases seem to be rare enough, such that by virtue of the Murrel-Laidler-Theorem, transition states are commonly defined to be saddle points of index one. At a transition state the potential energy is at a maximum with respect to the direction of the Hessian eigenvector corresponding to the negative eigenvalue and at a minimum with respect to all other directions. Therefore, a MEP can be mapped out by stepping away from the transition state in positive and negative direction of the Hessian eigenvector corresponding to the negative eigenvalue, followed by steepest descent iterations with small step sizes (see Sec. 2.1.1).[3,77,78] Examples for such MEPs are given by the blue pathways shown in Fig. 1.1.

One must note that the interconversion between minima is a dynamical process and, therefore,

a system is not constrained to move exclusively along the MEP. However due to Boltzmann's distribution there is an exponential preference towards low energy configurations and, in this sense, the MEP is a reasonable mathematical model. Indeed, already in 1976, Pechukas stated: "There is no dynamical significance to a path of steepest descent. It is a convenient mathematical device to get from high ground, around the transition state, to low ground where the stable molecules are."[79]

The connectivity of a PES is defined by a sequence of minima and transition states connected by a MEP. Two minima are considered to be neighbored (or directly connected) if there exist MEPs between them that only cross one intermediate transition state. The connectivity of a PES can, therefore, be established by sampling all minima, transition states and computing the information which minima are connected by which transition states. For PESs that are expensive to evaluate with respect to the computing time, an explicit tracing of the steepest descent pathways is frequently computationally not feasible. Therefore, one often establishes the connectivity by following pathways defined by, for example, quasi-Newton optimizers, or other advanced minimization techniques.[17,37,40,80,81] Energy minimized and steepest descent pathways usually connect the same minima.[3] Consequently, if one is interested in the connectivity, but not in the details of the MEP itself, it is often sufficient to approximate the MEP by an energy minimized pathway.[3]

Finding all the transition states in the relevant low energy region of a PES usually is even more demanding than global optimization. For one, converging to a transition state is in general more difficult and computationally more expensive than a minimization (see Chap. 2), for another, the number of transition states on a PES is even larger than the number of local minima. The latter has been discussed by Doye and Wales.[74] Here, their argument, which is based on the same idea as the above estimation of the number of minima, is reproduced. Again, a system consisting of m subsystems, each with N atoms, is considered. If the system is large enough, it is reasonable to assume that a transition state can form in one subsystem, while all other subsystems reside in local minima. For the number of transition states n_{ts} this gives the equation

$$n_{\mathrm{ts}}(mN) = m n_{\mathrm{min}}(N)^{m-1} n_{\mathrm{ts}}(N), \tag{1.4}$$

which is solved by

$$n_{\mathrm{ts}}(N) \propto N \exp(\alpha N). \tag{1.5}$$

Under free boundary conditions, the potential energy of a multi-atom system is invariant under overall translations and rotations (assuming the absence of any external potential). The translational invariances cause three eigenvalues of the Hessian matrix to be zero. In Ref. [3] it is shown that the three corresponding eigenvectors $\mathbf{t}_x$, $\mathbf{t}_y$ and $\mathbf{t}_z$ have the form of overall

translations in the $x-$, $y-$ and $z-$direction:

$$\mathbf{t}_x \propto \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \qquad \mathbf{t}_y \propto \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \qquad \mathbf{t}_z \propto \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ \vdots \end{pmatrix}. \tag{1.6}$$

At a stationary point $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots)$ the Hessian of a free molecule has three additional eigenvectors $\boldsymbol{\rho}_x$, $\boldsymbol{\rho}_y$ and $\boldsymbol{\rho}_z$ with vanishing eigenvalues (two for a linear molecule) that correspond to overall rotations around the $x-$, $y-$ and $z-$axis:[3]

$$\boldsymbol{\rho}_x \propto \begin{pmatrix} 0 \\ z_1 \\ -y_1 \\ 0 \\ z_2 \\ -y_2 \\ \vdots \end{pmatrix}, \qquad \boldsymbol{\rho}_y \propto \begin{pmatrix} z_1 \\ 0 \\ -x_1 \\ z_2 \\ 0 \\ -x_2 \\ \vdots \end{pmatrix}, \qquad \boldsymbol{\rho}_z \propto \begin{pmatrix} -y_1 \\ x_1 \\ 0 \\ -y_2 \\ x_2 \\ 0 \\ \vdots \end{pmatrix}. \tag{1.7}$$

There are various applications that need movements along overall translations and rotations to be eliminated. For example, in (thermostated) molecular dynamics simulations, numerical integration artifacts and periodically rescaling of velocities transfers energy from higher frequency modes to the overall translational and rotational degrees of freedom.[82] The Minima Hopping[18] global optimization method, uses (softened) random velocities as escape directions from a local minimum in a short molecular dynamics simulation. If naively generated, these random directions contain components corresponding to overall translations and overall rotations. Movements along those directions are not of interest for the purpose of global minimization. Another example is the saddle finding method described in Sec. 2.3.6. This method makes use of the fact that a direction of minimal curvature on the PES can be found by minimizing the directional curvature function. This is done with the help of finite differences. Contamination of these small finite difference displacements with translations or rotations can slow down convergence towards the direction of minimal curvature. Therefore, in Appx. A several methods for the elimination of overall translations and rotations are discussed.

1.2 Computation of Potential Energy Surfaces

As outlined in the previous chapter, the PES is a potential that describes the interactions of the atomic nuclei. In the framework of the Born-Oppenheimer-Approximation[83], such

a potential arises from an approximate decoupling of the electronic and nuclear degrees of freedom. The potential is given by the eigenvalues of an electronic Schrödinger equation, which in most cases has to be solved numerically. With respect to computational cost and accuracy of the results, one of the most efficient methods to solve the electronic problem is the density functional theory (DFT). At the expense of drastic accuracy losses, but at the gain of several orders of magnitudes of faster computation, the PES can also be modeled by means of classical empirical approximations, so called force fields. The focus of this thesis is the exploration and probing of topological features of PESs. The PES is thus the fundamental object being studied and as such it seems appropriate to outline its calculation, even though the calculation itself is not central to this work. In this section, a brief overview on important theories for the computation of the PES is given.

1.2.1 The Born-Oppenheimer-Approximation

In atomic units, the molecular Schrödinger equation reads

$$\underbrace{[\mathcal{T}_\mathrm{N} + \mathcal{T}_\mathrm{E} + \mathcal{V}_\mathrm{EE} + \mathcal{V}_\mathrm{EN} + \mathcal{V}_\mathrm{NN}]}_{:=\mathcal{H}} \Psi\left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) = E\Psi\left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right), \tag{1.8}$$

where the $\{\mathbf{r}_i\}$ are the coordinates of all N electrons and the $\{\mathbf{R}_j\}$ represent the coordinates of all N_at nuclei. The Hamilton operator is a sum of the operators of the nuclear kinetic energy, the electronic kinetic energy, the electronic Coulomb repulsion, the Coulomb attraction of the electrons and nuclei and the nuclear Coulomb repulsion:

$$\mathcal{T}_\mathrm{N} = -\sum_{i=1}^{N_\mathrm{at}} \frac{1}{2M_i} \nabla^2_{\mathbf{R}_i} \tag{1.9}$$

$$\mathcal{T}_\mathrm{E} = -\sum_{i=1}^{N} \frac{1}{2} \nabla^2_{\mathbf{r}_i}, \tag{1.10}$$

$$\mathcal{V}_\mathrm{EE} = \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i\neq j}}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{1.11}$$

$$\mathcal{V}_\mathrm{EN} = -\sum_{i}^{N} \sum_{j}^{N_\mathrm{at}} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|}, \tag{1.12}$$

$$\mathcal{V}_\mathrm{NN} = \frac{1}{2} \sum_{i=1}^{N_\mathrm{at}} \sum_{\substack{j=1 \\ i\neq j}}^{N_\mathrm{at}} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}. \tag{1.13}$$

Here M_i is the mass of nucleus i and Z_i is the atomic number of the $i-$th nucleus. By neglecting the nuclear kinetic energy, an electronic Hamiltonian $\mathcal{H}_e$ is defined:

$$\mathcal{H}_e = [\mathcal{T}_\mathrm{E} + \mathcal{V}_\mathrm{EE} + \mathcal{V}_\mathrm{EN} + \mathcal{V}_\mathrm{NN}]. \tag{1.14}$$

The corresponding Schrödinger equation reads

$$\mathscr{H}_e \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) = E_k^e \left(\{\mathbf{R}_i\}\right) \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right). \tag{1.15}$$

The electronic Hamiltonian $\mathscr{H}_e$ parametrically depends on the nuclear coordinates and so do its eigenvalues $E_k^e(\{\mathbf{R}_i\})$. Because $\mathscr{H}_e$ is hermitian, its eigenstates $\varphi_k\left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right)$ form a complete (orthonormal) set and any function depending on the electronic coordinates can be expanded in terms of them. In particular, this is true for the electronic dependence of the eigenstates $\Psi\left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right)$ of the complete molecular Schrödinger equation in Eq. 1.8. $\{\mathbf{R}_j\}$:

$$\Psi\left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) = \sum_k \Phi_k \left(\{\mathbf{R}_j\}\right) \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right). \tag{1.16}$$

The $\Phi\left(\{\mathbf{R}_j\}\right)_k$ are the expansion coefficients that depend on the nucleon coordinates. In the following, it is shown that these expansion coefficients can be interpreted as nuclear wave functions. To do so, Eq. 1.16 is inserted into Eq. 1.8. Using the product rule

$$\begin{aligned} \nabla^2_{\mathbf{R}_i} \Phi_k \left(\{\mathbf{R}_j\}\right) \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) = {} & \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) \nabla^2_{\mathbf{R}_i} \Phi_k \left(\{\mathbf{R}_j\}\right) \\ & + 2\nabla_{\mathbf{R}_i} \Phi_k \left(\{\mathbf{R}_j\}\right) \nabla_{\mathbf{R}_i} \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) \\ & + \Phi_k \left(\{\mathbf{R}_j\}\right) \nabla^2_{\mathbf{R}_i} \varphi_k \left(\{\mathbf{r}_i\}, \{\mathbf{R}_j\}\right) \end{aligned} \tag{1.17}$$

and the orthonormality of the electronic eigenstates, one obtains:

$$E_l^e \Phi_l - \sum_{i=1}^{N_{\text{at}}} \frac{1}{2M_i} \nabla^2_{\mathbf{R}_i} \Phi_l + \sum_k \Lambda_{lk} = E\Phi_l, \tag{1.18}$$

where

$$\Lambda_{lk} := \sum_{i=1}^{N_{\text{at}}} \frac{-1}{2M_i} \left(A_{lk} + B_{lk}\right). \tag{1.19}$$

are the so-called non-adiabatic coupling terms. Here, the A_{lk} and B_{lk} terms are defined by:

$$A_{lk} := 2 \int \varphi_l \nabla_{\mathbf{R}_i} \varphi_k d^N\mathbf{r} \nabla_{\mathbf{R}_i} \Phi_k \tag{1.20}$$

$$B_{lk} := \Phi_k \int \varphi_l \nabla^2_{\mathbf{R}_i} \varphi_k d^N\mathbf{r}. \tag{1.21}$$

Now, because the masses of the nuclei are at least three orders of magnitude larger than that of an electron and the electronic wave functions can be assumed to vary only slowly compared to the nuclear wave functions, the non-adiabatic coupling Λ_{lk} terms are neglected.[84] This transforms Eq. 1.18 into Schrödinger equations for the nucleonic wave functions Φ_l, in which the electronic energies $E_l^e(\{\mathbf{R}_i\})$ act as a potential for the nuclei:

$$-\sum_{i=1}^{N_{\text{at}}} \frac{1}{2M_i} \nabla^2_{\mathbf{R}_i} \Phi_l + E_l^e \left(\{\mathbf{R}_j\}\right) \Phi_l = E\Phi_l, \tag{1.22}$$

It is exactly these $E_l^e\left(\{\mathbf{R}_j\}\right)$ that define the PESs introduced above. Care must be taken, if two electronic surfaces are separated by a small energy gap. To see this, the commutator of the nuclear gradient with the electronic Hamiltonian is considered:

$$\int \varphi_l \left[\nabla_{\mathbf{R}_i}\,,\,\mathscr{H}_e\right]\varphi_k d^N\mathbf{r} = \left(E_k^e - E_l^e\right)\int \varphi_l \nabla_{\mathbf{R}_i}\varphi_k d^N\mathbf{r} \tag{1.23}$$

$$= \int \varphi_l \left(\nabla_{\mathbf{R}_i}\mathscr{H}_e\right)\varphi_k d^N\mathbf{r}. \tag{1.24}$$

The last line is just a number (depending on the nuclear positions) and, therefore, the coupling of different electronic states depends inversely on the gap between the electronic surfaces.[85]

All the methods outlined or developed in this thesis assume electronic excitations to be negligible and, therefore, operate on the ground state PES $E_0^e\left(\{\mathbf{R}_j\}\right)$. The corresponding electronic problem is assumed to be solved. In fact, due to the high dimensionality of the electronic wave functions this problem is quite a difficult one and the basis of several research areas. Numerous methods like Configuration Interaction, which in its simplest form reduces to Hartree Fock, Coupled Cluster, Møller-Plesset perturbation theory, Quantum Monte Carlo or Density Functional theory have been developed to solve the electronic problem.[86–88] Due to its favorable balance between accuracy and computational cost, the Density Functional Theory is probably one of the most used approaches. In the following section, a short review of this theory is given.

1.2.2 Density Functional Theory

The electronic problem reads

$$\underbrace{\left(\mathscr{T}_\mathrm{E} + \mathscr{V}_\mathrm{EE} + \mathscr{V}_\mathrm{EN}\right)}_{\mathscr{H}^{e\prime}}\varphi_j\left(\{\mathbf{x}_i\}\right) = \varepsilon_j\varphi_j\left(\{\mathbf{x}_i\}\right). \tag{1.25}$$

Here, the constant energy shift introduced by $\mathscr{V}_\mathrm{NN}$ has been transferred from the electronic Hamiltonian $\mathscr{H}^e$ to the eigenvalue ε_j. The corresponding shift has been emphasized by using a prime at the electronic Hamiltonian $\mathscr{H}^{e\prime}$ in Eq. 1.25. However, for simplicity, the prime will be omitted from now on and $\mathscr{H}^e$ is written instead of $\mathscr{H}^{e\prime}$. Additionally, the spin-dependency is introduced via the collective variables $\mathbf{x}_i$ that represent the continuous spatial coordinates $\mathbf{r}_i$ and the discrete spin coordinates s_i. The parametric dependence on the nuclear coordinates has been omitted in the above equation. Henceforth, in agreement with common practice, integrals like $\int ds_i$ will be understood as a summation over the spin coordinates.

In principle, the ground state of the electronic problem could be found by searching for the wave function that minimizes the expectation value of the energy

$$E[\Psi] := \frac{\langle\Psi|\mathscr{H}^e|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{1.26}$$

However, in terms of the $3N$ dimensional wave functions (considering the spatial part), such a minimization is computationally not feasible. For example, a simple discretization of the wave function into K points for each degree of freedom results in a memory requirement that scales like K^{3N}. Even for moderately sized systems, the required amount of memory is not available on even the largest existing computers.

To circumvent this problem, one could try to express the energy as a functional depending on the electron density $\rho(\mathbf{r})$*

$$\rho(\mathbf{r}) = N \int \dots \int |\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)|^2 d\mathbf{s}_1 d\mathbf{x}_2 d\mathbf{x}_3 \dots d\mathbf{x}_N, \tag{1.27}$$

which gives the probability $\rho(\mathbf{r})\, d\mathbf{r}$ to find an electron in a volume element $d\mathbf{r} = dxdydz$ around $\mathbf{r}$. This way, the variational problem of Eq. 1.26 could be recast into a minimization over densities which is a problem depending on just three degrees of freedom. Indeed, already in 1927, this idea was followed by Thomas and Fermi who developed the Thomas-Fermi-Method.[89,90] However, mainly due to the difficulties that exist with expressing the exact kinetic energy as an explicit functional of the electron density, the Thomas-Fermi-Method is not very accurate. Furthermore, the rigorous mathematical footing of replacing the wave function by the electron density was not established until 1964, when the two famous theorems of Hohenberg and Kohn had been published. The two theorems read:[91]

Theorem 1 (The Density as Basic Variable)**.** *The external potential V_{EN} is a unique functional of the ground state electron density $\rho_0(\mathbf{r})$, apart from a trivial additive constant.*

Theorem 2 (The Variational Principle)**.** *Define for a given external potential the energy functional*

$$E[\rho] := \int V_{EN}(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r} + F[\rho], \tag{1.28}$$

where $F[\rho]$ is a universal functional independent of the external potential and thus applicable to any many-electron system and ρ is the ground state density of some external potential.† *Let E_0 be the ground-state energy of $\mathcal{H}^e$. Then, for any density $\tilde{\rho}(\mathbf{r})$ in the domain of $E[\rho]$ such that*[94] *$\tilde{\rho}(\mathbf{r}) \geq 0$ and $\int \tilde{\rho}(\mathbf{r})\, d\mathbf{r} = N$,*

$$E_0 \leq E[\tilde{\rho}]. \tag{1.29}$$

For the proof of the theorems, it is first noted that the solution of the Schrödinger equation defines a surjective map A from the set of external potentials $\{\mathcal{V}_{\mathrm{EN},i}\}$ (that are assumed to be mutually different by more than a constant) to the set of corresponding ground state wave functions $\{\varphi_{0,i}\}$. These ground state functions are assumed to be non-degenerate. Via 1.27, a second surjective map B from these ground state wave functions to the set of their

*Normalized wave functions are assumed.

†Levy and Lieb[84,92,93] extended the definition of the universal functional by showing that it can be defined for any electron density $\rho(\mathbf{r})$, which can be derived from a N-electron wavefunction.

corresponding electron densities is defined. To show that A is also injective, it is assumed that two different external potentials $\mathcal{V}_{\mathrm{EN},i}$ and $\mathcal{V}_{\mathrm{EN},j}$ with $\mathcal{V}_{\mathrm{EN},i} - \mathcal{V}_{\mathrm{EN},j} \neq c$ for some constant c lead to the same ground state wave functions $\varphi_{0,i} = \varphi_{0,j}$. Then

$$\mathcal{H}_i^e \varphi_{0,i} - \mathcal{H}_j^e \varphi_{0,j} = \varepsilon_i \varphi_{0,i} - \varepsilon_j \varphi_{0,j} \tag{1.30}$$

$$\Rightarrow \quad \left(\mathcal{H}_i^e - \mathcal{H}_j^e\right) \varphi_{0,i} = \left(\varepsilon_i - \varepsilon_j\right) \varphi_{0,i}. \tag{1.31}$$

However, this contradicts the assumption of $\mathcal{V}_{\mathrm{EN},i} - \mathcal{V}_{\mathrm{EN},j} \neq c$ and, therefore, A is also injective.[95] For the injectiveness of the map B, the proof of Hohenberg and Kohn[91] is reproduced in the following. One assumes two different (non-degenerate) ground state functions $\varphi_{0,i}$ and $\varphi_{0,j}$, corresponding to external potentials $\mathcal{V}_{\mathrm{EN},i}$ and $\mathcal{V}_{\mathrm{EN},j}$, to produce the same ground state density ρ_0. Then by virtue of the variational principle

$$\varepsilon_{0,i} = \langle \varphi_{0,i} | \mathcal{H}_i^e | \varphi_{0,i} \rangle \tag{1.32}$$

$$< \langle \varphi_{0,j} | \mathcal{H}_i^e | \varphi_{0,j} \rangle \tag{1.33}$$

$$= \langle \varphi_{0,j} | \mathcal{H}_j^e + \mathcal{V}_{\mathrm{EN},i} - \mathcal{V}_{\mathrm{EN},j} | \varphi_{0,j} \rangle \tag{1.34}$$

$$= \varepsilon_{0,j} + \langle \varphi_{0,j} | \mathcal{V}_{\mathrm{EN},i} - \mathcal{V}_{\mathrm{EN},j} | \varphi_{0,j} \rangle \tag{1.35}$$

$$= \varepsilon_{0,j} + \int \rho_0(\mathbf{r}) \left[V_{\mathrm{EN},i}(\mathbf{r}) - V_{\mathrm{EN},j}(\mathbf{r}) \right] d\mathbf{r}. \tag{1.36}$$

The same arguments apply if i and j are interchanged and thus

$$\varepsilon_{0,j} < \varepsilon_{0,i} - \int \rho_0(\mathbf{r}) \left[V_{\mathrm{EN},i}(\mathbf{r}) - V_{\mathrm{EN},j}(\mathbf{r}) \right] d\mathbf{r}. \tag{1.37}$$

Adding Eq. 1.36 to Eq. 1.37 leads to the contradiction $\varepsilon_{0,i} + \varepsilon_{0,j} < \varepsilon_{0,j} + \varepsilon_{0,i}$. Hence, A and B are both surjective and injective and, therefore, there exists a unique one-to-one map between the ground state density ρ_0 and the external potential $\mathcal{V}_{\mathrm{EN}}$. Thereby, all wave functions are determined by the ground state density and all properties of the system can be written as a functional of ground state densities that are obtained by the solution of the Schrödinger equation. In particular, this is true for the total energy expectation value

$$E[\rho_0] = \underbrace{T_{\mathrm{E}}[\rho_0] + E_{\mathrm{EE}}[\rho_0]}_{:=F[\rho_0]} + \int V_{\mathrm{EN}}(\mathbf{r}) \rho_0(\mathbf{r}) \, d\mathbf{r}. \tag{1.38}$$

From this, the variational principle of Theorem 2 follows as a direct consequence of the one-to-one relation between the ground state density and the ground state wave function.

The Kohn-Sham Ansatz

Following the Hohenberg-Kohn theorems, the ground state density of a many-electron system and, with it, all its properties, can be found by minimizing the energy functional of Eq. 1.28. However, the exact form of the universal functional $F[\rho_0]$ is unknown. Furthermore, explicit

functionals of the density for the kinetic energy and the non-classical parts of the electron-electron interaction are unknown.[84] By introducing an auxiliary system of non-interacting electrons, Kohn and Sham circumvented these problems in their seminal work of 1965.[96] This non-interacting system is characterized by a Slater determinant comprised of Kohn-Sham orbitals ϕ_i. This system is *assumed* to have exactly the same ground state density as the interacting system of Eq. 1.25. Then, the total energy functional of Eq. 1.28 can be rewritten as:[95]

$$E\left[\rho\right] = T_{\mathrm{S}}\left[\rho\right] + E_{\mathrm{H}}\left[\rho\right] + E_{\mathrm{EN}}\left[\rho\right] + E_{\mathrm{XC}}\left[\rho\right], \tag{1.39}$$

where

$$T_{\mathrm{S}}\left[\rho\right] := -\frac{1}{2}\sum_{i,s}\langle\phi_i|\nabla^2|\phi_i\rangle, \tag{1.40}$$

$$E_{\mathrm{H}}\left[\rho\right] := \frac{1}{2}\int\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}', \tag{1.41}$$

$$E_{\mathrm{EN}}\left[\rho\right] := \int V_{\mathrm{EN}}\left(\mathbf{r}\right)\rho\left(\mathbf{r}\right)d\mathbf{r}, \tag{1.42}$$

$$E_{\mathrm{XC}}\left[\rho\right] := T_{\mathrm{E}}\left[\rho\right] - T_{\mathrm{S}}\left[\rho\right] + E_{\mathrm{EE}}\left[\rho\right] - E_{\mathrm{H}}\left[\rho\right], \tag{1.43}$$

$$\rho\left(\mathbf{r}\right) := \sum_{i,s}|\phi_i\left(\mathbf{r},s\right)|^2. \tag{1.44}$$

Here, T_{S} is the kinetic energy of a non-interacting system. Via the Kohn-Sham orbitals it is an implicit functional of the electron density. The Hartree energy E_{H} is the classical Coulomb interaction energy of the electron density interacting with itself. The exchange correlation functional E_{XC} serves as a container for everything that is unknown and that cannot be treated rigorously. For example, the independent-particle kinetic energy $T_{\mathrm{S}}\left[\rho\right]$ is not the exact kinetic energy of the interacting system. However, its correction to the exact kinetic energy of the interacting system, which is stowed away in E_{XC}, is usually small.[95] Similar considerations apply to the non-classical electron-electron interaction that cannot be represented by the Hartree energy.

Varying Eq. 1.39 with respect to a ϕ_i^* under the constraint of normalized orbitals yields, after a unitary transformation of the orbitals, Schrödinger-like Kohn-Sham differential equations for the Kohn-Sham orbitals[94]

$$\left(-\frac{1}{2}\nabla^2 + v\left(\mathbf{r}\right)\right)\phi_i = \epsilon_i\phi_i, \tag{1.45}$$

where the Kohn-Sham potential $v\left(\mathbf{r}\right)$ is given by

$$v\left(\mathbf{r}\right) := V_{EN}\left(\mathbf{r}\right) + \int\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \underbrace{\frac{\delta E_{\mathrm{XC}}\left[\rho\left(\mathbf{r}\right)\right]}{\delta\rho\left(\mathbf{r}\right)}}_{:=V_{\mathrm{XC}}(\rho(\mathbf{r}))}. \tag{1.46}$$

Solving Eq. 1.45 gives, via the Kohn-Sham orbitals, the electron density:

$$\rho(\mathbf{r}) = \sum_{i,s} |\phi_i(\mathbf{r}, s)|^2. \tag{1.47}$$

However, the left-hand side operator of Eq. 1.45 depends on the density and, therefore, the problem has to be solved self-consistently.

What remains to discuss is how to approximate the exchange-correlation functional E_{XC}. In Eq. 1.39, the independent electron kinetic energy T_{S} and the long-range Hartree energy E_{H} have been separated from the exchange-correlation functional E_{XC}, which allows to approximate E_{XC} as a local functional of the density.[84]

In case of the local density approximation (LDA), the exchange correlation energy is written as (neglecting spin polarization)

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\, \epsilon_{\mathrm{XC}}^{\mathrm{hom}}(\rho(\mathbf{r}))\, d\mathbf{r}, \tag{1.48}$$

where $\epsilon_{\mathrm{XC}}^{\mathrm{hom}}$ is the exchange and correlation energy per electron of a homogeneous electron gas with density ρ.[96] It can be split in a sum corresponding to contributions of exchange and correlation $\epsilon_{\mathrm{XC}}^{\mathrm{hom}}(\rho(\mathbf{r})) = \epsilon_{\mathrm{X}}^{\mathrm{hom}}(\rho(\mathbf{r})) + \epsilon_{\mathrm{C}}^{\mathrm{hom}}(\rho(\mathbf{r}))$.[97] An expression for the exchange part $\epsilon_{\mathrm{X}}^{\mathrm{hom}}$ of the homogeneous electron gas is analytically known and the correlation part $\epsilon_{\mathrm{C}}^{\mathrm{hom}}$ of the homogeneous electron gas was fitted to highly accurate quantum Monte Carlo results.[98–102] The LDA exchange-correlation functional works best for slowly varying electron densities.[84] A more accurate description, in particular for systems with a more rapidly varying electron density, is available with the exchange-correlation functionals of the generalized gradient approximation (GGA) family. Here, the exchange correlation density also depends on the magnitude of the electron density gradient $|\nabla\rho|$:[103–105]

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho] = \int \rho(\mathbf{r})\, \epsilon_{\mathrm{XC}}^{\mathrm{hom}}(\rho(\mathbf{r}, |\nabla\rho|))\, d\mathbf{r}. \tag{1.49}$$

By additionally incorporating the kinetic energy density (meta-GGA functionals),[97] or by mixing a portion of the exact Hartree Fock exchange with exchange correlation contributions from GGA and / or LDA (hybrid functionals)[106–108] even more accurate exchange correlation functionals can be obtained.

1.3 Force Fields

The Kohn-Sham DFT discussed in the previous chapter provides an excellent compromise between computational efficiency and physical accuracy. For example, nowadays unbiased and systematic searches for the most stable atomic configuration or the systematic search for a reaction pathway is possible at DFT level for reasonably sized system. Nevertheless, DFT simulations that go beyond a few hundred atoms quickly become computationally intractable,

especially, if more than only a few energy and force evaluations have to be performed. At the loss of accuracy, one can restore to significantly faster methods such as semi-empirical tight-binding methods or even to the completely empirical force fields. Force fields are analytic parametrizations of PESs, which are fitted to experimental data or to the results of very accurate quantum mechanical calculations. As a consequence of their empirical or semi-empirical foundation, force fields and tight-binding methods are not available for many materials. Furthermore, it was demonstrated for silicon that, in particular, force fields exhibit numerous spurious local minima that do not exist at the DFT level of theory.[109] Also the energetic ordering of the local minima can change dramatically if going from force field level to more accurate descriptions of the PES.[110] Therefore, in the case of structure or reaction pathway prediction the usefulness of force fields is limited.

Nonetheless, force fields were essential for the present work, because energy and force evaluations based on force fields are several orders of magnitudes faster than their DFT counterparts. Furthermore, force field based interactions can feature significant properties of real systems such as frustration or, as in biomolecules, the concurrent existence of tiny and large force constants. Therefore, force fields are ideal testing grounds for novel atomistic methods.

There exists a whole zoo of different force fields. In most cases, each force field is tinkered to a specific system or class of systems. It is beyond the scope of this thesis to give a broad review of the existing force fields and, therefore, only force fields that were used in the present work are mentioned in the following.

Presumably one of the simplest force fields is the Lennard-Jones (LJ) potential[111,112]

$$E_{\mathrm{LJ}} = 4 \sum_{i<j} \epsilon_{i,j} \left\{ \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right\}, \tag{1.50}$$

here $\epsilon_{i,j}$ defines the pair-well depth, $2^{1/6}\sigma_{i,j}$ is the pair-well equilibrium distance and $r_{i,j}$ the distance between the atoms i and j. In this thesis, the energy and distance parameters are understood to be independent of the particles indices if it is spoken of the LJ potential and in these cases energies and distances are reported in units of ϵ and σ. In Chap. 5 the binary LJ (BLJ) potential is used. Here A and B type particles exist and the $\epsilon_{i,j}$ and $\sigma_{i,j}$ can take on three different values, depending on whether they are related to a $A-A$, $B-B$ or $A-B$ interaction. The van der Waals interaction of noble gases is reasonably described by the LJ potential and various parameters can be found in the literature.[113–117]

A slightly more complex function, the Born-Mayer-Huggins-Tosi-Fumi potential (BMHTF),[118–123] can be used to describe the ionic interactions in alkali halides:

$$E_{\mathrm{BMHTF}} = \sum_{i<j} c_{i,j}\, b \exp\left[\frac{\sigma_i + \sigma_j - r_{i,j}}{\rho} \right] - \frac{C_{i,j}}{r_{i,j}^6} + \frac{D_{i,j}}{r_{i,j}^8} - \frac{Z_i Z_j}{r_{i,j}}, \tag{1.51}$$

where b is the same for all salts and $c_{i,j}$ are Pauling's numerical parameters.[123,124] The lengths

scale σ_i and σ_j, the hardness parameters ρ, the dipole-dipole coefficients $C_{i,j}$ and the dipole-quadrupole coefficients $D_{i,j}$ are adjustable parameters depending on the system. Z_i and Z_j are the ionic charges.

For most systems, the potential energy is a function that depends on more than just pairwise distances. Indeed, besides bond-stretchings, many chemical force fields include bond-bending and torsional terms. A good example for a force field that contains all the common terms is the Assisted Model Building with Energy Refinement (AMBER) force field that is widely used for biomolecules:[125,126]

$$E_{\text{AMBER}} = \sum_{\text{bond}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsional}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{i,j}}{r_{r,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^6} + \frac{Z_i Z_j}{\epsilon r_{i,j}} \right]. \tag{1.52}$$

The first two terms describe harmonic approximations to the bond-stretchings and bond-bendings. K_r and K_θ are the corresponding force constants and the equilibrium bond distances and bond angles are given by r_{eq} and θ_{eq}, respectively. The torsional rotations in the third term are expanded in terms of short Fourier series and the last term accounts for the van der Waals interaction and electrostatic Coulomb energies, including a dielectric constant ϵ that allows for implicit non-vacuum environments.

For technologically important materials, such as silicon or carbon, there also exist force fields that allow for bond breaking or changes in hybridization. Neither bonds between atoms nor hybridization angles have to be specified explicitly. However, they typically exist only for systems that consist of only a single type of atoms. One of the best available force fields for silicon is the Lenosky force field.[127,128] Both the Lenosky as well as the AMBER force field were used for the benchmarks in Chap. 2.

1.4 Disconnectivity Graphs

Disconnectivity graphs introduced by Becker and Karplus[73] and frequently used and illustrated by Wales et al.[1,3,40,129] can be used to visualize multidimensional PESs. They, therefore, allow to obtain a rough, intuitive insight into dynamic properties. In this section, the theory of disconnectivity graphs is briefly recapitulated.

Disconnectivity graphs illustrate which minima are connected by reaction pathways where the energy never exceeds a given threshold energy. Such mutually accessible regions are called 'superbasins'.[40] The number of superbasins depends on the threshold energy. The vertical axis of a disconnectivity graph is partitioned into a predefined and freely chosen number of equidistant energy threshold levels E_i, such that $E_{i+1} > E_i$. At each threshold energy it is analyzed which minima form which superbasins. The superbasins are represented by nodes on the graph and are arranged along the horizontal axis that corresponds to their threshold

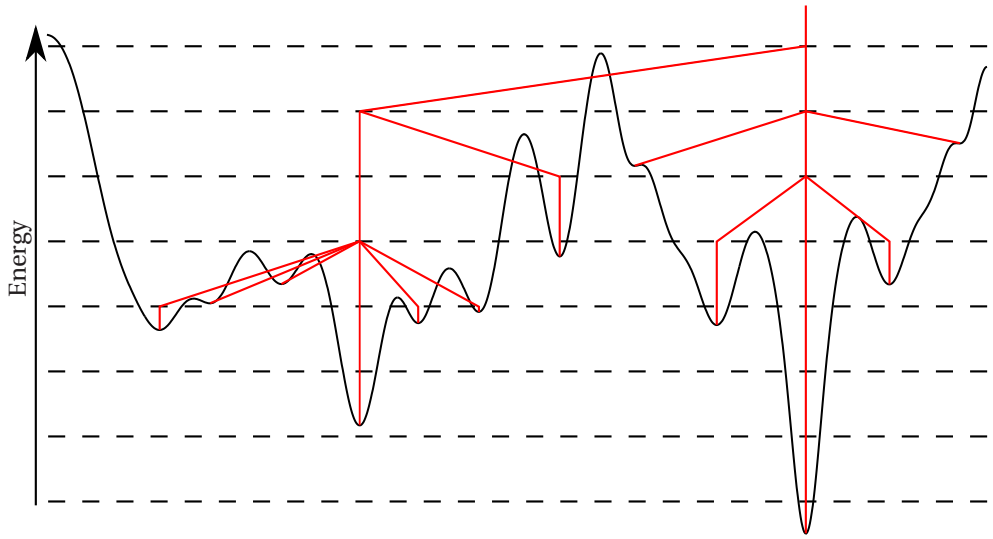

Figure 1.3: This figure shows the construction of a disconnectivity graph (red tree-like graph) for a one-dimensional model energy landscape. The energy thresholds are visualized by the horizontal black dashed lines.

energy. Two nodes at energy levels E_{i+1} and E_i are connected with a line, if they belong to the same superbasin at the higher energy E_{i+1}. Finally, all the single minima at the bottoms of the superbasins are represented separately by drawing lines down to their respective energies. The horizontal position of the nodes and minima is arbitrary. Typically there are too many minima to visualize, hence only the lowest n minima are usually plotted. Nevertheless, all minima and transition states contained in the underlying stationary point database contribute to the superbasin and barrier analysis. At the examples of the one-dimensional model potential energy landscape that already was used in Fig. 1.2, the construction of a disconnectivity graph is visualized in Fig. 1.3.

The plots of all disconnectivity graphs in this thesis were generated using the disconnectionDPS[129] software.

2 Local Optimization*

Stationary points are the most interesting and most important points of potential energy surfaces (PESs). The relative energies of local minima and their associated configuration space volumes determine thermodynamic equilibrium properties.[3] According to transition state theory, dynamical properties can be deduced from the energies and the connectivity of minima and transition states.[130] Therefore, the efficient determination of stationary points of PESs is of great interest to the communities of computational chemistry, physics, and biology. Clearly, optimization and, in particular, minimization problems are present in virtually any field. This explains why the development and mathematical characterization of iterative optimization techniques are important and longstanding research topics, which resulted in a number of highly sophisticated methods like, for example, conjugate gradient (CG),[131], the fast inertial relaxation engine (FIRE)[132] or quasi-Newton methods like the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[133–136] and its limited memory variant (L-BFGS).[137,138] Since for a quadratic function Newton's method is guaranteed to converge within a single iteration, it is not surprising that the BFGS and L-BFGS algorithms belong to the most efficient methods for minimizations of atomic systems.[3]

Atomic interactions are bounded from below. Therefore, in practice, descent directions are safe routes towards close-by local minima. Furthermore, the curvature at a minimum is positive in all directions. This means, all directions can be treated on the same footing during a minimization. The situation is different for saddle point optimizations. A saddle point is a stationary point at which the PES is at a maximum with respect to one or more particular directions, and at a minimum with respect to all other directions. Close to a saddle point, it is, therefore, not possible to treat all directions on the same footing. Instead, one has to single out the directions that have to be maximized. Furthermore, far away from a saddle point it is usually impossible to tell which search direction guarantees to finally end up in a saddle point. Therefore, saddle point optimizations typically are more demanding and significantly less reliable than minimizations. Saddle point finding algorithms can be roughly classified into

*Parts of this chapter have been published in B. Schaefer, S. A. Ghasemi, S. Roy, and S. Goedecker, "Stabilized Quasi-Newton Optimization of Noisy Potential Energy Surfaces", The Journal of Chemical Physics **142**, 034112 (2015). Reprinted with permission. Copyright 2015 by the American Institute of Physics.

single-ended and double-ended methods. Single-ended methods like the dimer method[55–58] or Wales' hybrid eigenvector following[49,50] start their search for a saddle point at some location on the PES, whereas double-ended methods find one or multiple saddle points between two given structures, which, in most cases, are local minima. Most single ended searches exploit in some way the idea of following the lowest eigenvector of the Hessian matrix, an idea that dates back to a 1971 publication of Crippen and Scheraga.[139] This rough idea of how to find a saddle point gives ample scope for the actual realization of such a method. Consequently, there are large performance differences between different eigenvector following methods, as is demonstrated in Sec .2.3.7.

In this chapter, first an overview of existing methods for finding minima and saddle points is given. Finally, in Sec. 2.3, a novel stabilized quasi-Newton method suitable for finding minima and saddle points on noisy PESs is presented and benchmarked.[64]

2.1 Local Minimization

2.1.1 Steepest Descent

As has been outlined in Chap. 1, the PES of an N-atomic system is a map $E : \mathbb{R}^{3N} \mapsto \mathbb{R}$ that assigns to each atomic configuration $\mathbf{R}$ a potential energy. Probably the most obvious approach to minimize the energy of a given point $\mathbf{R}^i$ on the PES is by going a small step into the direction of greatest descent, which is given by the negative gradient $-\nabla E\left(\mathbf{R}^i\right)$ at this point. The iterations of the steepest descent method are given by

$$\mathbf{R}^{i+1} = \mathbf{R}^i - \alpha_i \nabla E\left(\mathbf{R}^i\right), \tag{2.1}$$

where the positive real number α_i is denoted as the "step size". The step size α_i must be chosen such that the energy decreases in all steps. The optimal step size can be found by means of a line search, that is, by minimizing the function $\varepsilon(\alpha_i) = E\left(\mathbf{R}^i - \alpha_i \nabla E\left(\mathbf{R}^i\right)\right)$. In most cases this cannot be done analytically and, therefore, an iterative method has to be used. However, such iterative line searches require several energy evaluations, which frequently is computationally not efficient, in particular, if the evaluation of the PES is computationally expensive. In this case, reasonable step sizes can be found by an energy or gradient feedback. In the case of the energy feedback, α is slightly increased (for example by 5%) if the energy decreases. If the energy increases, the step size is at least twice as large as the optimal step size (assuming a quadratic function). Therefore, the stepsize should be decreased by a factor of $\frac{1}{2}$. For the gradient feedback, the basic idea is that consecutive gradients should point into similar directions, if the step size is not too large. Therefore, the step size is slightly increased if the angle between consecutive gradients is smaller than, for example, 60°. Otherwise, the step size is reduced by a factor of $\frac{1}{2}$. In any case, the initial step size should be chosen as the inverse of the largest eigenvalue.

The steepest descent method is straight forward to implement and, in combination with

conservative step sizes, usually more reliable than other, more advanced, methods.[140] Unfortunately, steepest descent becomes very inefficient if the optimization problem is ill-conditioned, that is, if the spectrum of the Hessian matrix spans a range of several orders of magnitude. The problem is that the number of iterations needed by the steepest descent method scales linearly with the condition number κ in the quadratic region of a function. The condition number is defined as the ratio of the largest to the smallest Hessian eigenvalue.[141] The poor efficiency of the steepest descent method is intuitively accessible. When assuming a quadratic form (see Fig. 2.1), the gradients at points on the principal axis are collinear to the principal axis. This means, the optimal step size for points on the principal axis is simply the inverse of the corresponding Hessian eigenvalue (cf. Sec. 2.1.2). For points that are not located on a principal axis, one has to be conservative and the step size has to be chosen as the inverse of the largest Hessian eigenvalue. For ill-conditioned problems this conservative step size will be much too small for directions corresponding to small Hessian eigenvalues and the steepest descent method tends to approach the minimum in excessive "zigzag" moves (see Fig. 2.1). The problem of ill-conditioning can be alleviated by transforming the coordinates such that, in the best case, all the curvatures equalize after the transformation. For this, a linear coordinate transformation $\mathbf{R} = A\mathbf{S}$, defined by an invertible $3N \times 3N$ square matrix A, is considered.[142,143] The gradient of the coordinate transformed PES $\tilde{E}(\mathbf{S}) = E(A\mathbf{S})$ is given by

$$\nabla_{\mathbf{S}} \tilde{E} = A^T \nabla_{\mathbf{R}} E. \tag{2.2}$$

In the new coordinates, the steepest descent method looks like

$$\mathbf{S}^{i+1} = \mathbf{S}^i - A^T \nabla_{\mathbf{R}} E. \tag{2.3}$$

Going back to the old coordinates, one obtains

$$\mathbf{R}^{i+1} = \mathbf{R}^i - \underbrace{AA^T}_{:=P} \nabla_{\mathbf{R}} E, \tag{2.4}$$

where P is commonly denoted as the preconditioning matrix.[144] The optimal preconditioning matrix corresponds to a coordinate transformation that leads to a condition number $\kappa = 1$. As will be seen in Sec. 2.1.2, the optimal preconditioning matrix for an exactly quadratic function is given by the inverse Hessian.

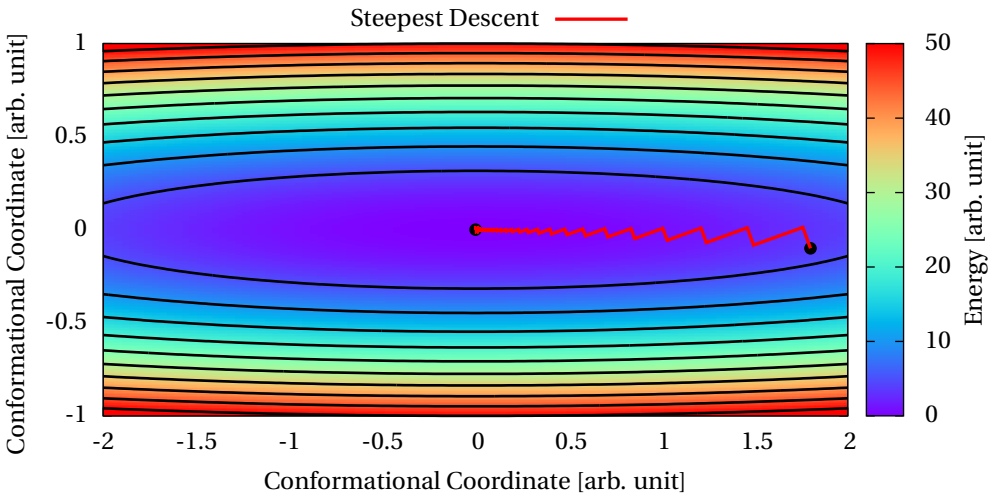

Figure 2.1: Sequence of steepest descent iterates using optimal step sizes on a two-dimensional quadratic model function. The condition number for this model function is $\kappa = 50$. The "zigzag" pattern is characteristic of the steepest descent method.

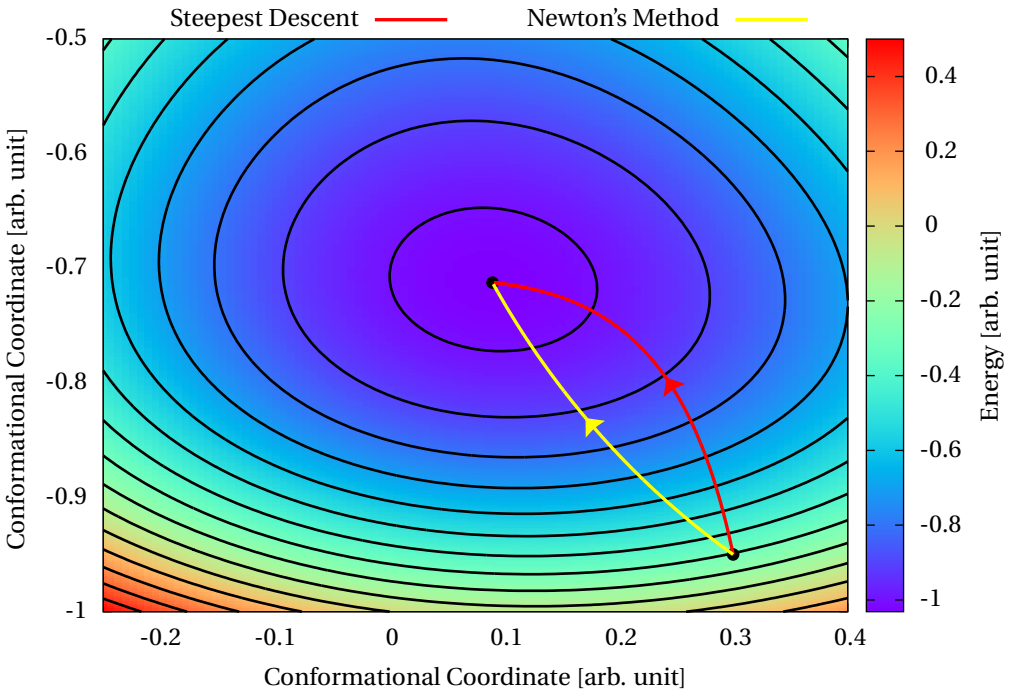

Figure 2.2: Trajectories for the steepest descent (red) and Newton's method (yellow) in the limit of small step sizes on a non-quadratic model PES. On a perfectly quadratic PES, the trajectory of Newton's method would be transformed to a straight line.

2.1.2 Newton's and Quasi-Newton's Method

For the derivation of Newton's Method, it is assumed that a second order expansion of the PES $E(\mathbf{R})$ about a point $\mathbf{R}^i$ is possible:

$$E(\mathbf{R}) \approx E\left(\mathbf{R}^i\right) + \left[\mathbf{R}-\mathbf{R}^i\right]^T \nabla E\left(\mathbf{R}^i\right) + \frac{1}{2}\left[\mathbf{R}-\mathbf{R}^i\right]^T H_{\mathbf{R}^i}\left[\mathbf{R}-\mathbf{R}^i\right] \tag{2.5}$$

$$\nabla E(\mathbf{R}) \approx \nabla E\left(\mathbf{R}^i\right) + H_{\mathbf{R}^i}\left[\mathbf{R}-\mathbf{R}^i\right]. \tag{2.6}$$

Here, $H_{\mathbf{R}^i}$ is the Hessian of the PES evaluated at $\mathbf{R}^i$. If $\mathbf{R}$ is a stationary point, the left-hand side gradient of Eq. 2.6 vanishes and Newton's method for minimization follows:

$$\mathbf{R}^{i+1} = \mathbf{R}^i - H_{\mathbf{R}^i}^{-1} \nabla E\left(\mathbf{R}^i\right) \tag{2.7}$$

In the previous equation $\mathbf{R}$ was renamed to $\mathbf{R}^{i+1}$ in order to emphasize the iterative character of Newton's Method for non-quadratic PESs. For a quadratic form, the expansion in Eq. 2.5 and Eq. 2.6 is exact and Eq. 2.7 will solve the minimization problem in a single step. This is equivalent to the perfectly preconditioned steepest descent mentioned at the end of Sec. 2.1.1.

Care must be taken when starting the minimization in a region in that the Hessian is not positive definite. In these cases, the step direction defined in Eq. 2.7 is not a descent direction and, if the current region is too steep, the displacements $|\mathbf{R}^{i+1} - \mathbf{R}^i|$ may become too large. The introduced instabilities can be eliminated by replacing the Hessian eigenvalues with their absolute value and by explicitly limiting the maximum displacement.

Fig. 2.2 displays the trajectories of Newton's method and of the steepest descent approach on a non-quadratic model PES in the limit of small step sizes. Newton's method tends to take the more direct route towards the minimum, which in case of a perfectly quadratic PES were a straight line. Methods exploiting the step directions defined by Eq. 2.7 often converge superlinearly, or even quadratically.[141]

In practice, it is in many cases either impossible to calculate an analytic Hessian or it is too time consuming to compute it numerically at every iteration by means of finite differences. Therefore, quasi-Newton methods use an approximation to the exact Hessian that is computationally less demanding. Using a constant multiple of the identity matrix as an approximation to the Hessian results in the simple steepest descent method of Sec. 2.1.1. In most cases, such a choice is a very poor approximation to the true Hessian. However, improved approximations can be generated from local curvature information which is contained in the history of the last n_{hist} displacements and gradient differences

$$\Delta\mathbf{R}^i = \mathbf{R}^i - \mathbf{R}^{i-1}, \tag{2.8}$$

$$\Delta\mathbf{g}^i = \nabla E\left(\mathbf{R}^i\right) - \nabla E\left(\mathbf{R}^{i-1}\right), \tag{2.9}$$

where $i = 1 \dots n_{\text{hist}}$. Probably the most prominent example for such a quasi-Newton method is the Broyden-Fletcher-Goldfarb-Shanno method, outlined in the following section.

The Broyden-Fletcher-Goldfarb-Shanno Method

The Broyden-Fletcher-Goldfarb-Shanno[133–136] (BFGS) method is one of the most used[141] and one of the most efficient minimization methods.[132,137,145,146] In this approach, the iterations look very similar to those of Newton's method:

$$\mathbf{R}^{i+1} = \mathbf{R}^i - \alpha_i B_i^{-1} \nabla E\left(\mathbf{R}^i\right). \tag{2.10}$$

Here, B_i^{-1} is an approximation to the inverse Hessian $H_{\mathbf{R}^i}^{-1}$ of Eq. 2.7. One of the fundamental ideas of the BFGS method is to build the approximation successively by means of additive updates U_i, instead of computing it from scratch at each iteration:

$$B_i^{-1} = B_{i-1}^{-1} + U_i. \tag{2.11}$$

Formulas for the updates U_i can be obtained by requiring the B_i^{-1} to be positive definite and symmetric

$$B_i^{-1} = \left(B_i^{-1}\right)^T, \tag{2.12}$$

and to fulfill the secant equation (cf. Eq. 2.6)[141]

$$B_i^{-1}\left(\nabla E\left(\mathbf{R}^{i+1}\right) - \nabla E\left(\mathbf{R}^i\right)\right) = \mathbf{R}^{i+1} - \mathbf{R}^i. \tag{2.13}$$

It is this secant equation, that provides finite difference curvature information for the approximate Hessian. Both the symmetry condition, as well as the secant equation are fulfilled by the exact Hessian, which makes it natural to require them for the approximate Hessians, too. Additionally, to obtain a unique formula for U_i, one must require that B_{i+1}^{-1} is, in some sense, close to B_i^{-1}. If the distance between two matrices is defined by a weighted Frobenius norm $||\cdot||_{F,w}$, the following BFGS update formula can be obtained by minimizing $||B^{-1} - B_i^{-1}||_{F,w}$ with respect to B^{-1} and subject to the constraints given by Eq. 2.12 and Eq. 2.13:[133–136,141]

$$U_i = \left(\rho_i \Delta\mathbf{R}^i (\Delta\mathbf{g}^i)^T - Id\right) B_i^{-1} \rho_i \Delta\mathbf{g}^i (\Delta\mathbf{R}^i)^T - \rho_i \Delta\mathbf{R}^i (\Delta\mathbf{g}^i)^T B_i^{-1} + \rho_i \Delta\mathbf{R}^i (\Delta\mathbf{R}^i)^T, \tag{2.14}$$

where $\rho_i := 1/\left((\Delta\mathbf{g}^i)^T \Delta\mathbf{R}^i\right)$. The $\Delta\mathbf{R}^i$ and $\Delta\mathbf{g}^i$ have been defined in Eq. 2.8 and Eq. 2.9. The initial approximation to the Hessian matrix can be set to a multiple of the identity matrix.

The BFGS algorithm needs to store and manipulate the complete Hessian approximation, which can become prohibitive for systems with a huge number of atoms and a PES that is computationally inexpensive to evaluate. To circumvent this problem, there exists L-BFGS, a limited memory variant of the BFGS algorithm.[137,138,141] The basic idea of L-BFGS is to store

a history consisting of only the last n_{hist} displacements $\mathbf{\Delta R}^i$ and gradient differences $\mathbf{\Delta g}^i$. From this history the approximation to the current Hessian can be obtained by a recursive procedure. L-BFGS is equivalent to BFGS if the number of the current iteration is smaller than n_{hist}, assuming consistent choices of the initial approximate Hessians.[141]

Although the size of the approximate BFGS Hessian is not significant with respect to the computing time for the problems considered in this thesis, the L-BFGS and not the BFGS algorithm has been used. This was done, because there exists an excellent implementation that is available from Nocedal's website,[147] which already was included in the BigDFT code.[148,149]

For non-convex functions, it is important for the step size α_i of Eq. 2.10 to be determined by a line search based on conditions like the Wolfe or strong Wolfe conditions, which guarantee updates such that each approximate Hessian is positive definite.[141] However, the experience was made that this line search is problematic when there is a relatively large amount of computational noise on the forces.[64] Instabilities and inefficiencies of BFGS applied to noisy PESs or problems related to the line search have been reported by others, too.[132,140,150,151] Recently, the fast inertial relaxation engine has become popular in the field of ab-initio structure prediction. Although not as efficient as the BFGS method, it was reported to be robust with respect to computational noise.[132]

2.1.3 Fast Inertial Relaxation Engine

Typically, steepest descent based minimizers are very stable, but not very efficient. On the other hand, the quasi-Newton methods that have been available so far are efficient, but, at times, are instable if energies and forces are inconsistent with each other. Recently, Bitzek et al. developed the fast inertial relaxation engine (FIRE), that combines efficiency with robustness.[132] FIRE is not a quasi-Newton Method, but belongs to the class of damped MD optimizers.[152,153] The idea is appealingly simple. On a hill, at an initial position with non-zero slope, a (frictionless) ball with zero initial velocity is released to roll downwards. If, at some time, the ball starts to roll upwards, it is stopped and immediately released again. This procedure can be repeated until the ball has arrived at a local minimum. In general the ball's velocity does not point into the direction of steepest descent. However, if it could actively steer to a direction steeper than the direction given by the current velocity, the convergence to a minimum might be more efficient. The equations of motions for such a steered motion is given by[132]

$$\frac{d}{dt}\mathbf{v}(t) = \frac{\mathbf{F}(t)}{m} - \gamma(t)|\mathbf{v}(t)|\left(\hat{\mathbf{v}}(t) - \hat{\mathbf{F}}(t)\right). \tag{2.15}$$

Here, $\mathbf{v}(t) = \frac{d}{dt}\mathbf{R}(t)$ is the velocity and $\mathbf{F}(t)$ are the forces acting on all the particles. Hats indicate normalized vectors.

By adding a small number of additional lines of code, any MD algorithm can be modified to obtain the FIRE algorithm. In Fig. 2.3 a pseudocode for the fire algorithm is given. The param-

FIRE

1. initialize variables $it \leftarrow 0$; $\Delta t \leftarrow \Delta t_0$; $\alpha \leftarrow \alpha_{\text{start}}$; $\mathbf{v} \leftarrow \mathbf{0}$;
2. `repeat`
3. $\quad it \leftarrow it + 1$;
4. $\quad$ Using any MD integrator, calculate $\mathbf{R}$, $\mathbf{F} = -\nabla E(\mathbf{R})$ and $\mathbf{v}$;
5. $\quad$ compute power $P = \mathbf{F} \cdot \mathbf{v}$;
6. $\quad$ modify velocities $\mathbf{v} \leftarrow (1-\alpha)\mathbf{v} + \alpha \hat{\mathbf{F}}|\mathbf{v}|$;
7. $\quad$ `if` $P \leq 0$ `then`
8. $\quad\quad$ freeze system $\mathbf{v} \leftarrow \mathbf{0}$;
9. $\quad\quad$ decrease time step $\Delta t \leftarrow \Delta t f_{\text{dec}}$;
10. $\quad\quad$ reset damping factor $\alpha \leftarrow \alpha_{\text{start}}$;
11. $\quad\quad$ $it_{\text{cut}} \leftarrow it$;
12. $\quad$ `else if` $P > 0$ and $it - it_{\text{cut}} > N_{\text{min}}$ `then`
13. $\quad\quad$ increase time step $\Delta t \leftarrow \min(\Delta t f_{\text{inc}}, \Delta t_{\text{max}})$;
14. $\quad\quad$ decrease damping factor $\alpha \leftarrow \alpha f_\alpha$;
15. $\quad$ `endif`
16. `until convergence.`

Figure 2.3: Pseudocode of the FIRE algorithm as described in Ref. [132]. The vectors **R**, **v** and **F** are elements of $\mathbb{R}^{3N}$. Values for parameters that usually do not need any adjustments are $N_{\text{min}} = 5$, $f_{\text{inc}} = 1.1$, $f_{\text{dec}} = 0.5$, $\alpha_{\text{start}} = 0.1$, $f_\alpha = 0.99$. The maximum time step Δt_{max} is system dependent. Bitzek et al. suggest to set $\Delta t_{\text{max}} \approx 10 \Delta t_{\text{MD}}$, where Δt_{MD} is a typical time step used in the MD simulation part.[132]

eter α used in the pseudocode is defined by $\alpha := \gamma \Delta t$. It can be seen from the discretization of Eq. 2.15 given in Fig. 2.3 that the damping term in Eq. 2.15 can be looked at as a mixing of the velocities with the forces (steepest descent directions).

It is intuitively clear that MD-type optimizers should be noise tolerant, since the inertia introduced by the simulation of Newton's equations of motion smoothens the erratic "bumps" introduced by noisy forces. In fact, similar damped MD based optimizers have been known before the advent of FIRE.[152,153]

Often, FIRE is more efficient than other advanced, but more complicated algorithms, like for example the conjugate gradient method.[132] However, already in the original publication it was demonstrated by means of benchmarks that FIRE is inferior to the (L-)BFGS method.[132] This was the motivation of the development of the noise tolerant quasi-Newton approach, which is described in Sec. 2.3.

2.2 Mode Following Methods for Optimization of Saddle Points

A first order saddle point is a stationary point with one negative eigenvalue. That is, it is a maximum along the direction of the Hessian eigenvector corresponding to the lowest eigenvalue and a minimum in all other directions. For convenience, the Hessian eigenvector corresponding to the lowest eigenvalue will henceforth be denoted as minimum mode. With this in mind, the basic working principle of mode following methods[49,55,154] is obvious. Starting at a given point on the PES, saddle points are found by iteratively translating the point along a modified force $\mathbf{F}^{\dagger} = \mathbf{F} - 2(\mathbf{F}\cdot\hat{\mathbf{d}}_{\text{min}})\hat{\mathbf{d}}_{\text{min}}$. Here, $\hat{\mathbf{d}}_{\text{min}}$ is the minimum mode. In this modified force, the force component that is parallel to the minimum mode is simply flipped. Therefore, the energy is maximized along the minimum mode and minimized in all other directions and it is clear that this procedure should converged to a saddle point, at least, if it is started close enough to a saddle point.

2.2.1 Dimer Method

The dimer method is a mode following method that was first described in a 1999 publication by Henkelman and Jónsson.[55] In the subsequent decade it was significantly improved by contributions of Olsen et al.[56], Heyden et al.[57] and Kästner and Sherwood.[58] This section begins with an outline on how the original version of the dimer method implements the above mentioned mode following idea and it concludes with a review of the significant improvements that have been made by the just mentioned authors.

The central object in the dimer method is a system consisting of two images $\mathbf{R}_1$ and $\mathbf{R}_2$. The images are two close-by points on the PES with corresponding energies E_1, E_2 and forces $\mathbf{F}_1$, $\mathbf{F}_2$. They are separated from their midpoint $\mathbf{R}_0$ by a distance ΔR, such that $\mathbf{R}_1 := \mathbf{R}_0 + \Delta R\hat{\mathbf{N}}$ and $\mathbf{R}_2 := \mathbf{R}_0 - \Delta R\hat{\mathbf{N}}$, where the normalized vector $\hat{\mathbf{N}}$ defines the dimer axis. The dimer energy is defined as the sum of the image energies $E = E_1 + E_2$. In the original version of the dimer method, the energies and forces at the midpoint are not explicitly calculated, but they are interpolated by $\mathbf{F}_0 = (\mathbf{F}_1 + \mathbf{F}_2)/2$ and

$$E_0 = \frac{E}{2} + \frac{\Delta R}{4}(\mathbf{F}_1 - \mathbf{F}_2)\cdot\hat{\mathbf{N}}, \tag{2.16}$$

which is derived from a central difference approximation to the curvature

$$C \approx \frac{(\mathbf{F}_1 - \mathbf{F}_2)\hat{\mathbf{N}}}{2\Delta R} \tag{2.17}$$

$$\approx \frac{E - 2E_0}{(\Delta R)^2}. \tag{2.18}$$

In order to find the direction of minimal curvature, the dimer energy E is minimized by rotating the dimer around its midpoint. As is apparent from Eq. 2.18, this is equivalent to minimizing

the curvature function C. A rotational force $\mathbf{F}^{\perp}$, which is acting on $\mathbf{R}_1$, is defined by

$$\mathbf{F}^{\perp} := \mathbf{F}_1^{\perp} - \mathbf{F}_2^{\perp}, \tag{2.19}$$

where $\mathbf{F}_i^{\perp} := \mathbf{F}_i - (\mathbf{F}_i \cdot \hat{\mathbf{N}})\hat{\mathbf{N}}$ for $i = 1,2$. The dimer energy is minimal, if the rotational force vanishes. In the dimer method, E is not minimized by iteratively displacing $\mathbf{R}_1$ along the rotational force (and rescaling the distance to the dimer center), but it is done in a sophisticated two-step procedure that consists of fixing the plane of the rotation and determining the amount of the rotation by a modified Newton method.

The plane of the rotation is defined to be spanned by the dimer axis $\hat{\mathbf{N}}$ and a normalized vector $\hat{\mathbf{\Theta}}$ that is orthogonal to the dimer axis. In a steepest descent scheme, $\hat{\mathbf{\Theta}}$ is chosen to be parallel to the rotational force $\mathbf{F}^{\perp}$. In fact, the steepest descent approach is not very efficient and, already in the original publication, $\hat{\mathbf{\Theta}}$ is determined within a conjugate gradient scheme.[55]

After having fixed the plane of rotation, the angle of the rotation needs to be determined. To do so, Henkelman quadratically expanded the dimer energy in terms of the normal modes of the potential energy within the plane of rotation and obtained an expression for the dimer energy as a function of the rotational angle θ:[55]

$$E(\theta) = 2E_0 + c_1 \cos(2(\theta - \theta_0)) + c_2, \tag{2.20}$$

where c_1 and c_2 are some unknown constants depending on the curvature with respect to the normal modes in the rotational plane. Computing the ratio of the first $F = -E'(\theta) = A\sin[2(\theta - \theta_0)]$ and second derivative $F' = -E''(\theta) = 2A\cos[2(\theta - \theta_0)]$ of the dimer energy, allows the determination of the zero of F without knowing the constant A. This results in the following $\theta_{\min}$ that minimizes the dimer energy within the rotational plane[55–57]

$$\theta_{\min} = \theta_0 = -\frac{1}{2}\arctan\left(\frac{F}{F'}\right) - \frac{\delta\varphi}{2}, \tag{2.21}$$

where in the simulation F and F' are evaluated by means of gradient calculations at a dimer configuration that is rotated by $\delta\varphi$. Specific formulas for F and F' can be found in Refs. [56, 57]. Within one step, the dimer is not necessarily converged to the minimum mode. Therefore, several rotational steps might have to be repeated until the dimer is converged.[58]

So far, only the (approximate) minimum mode has been determined. Now, the dimer needs to be moved into the direction of a saddle point. This is done by translating the dimer along a modified force $\mathbf{F}^{\dagger}$

$$\mathbf{F}^{\dagger} = \begin{cases} -(\mathbf{F}_0 \cdot \hat{\mathbf{N}})\hat{\mathbf{N}} & \text{if } C_N > 0 \\ \mathbf{F}_0 - 2(\mathbf{F}_0 \cdot \hat{\mathbf{N}})\hat{\mathbf{N}} & \text{if } C_N < 0. \end{cases} \tag{2.22}$$

The step size for the translation is given by the absolute value of the inverse curvature along the direction defined by $\mathbf{F}^{\dagger}$, which is computed by finite differences, for which additional force evaluations are necessary (the forces at both ends of the final rotated dimer have not yet been

computed). Although the step length is determined in a Newton fashion, this is not true for the step direction, which, in the original dimer implementation, is found by means of a conjugate gradient scheme. To conclude, at least six force evaluations have to be performed for each dimer cycle.[57]

In later versions, the efficiency of the dimer method has been improved by computing Eq. 2.18 by means of a forward finite difference, which saves two force evaluations per rotation step.[56,57] Furthermore, Heyden et al.[57] realized that a more accurate parametrization of $E(\theta)$ is possible by means of a short Fourier series, which results in a more accurate estimate for θ_{min} and Kästner introduced the usage of the L-BFGS algorithm for the dimer rotation and translation.[58] It was this latest and improved version of the dimer method that has been used for the benchmarks described in Sec. 2.3.7.

A Stabilized Dimer Method

In Chap. 3 the eigenvector following exploration (EFE) method[2,35] is used for exploring the energy landscape and sampling the connectivity between minima and saddle points. The idea is to escape from a minimum to many different saddle points by following different eigenmodes of the Hessian. As is argued in Appx. C, the only local minimum of Eq. 2.18 corresponds to the minimum mode of the Hessian. All other Hessian eigenvectors represent saddle points, with the consequence that many gradient based optimizers are unstable at these points. As soon as the search mode deviates from the exact eigenmode, which inevitably happens during an actual simulation due to the finite optimization step size, there is a strong tendency to converge to the lowest mode. This is a significant problem if the systematic following of many different modes is desired.

To overcome this problem, Mohr suggested[62] to use the direct inversion of iterative subspace (DIIS)[155] method for the rotation of the dimer. In this scheme, the $-\alpha\mathbf{F}_i^{\perp}$ of the dimer method are used as the residual vectors, required for the construction of the DIIS matrix. Here, α is some positive constant, the integer i denotes the iteration number and $\mathbf{F}^{\perp}$ is given by Eq. 2.19. The DIIS scheme has the tendency to converge to close-by stationary points[16] and, therefore, will not converge to the lowest mode, but rather to the mode that has the largest overlap with the previous dimer orientation. In this sense, the dimer method is stabilized and can be used to systematically follow different modes out of a local minimum.

Fig. 2.4 compares the DIIS stabilized dimer rotation to a steepest descent based variant. It can be clearly seen that the stabilized dimer method reliably attaches to one mode, whereas the steepest descent based approach exhibits a sudden switch to a lower mode (around iteration 17).

Finally, also the stabilized dimer method should converge to a first order saddle point. Therefore, at some point the dimer rotation must be allowed to converge to the minimum mode. In the implementation that was used for the EFE method, this was done as soon as the second

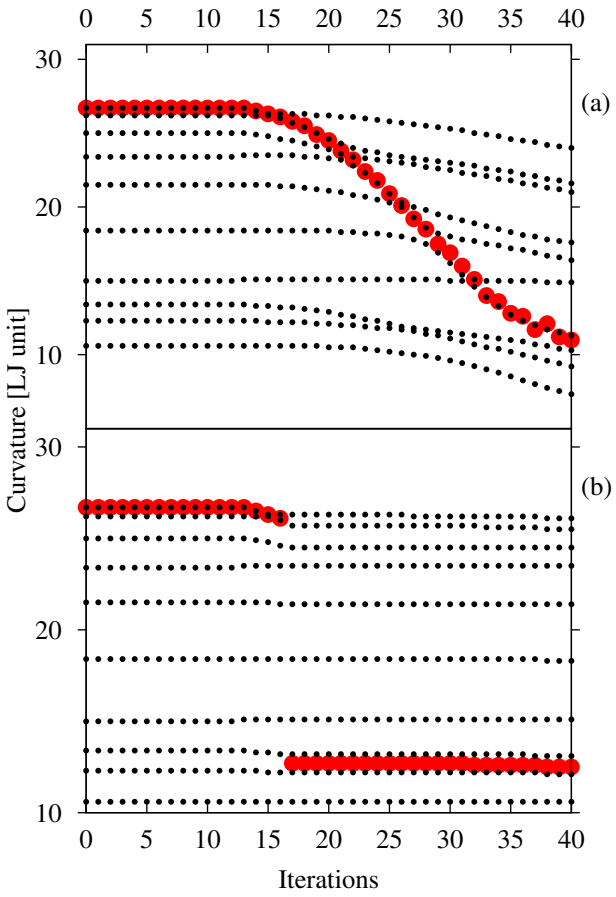

Figure 2.4: A visualization of how using DIIS for the dimer rotations helps to stay on a given mode. Panel (a) visualizes data obtained by using DIIS, panel (b) shows data obtained by using steepest descent. The small black dots are the ten lowest eigenvalues of the Hessian at each step of a trajectory starting at a local minimum, whereas the large red dots are the curvature along the search direction. The DIIS procedure in panel (a) stays in general on the mode that has the largest overlap with the dimer direction, and thus stays on the initial mode for quite a long time. In contrast to this, the steepest descent procedure in panel (b) becomes unstable as soon as the 9th and 10th mode cross and switches to a low curvature mode, as a consequence. Reprinted with permission from Ref. [62]. Copyright 2014 by the American Institute of Physics.

derivative of the energy with respect to the iteration number became negative. At this point, the lowest mode is determined by using the Lanczos method,[156] as presented in Ref. [56]. Of course, any other method that does not converge to a saddle point of the curvature function (Eq. 2.18) could be used.

2.2.2 Bar-Saddle

As already explained in the introduction to this chapter, finding saddle points is much more difficult than finding minima. As a consequence, saddle finding techniques tend to suffer from a comparatively high failure rate. An attempt to increase the reliability of saddle finding methods was undertaken by Amsler which resulted in the bar-saddle method.[62] At the time of its development, it was the most reliable and most efficient method available to the author. Consequently, the initial implementation of the minima hopping guided path search (MHGPS) approach, presented in Sec. 3.5, was based on the Bar-Saddle method. Later, the usage of bar-saddle within MHGPS was superseded by the stabilized quasi-Newton saddle (SQNS) search method presented in Sec. 2.3.6. The description of bar-saddle in this chapter follows the outline given in Ref. [62].

The fundamental idea of bar-saddle is a solid, horizontal bar placed at a point that is higher in energy than a close-by saddle point. Such a bar would roll towards the saddle point, if its point of contact with the PES is kept at the bar's center. This is in contrast to a ball that, at the presence of friction, would roll towards a local minimum.

Formally, the bar-saddle method is similar to the dimer method,[55] it is, however, based on a different usage paradigm. By means of a simple linear interpolation in Cartesian coordinates, an initial path is generated, from which the highest energy configuration is taken. The highest energy configuration along this path is found by means of Brent's algorithm,[157] and is used as starting configuration for the bar-saddle method. Later, the author of this thesis replaced the linear interpolation by the freezing string method with Cartesian interpolation to avoid atomic clashes.[158] It was this latter version that was used in Sec. 3.5.

Just like the dimer of the dimer method, the bar is defined by two points $\mathbf{R}_A$ and $\mathbf{R}_B$ on the PES. Both points are assumed to be in close vicinity to each other and their separation – the length of the bar – is denoted as $h = ||\mathbf{R}_A - \mathbf{R}_B||$. Starting at a suitable high-energy configuration, the bar is displaced such that the maximum energy along the bar is at its center and such that the energy at the bar's center is minimized with respect to all other directions that are perpendicular to the bar. To do so, the forces at both bar ends are decomposed into a component parallel to the bar $\mathbf{F}_\mathrm{i}^\parallel = \left(\mathbf{F}_\mathrm{i} \cdot \hat{\mathbf{h}}\right)\hat{\mathbf{h}}$ and a component perpendicular to the bar $\mathbf{F}_\mathrm{i}^\perp = \mathbf{F}_\mathrm{i} - \mathbf{F}_\mathrm{i}^\parallel$. Here $\hat{\mathbf{h}} := (\mathbf{R}_A - \mathbf{R}_B)/h$ is the unit vector in bar direction and $i = A, B$. At every iteration, the energies and forces at both bar ends must be evaluated. Both, the energy and

the force in bar direction at the bar center are computed by means of a cubic interpolation

$$E_{h/2} = \frac{1}{8}(4E_A + 4E_B + (f_B - f_A)h), \tag{2.23}$$

$$\mathbf{F}^{\parallel}_{h/2} = \frac{6E_A - 6E_B - (f_A + f_B)h}{4h}\hat{\mathbf{h}}, \tag{2.24}$$

where $f_i = \mathbf{F}_\mathrm{i} \cdot \hat{\mathbf{h}}$. This interpolated parallel force is used to flip the corresponding force component. That is, the translational forces acting on both bar ends are defined as $\mathbf{F}^{\mathrm{Trans}}_A = \mathbf{F}^{\mathrm{Trans}}_B = \mathbf{F}^{\perp}_{\mathrm{h}/2} - 2\mathbf{F}^{\parallel}_{h/2}$.

The rotational forces acting on both bar ends are defined by $\mathbf{F}^{\mathrm{Rot}}_A = \frac{1}{2}(\mathbf{F}^{\perp}_A - \mathbf{F}^{\perp}_B)$ and $\mathbf{F}^{\mathrm{Rot}}_B = \frac{1}{2}(\mathbf{F}^{\perp}_B - \mathbf{F}^{\perp}_A)$. Like in the dimer method, these rotational forces approximately align the bar with the minimum mode.

In contrast to the bar-saddle approach, the dimer method estimates both the parallel and perpendicular components of the translational force by the arithmetic mean of the forces at the dimer endpoints. The force responsible for the rotation only acts on one endpoint in case of the dimer method and the rotation is implemented by using the parametrization of a circle in a 2-dimensional plane (cf. Eq. 2.20). In principle, the dimer is rotated in a single step by an angle that is estimated by means of a modified one-dimensional Newton method (cf. Eq. 2.21).[55,62]

Finally, in a steepest descent approach, the bar ends are displaced by $\alpha\mathbf{F}^{\mathrm{Trans}}_i + \beta\mathbf{F}^{\mathrm{Rot}}_i$, where $\alpha > 0$ and $\beta > 0$ define the translational and rotational step sizes. Because a finite sized rotational step size is used, it is necessary to rescale the bar length after each iteration such that $|\mathbf{R}^{\mathrm{New}}_B - \mathbf{R}^{\mathrm{New}}_A| \stackrel{!}{=} h$. To increase the efficiency, the step sizes α and β can be adjusted by means of a simple energy or gradient feedback. In the actual implementation, the steepest descent moves only have been used within the first few iterations, after which the BFGS method has been used for the translational part of the displacement. It should be noted that the bar-saddle method will also converge to a saddle point if the initial configuration is lower in energy than the saddle point, albeit at the cost of efficiency.

2.3 Stabilized Quasi-Newton Optimization[†]

If the PES can be computed with an accuracy on the order of the machine precision, the above mentioned L-BFGS algorithm usually works very well. In practice, however, computing the PES at this high precision is not possible for physically accurate but computationally demanding levels of theory, like, for example, density functional theory (DFT). At DFT level, this is due to the finitely spaced integration grids and self consistency cycles that have to be stopped at small, but non-vanishing thresholds. Therefore, optimization algorithms that are

[†]The research presented in this section has been published in B. Schaefer, S. A. Ghasemi, S. Roy, and S. Goedecker, "Stabilized Quasi-Newton Optimization of Noisy Potential Energy Surfaces", The Journal of Chemical Physics **142**, 034112 (2015). Reprinted with permission. Copyright 2015 by the American Institute of Physics.

used at these accurate levels of theory must not only be computationally efficient but also tolerant of noise in forces and energies.

In this chapter a technique is presented that allows the extraction of curvature information from noisy PESs. It is explained how to use this technique for the construction of a stabilized quasi-Newton minimizer (SQNM) and a stabilized quasi-Newton saddle (SQNS) finding method. With the help of benchmarks, both optimizers are demonstrated to be robust and efficient. The comparison of SQNM to L-BFGS and FIRE and the comparison of SQNS to the improved dimer method[55,58] mentioned in Sec. 2.2.1 reveals that SQNM and SQNS are superior to their existing alternatives.

2.3.1 Significant Subspace in Noisy Optimization Problems

In noisy optimization problems, the noisy components of the gradients can lead to displacement components that correspond to erratic movements on the PES. Consequently, curvature information that comes from the subspace spanned by these displacement components must not be used for the construction of an approximate Hessian. In contrast to this, the non-noisy gradient components promote locally systematic net-movements, which do not tend to cancel each other. In this sense, the displacement components that correspond to these well defined net-movements span a significant subspace from which meaningful curvature information can be extracted and used for building an approximate Hessian.

The situation is depicted in Fig. 2.5 where the red solid vectors represent the history of normalized displacements and the blue dashed vectors constitute a basis of the significant subspace. All the red solid vectors in Fig. 2.5a point into similar directions. Therefore, curvature information should only be extracted from a one-dimensional subspace, as, for example, is given by the blue dashed vector. Displacement components perpendicular to this blue dashed vector come from the noise in the gradients. In contrast to Fig. 2.5a, Fig. 2.5b shows a displacement that points into a considerably different direction than all the other displacements. For this reason, significant curvature information can be extracted from the full two-dimensional space.

To define the significant subspace more rigorously, first the set of normalized displacements is introduced (cf. Eq. 2.8)

$$\widehat{\mathbf{\Delta R}}^i := \frac{\mathbf{\Delta R}^i}{|\mathbf{\Delta R}^i|}, \tag{2.25}$$

where $i = 1 \dots n_{\text{hist}}$. With $\sum_k |\boldsymbol{\omega}_k|^2 = 1$, linear combinations $\mathbf{w}$ of the normalized displacements are defined as:

$$\mathbf{w} := \sum_{k=1}^{n_{\text{hist}}} \boldsymbol{\omega}_k \widehat{\mathbf{\Delta R}}^k, \tag{2.26}$$

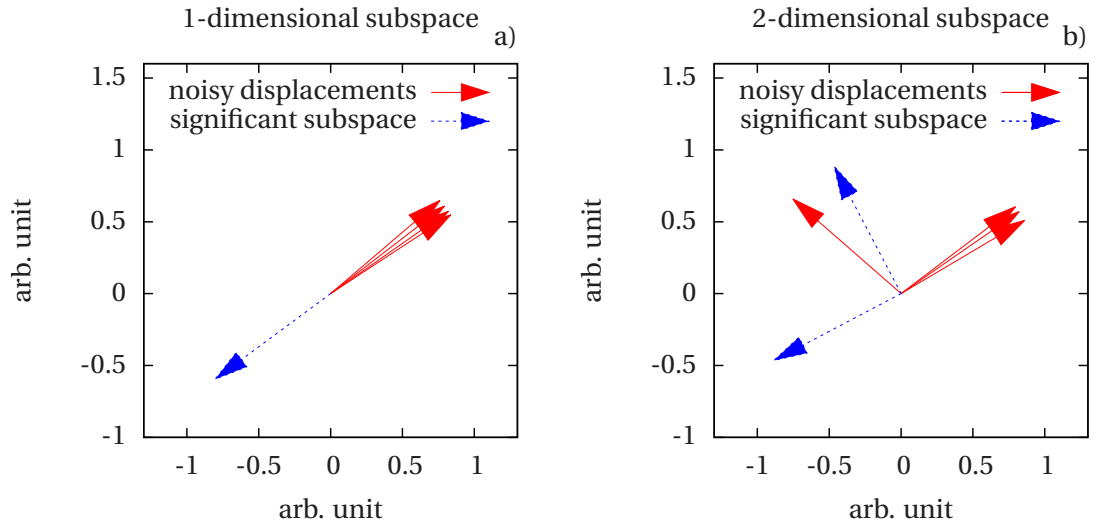

Figure 2.5: Illustrated are significant subspaces spanned by the displacements in a model atomic coordinate space. Only from the significant subspace, it is meaningful to extract curvature information. The red solid arrows simulate displacements made under the influence of noisy forces. The blue dashed arrows show significant subspaces from which it is meaningful to extract curvature information. Panel (a) shows a case in which the significant subspace is only one-dimensional. Panel (b) shows an example in which curvature information can be extracted from the full 2-dimensional space. The significant subspaces that are shown here were computed by using the method outlined in Sec. 2.3.1. Reprinted with permission from Ref. [64]. Copyright 2015 by the American Institute of Physics.

Furthermore, a real symmetric overlap matrix S is defined as

$$S_{kl} := \widehat{\mathbf{\Delta R}}^k \cdot \widehat{\mathbf{\Delta R}}^l. \tag{2.27}$$

It can be seen from,

$$\mathbf{w} \cdot \mathbf{w} = \boldsymbol{\omega}^T S \boldsymbol{\omega}, \tag{2.28}$$

that $|\mathbf{w}|$ is made stationary by coefficient vectors $\boldsymbol{\omega}^i$ that are eigenvectors of the overlap matrix. In particular, the longest and shortest vectors that can be generated by linear combinations with normalized coefficient vectors $\boldsymbol{\omega}$ correspond to those eigenvectors of the overlap matrix that have the largest and smallest eigenvalues. As motivated above, the shortest linear combinations of the normalized displacements correspond to noise.

From now on, let the $\boldsymbol{\omega}^i$ be eigenvectors of (S_{kl}) and let λ_i be the corresponding eigenvalues. With the orthonormal

$$\widetilde{\widehat{\mathbf{\Delta R}}}^i := \frac{1}{\sqrt{\lambda_i}} \sum_{k=1}^{n_{\text{hist}}} \omega_k^i \widehat{\mathbf{\Delta R}}^k, \tag{2.29}$$

the *significant subspace* $\mathfrak{S}$ is finally defined as

$$\mathfrak{S} := \mathrm{span}\left(\left\{\widetilde{\mathbf{\Delta R}}^i \;\middle|\; \lambda_i / \max_j\{\lambda_j\} > \epsilon\right\}\right), \tag{2.30}$$

where $0 \leq \epsilon \leq 1$. In all applications presented in this chapter, $\epsilon = 10^{-4}$ has proven to work well. Henceforth, the dimension of $\mathfrak{S}$ will be denoted as n_{dim}. By construction it is guaranteed that $n_{\mathrm{dim}} \leq 3N$. It should be noted that at each iteration of the optimization algorithms that are to be introduced below, the significant subspace and its dimension n_{dim} can change. The history length n_{hist} usually lies between 5 and 20.

The above procedure is analogous to Löwdins canonical orthogonalization,[159–161] which is used in the electronic structure community to remove linear dependencies from chemical basis sets.

2.3.2 Obtaining Curvature Information on the Significant Subspace

A projection $\tilde{H}$ of the Hessian H onto $\mathfrak{S}$ is defined as

$$\begin{aligned}\tilde{H} &:= PHP \\ &= \sum_{ij} H_{ij}\, \widetilde{\mathbf{\Delta R}}^i \left(\widetilde{\mathbf{\Delta R}}^j\right)^T,\end{aligned} \tag{2.31}$$

where for all $\widetilde{\mathbf{\Delta R}}^i \in \mathfrak{S}$

$$P := \sum_{i=1}^{n_{\mathrm{dim}}} \widetilde{\mathbf{\Delta R}}^i \left(\widetilde{\mathbf{\Delta R}}^i\right)^T \tag{2.32}$$

and

$$H_{ij} := \left(\widetilde{\mathbf{\Delta R}}^i\right)^T H\, \widetilde{\mathbf{\Delta R}}^j. \tag{2.33}$$

Using Eq. 2.6 and Eq. 2.9 and defining

$$\widetilde{\mathbf{\Delta g}}^i := \frac{1}{\sqrt{\lambda_i}} \sum_{k=1}^{n_{\mathrm{hist}}} \frac{\boldsymbol{\omega}_k^i}{|\Delta \mathbf{R}^k|} \Delta \mathbf{g}^k, \tag{2.34}$$

where $i = 1 \ldots n_{\mathrm{dim}}$, an approximation for each matrix element H_{ij} is obtained:

$$H_{ij} \approx \widetilde{\mathbf{\Delta g}}^i \cdot \widetilde{\mathbf{\Delta R}}^j. \tag{2.35}$$

In practice, H_{ij} is explicitly symmetrized to avoid asymmetries introduced by anharmonic effects:

$$H_{ij} \approx \frac{1}{2}\left(\tilde{\mathbf{\Delta g}}^i \cdot \widetilde{\mathbf{\Delta R}}^j + \tilde{\mathbf{\Delta g}}^j \cdot \widetilde{\mathbf{\Delta R}}^i\right). \tag{2.36}$$

Because the projection P is the identity operator on $\mathfrak{S}$, the curvature $C(\widehat{\mathbf{d}})$ on the PES along a normalized $\widehat{\mathbf{d}} \in \mathfrak{S}$ is given by

$$C(\widehat{\mathbf{d}}) = \widehat{\mathbf{d}}^T \tilde{H} \widehat{\mathbf{d}}. \tag{2.37}$$

Given the normalized eigenvectors $\mathbf{v}^i$ and corresponding eigenvalues κ_i of the $n_{\text{dim}} \times n_{\text{dim}}$ Matrix $\left(H_{ij}\right)$, the normalized eigenvectors $\tilde{\mathbf{v}}^i \in \mathfrak{S}$ of $\tilde{H}$ with eigenvalues κ_i can be written as

$$\tilde{\mathbf{v}}^i = \sum_{k=1}^{n_{\text{dim}}} \mathbf{v}^i_k \widetilde{\mathbf{\Delta R}}^k, \tag{2.38}$$

where $\mathbf{v}^i_k$ is the k-th element of $\mathbf{v}^i$. As can be seen from Eq. 2.37, the κ_i give the curvatures of the PES along the directions $\tilde{\mathbf{v}}^i$.

2.3.3 Using Curvature Information on the Significant Subspace for Preconditioning ∇E

The gradient ∇E can be decomposed into a component lying in $\mathfrak{S}$ and a component lying in its orthogonal complement:

$$\nabla E = \nabla E_{\mathfrak{S}} + \nabla E_{\perp}, \tag{2.39}$$

where $\nabla E_{\mathfrak{S}} := P' \nabla E$, $\nabla E_{\perp} := (I - P')\nabla E$ and $P' := \sum_i \tilde{\mathbf{v}}^i \left(\tilde{\mathbf{v}}^i\right)^T$. In this section it is motivated how the κ_i can be used to precondition $\nabla E_{\mathfrak{S}}$. Furthermore, it is explained how $\nabla E_{\perp}$ can be scaled appropriately with the help of a feedback that is based on the angle between two consecutive gradients.

Let the Hessian H at the current point of the PES be non-singular and let ν_i and $\mathbf{V}^i$ be its eigenvalues and normalized eigenvectors. In Newton's Method (Eq. 2.7), the gradients are conditioned by the inverse Hessian. For the significant subspace component $\nabla E_{\mathfrak{S}}$ it follows:

$$H^{-1} \nabla E_{\mathfrak{S}} = \sum_{i=1}^{3N} \sum_{j=1}^{n_{\text{dim}}} \left[\left(\frac{\nabla E \cdot \tilde{\mathbf{v}}^j}{\nu_i} \right) \left(\tilde{\mathbf{v}}^j \cdot \mathbf{V}^i \right) \mathbf{V}^i \right] \tag{2.40}$$

As outlined in the previous section, the curvature κ_j along $\tilde{\mathbf{v}}^j$ is known. Therefore, at a first thought, Eq. 2.40 suggests to simply replace ν_i with κ_j where $i = 1 \ldots 3N$ and $j = 1 \ldots n_{\text{dim}}$. Indeed, if the optimization was restricted to the subspace $\mathfrak{S}$ this choice would be appropriate.

However, with respect to the complete domain of the PES, there is a risk to underestimate the curvature ν_i if the overlap $O_{ij} := \tilde{\mathbf{v}}^j \cdot \mathbf{V}^i$ is non-vanishing.

In particular, if O_{ij} is far from being negligible, underestimating the curvature ν_i can be particularly problematic because coordinate changes in the direction of $\mathbf{V}^i$ might be too large. This can render convergence difficult to obtain in practice.

Therefore, ν_i in Eq. 2.40 is replaced by

$$\kappa'_j := \sqrt{\kappa_j^2 + r_j^2}, \tag{2.41}$$

where r_j is chosen in analogy to the residue of Weinstein's Criterion[162,163] as

$$r_j := \left| H\tilde{\mathbf{v}}^j - \left((\tilde{\mathbf{v}}^j)^T H \tilde{\mathbf{v}}^j \right) \tilde{\mathbf{v}}^j \right|. \tag{2.42}$$

Using Eqn. 2.34, Eqn. 2.37 and Eqn. 2.38, this residue can be approximated by

$$r_j \approx \left| \sum_{k=1}^{n_{\text{dim}}} \left[\mathbf{v}_k^j \, \tilde{\mathbf{\Delta g}}^k \right] - \kappa_j \tilde{\mathbf{v}}^j \right|. \tag{2.43}$$

With this choice for κ'_j, the preconditioned gradient $\nabla E^{\text{P}}_{\mathfrak{S}}$ is finally given by:

$$\nabla E^{\text{P}}_{\mathfrak{S}} := \sum_{j=1}^{n_{\text{dim}}} \left(\frac{\nabla E \cdot \tilde{\mathbf{v}}^j}{\kappa'_j} \right) \tilde{\mathbf{v}}^j. \tag{2.44}$$

Clearly, the residue r_j can only alleviate the problem of curvature underestimation, but it does not rigorously guarantee that every single ν_i is estimated appropriately. However, in practice this choice works very well. The reason for this can be seen from Fig. 2.6. In Fig. 2.6a, a histogram of the quality and safety measure $q_{ij} := \sqrt{\kappa_j^2 + r_j^2} - \nu_i$ is shown. If $q_{ij} < 0$, the curvature ν_i is underestimated, if $q_{ij} \approx 0$ the curvature ν_i is well estimated and finally, if $q_{ij} > 0$, the curvature is overestimated. Overestimation leads to too small step sizes and, therefore, to a more stable algorithm, albeit at the cost of a performance loss. Critical underestimation of the curvature ($q_{ij} \ll 0$) is rare. Fig. 2.6b shows the averages of the overlap O_{ij} in the corresponding bins. If, on average, $\tilde{\mathbf{v}}^j$ has a large overlap with $\mathbf{V}^i$, the curvature along $\mathbf{V}^i$ is estimated accurately (histogram in Fig. 2.6a peaks at $q_{ij} \approx 0$).

What remains to be discussed is how the gradient component $\nabla E_\perp$ should be scaled. By construction, $\nabla E_\perp$ lies in the subspace for which no curvature information is available. Therefore, this gradient component is treated by a simple steepest descent approach that adjusts the step size $\alpha > 0$ at each iteration. For the minimizer that is outlined in Sec. 2.3.5, the adjustment is based on the angle between the complete gradient ∇E and the preconditioned gradient ∇E^{P}. If the cosine of this intermediate angle is larger than 0.2, α is increased by a factor of 1.1, otherwise α is decreased by a factor of 0.85. For the saddle search algorithm the feedback is slightly different and will be explained in Sec. 2.3.6. The exact numbers for the scaling factors

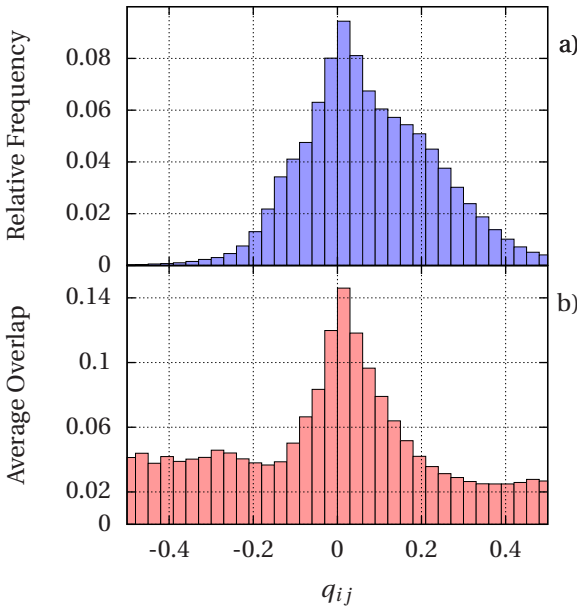

Figure 2.6: Panel a) is a histogram of $q_{ij} := \sqrt{\kappa_j^2 + r_j^2} - \nu_i$ for $i = 1 \ldots 3N$ and $j = 1 \ldots n_{\text{dim}}$. q_{ij} is a measure for the quality of the estimation of the eigenvalue ν_i of the exact Hessian. Panel b) shows the bin-averaged overlap O_{ij}. The frequency of severe curvature underestimation drops quickly in the region $q_{ij} < 0$. The histogram in panel a) peaks in the region of good estimation ($q_{ij} \approx 0$) which coincidences with the region of large overlap O_{ij}, shown in panel b). The data for this figure come from 100 minimizations of a Si_{20} system described by the Lenosky-Silicon[127,128] force field. Reprinted with permission from Ref. [64]. Copyright 2015 by the American Institute of Physics.

were determined heuristically. The only constraints are that the scaling factors must increase the step size if the complete gradient and the preconditioned gradient point into similar directions and decrease the step size otherwise. Based on experience, the above choices offer a good efficiency.

In conclusion, the total preconditioned gradient ∇E^{P} is given by

$$\nabla E^{\text{P}} := \nabla E^{\text{P}}_{\mathfrak{S}} + \alpha \nabla E_{\perp} \tag{2.45}$$

Sec. 2.3.4 explains how this preconditioned gradient can be further improved for biomolecules.

The preconditioned subspace gradient $\nabla E^{\text{P}}_{\mathfrak{S}}$ was obtained under the assumption of a quadratic PES. However, if the gradients at the current iteration are large, this assumption is probably not satisfied. Displacing along $\nabla E^{\text{P}}_{\mathfrak{S}}$ in these cases can reduce the stability of the optimization. Hence, if the $|\nabla E|$ exceeds a certain threshold, it can be useful to set the dimension of $\mathfrak{S}$ to zero for a certain number of iterations. This means that $\nabla E_{\perp} = \nabla E$ and, therefore, $\nabla E^{\text{P}} = \alpha \nabla E$. In that case, α is also adjusted according to the above described gradient feedback. As this fallback to steepest descent is intended as a last and final fallback, it should have the ability

to deal with arbitrarily large forces. Therefore, it is also checked that $\alpha\nabla E$ does not displace some atom by more than a user-defined trust radius. However, based on experience, this fallback is not necessary in most cases. Indeed, all the benchmarks presented in Sec. 2.3.7 were performed without this fallback.

2.3.4 Additional Efficiency for Biomolecules

Many large molecules like biomolecules or polymers are floppy systems in which the largest and smallest curvatures can be very different from each other. Steepest descent optimizers are very inefficient for these ill-conditioned systems, because the high curvature directions force to use step sizes that are far too small for an efficient optimization in the directions of small curvatures. Put more formally, the optimization is inefficient for those systems, because the condition number, which is the fraction of largest and smallest curvature, is large.[164] For biomolecules, the high-curvature directions usually correspond to bond stretchings, that is, movements along inter-atomic displacement vectors of bonded atoms. For the current purpose two atoms are regarded to be bonded if their inter-atomic distance is smaller than or equal to 1.2 times the sum of their covalent radii. For $i = 1 \dots N$, let $\mathbf{r}^i \in \mathbb{R}^3$ be the coordinate vector of the i-th atom. For a system with n_{bond} bonds a bond vector $\mathbf{b}^m \in \mathbb{R}^{3N}$, $m = 1 \dots n_{\text{bond}}$ is defined for each bond

$$\mathbf{b}^m := \begin{pmatrix} \overset{\sim m,1}{\mathbf{b}} \\ \overset{\sim m,2}{\mathbf{b}} \\ \vdots \\ \overset{\sim m,N}{\mathbf{b}} \end{pmatrix}. \tag{2.46}$$

Here the $\overset{\sim m,k}{\mathbf{b}} \in \mathbb{R}^3$, $k = 1 \dots N$ are defined as

$$\overset{\sim m,i}{\mathbf{b}} := -\overset{\sim m,j}{\mathbf{b}} := \begin{cases} \mathbf{r}^j - \mathbf{r}^i, & \text{if atoms i and j are} \\ & \text{bonded by the m-th bond,} \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{2.47}$$

The $\mathbf{b}^m$ are sparse vectors with six non-zero elements.

The total gradient ∇E is separated into its bond-stretching components ∇E_{str} and all the remaining components ∇E_{r}:

$$\nabla E = \nabla E_{\text{str}} + \nabla E_{\text{r}}. \tag{2.48}$$

Let $c_m \in \mathbb{R}$ be coefficients that allow the bond-stretching components to be expanded in terms

of the bond vectors

$$\nabla E_{\mathrm{str}} := \sum_{m=1}^{n_{\mathrm{bond}}} c_m \mathbf{b}^m. \tag{2.49}$$

Using definition Eq. 2.49, left-multiplying Eq. 2.48 with a bond vector $\mathbf{b}^n$ and requiring the ∇E_{r} to be orthogonal to all the bond vectors, the following linear system of equations is obtained, which determines the coefficients c_m and, with it, the bond stretching gradient defined in Eq. 2.49:

$$\mathbf{b}^n \cdot \nabla E = \sum_m c_m \mathbf{b}^n \cdot \mathbf{b}^m. \tag{2.50}$$

For the optimization of a biomolecule, the bond-stretching components are minimized in a simple steepest descent fashion. The atoms are displaced by $-\alpha_{\mathrm{s}} \nabla E_{\mathrm{str}}$. The bond-stretching step size α_{s} is a positive number, which is adjusted in each iteration of the optimization by simply counting the number of projections $\mathbf{b}^m \cdot \nabla E$ that have not changed signs since the last iteration. If more than two thirds of the signs of the projections have remained unchanged, the bond-stretching step size α_{s} is increased by 10 percent. Otherwise, α_{s} is decreased by a factor of 1/1.1. The non-bond-stretching gradients ∇E_{r} are preconditioned using the stabilized quasi-Newton approach presented in Secs. 2.3.1 to 2.3.3. It is important to note that in Secs. 2.3.1 to 2.3.3 all ∇E have to be replaced by ∇E_{r} when using this biomolecule preconditioner. In particular, this is also true for the gradient feedbacks that are described in Secs. 2.3.3 and 2.3.6.

2.3.5 Finding Minima – The SQNM method

The pseudocode in Fig. 2.7 demonstrates how the above presented techniques can be assembled into an efficient and stabilized quasi-Newton minimizer (SQNM). The pseudocode contains 4 parameters explicitly. α_{start} and $\alpha_{\mathrm{s,start}}$ are initial step sizes that scale $\nabla E_{\perp}$ and ∇E_{str}, respectively. m is the maximum length of the history list from which the significant subspace $\mathfrak{S}$ is constructed. E_{thresh} is an energy-threshold that is used to determine whether a minimization step is accepted or not. It should be adapted to the noise level of the energies and forces. The history list is discarded if the energy increases, because an increase in energy is an indication for inaccurate curvature information. In this case, the dimension of the significant subspace is considered to be zero. Furthermore, line 17 implicitly contains the parameter ϵ, which is described in Sec. 2.3.1. The optimization is considered to be converged if the norm of the gradient is smaller than a certain threshold value. Of course, other force criteria, like for example using the maximum force component instead of the force norm, are possible.

SQNM

1. $\alpha \leftarrow \alpha_{\text{start}}$; $\alpha_s \leftarrow \alpha_{\text{s,start}}$;
2. accepted $\leftarrow$ true;
3. $k \leftarrow 1$;
4. Initialize $\mathbf{R}_k$ with coordinates;
5. $E_k \leftarrow E(\mathbf{R}_k)$;
6. `repeat`
7. `if` optimizing biomolecule `then`
8. `if` accepted `then`
9. Compute ∇E_{str} for $\mathbf{R}_k$, as outlined in Sec. 2.3.4;
10. Adjust α_s based on the feedback described in Sec. 2.3.4;
11. $\mathbf{g}_k \leftarrow \nabla E(\mathbf{R}_k) - \nabla E_{\text{str}}$;
12. $\mathbf{R}_k \leftarrow \mathbf{R}_k - \alpha_s \nabla E_{\text{str}}$;
13. `end if`
14. `else`
15. $\mathbf{g}_k \leftarrow \nabla E(\mathbf{R}_k)$;
16. `end if`
17. Based on the $\{\mathbf{g}_j, \mathbf{R}_j\}_{j \le k}$ in the history list, compute the preconditioned gradient ∇E^{P} as outlined in Secs. 2.3.1 to 2.3.4;
18. $\mathbf{R}_{k+1} \leftarrow \mathbf{R}_k - \nabla E^{\text{P}}$;
19. `if` $E(\mathbf{R}_{k+1}) > E_k + E_{\text{thresh}}$ `and` $\alpha > \alpha_{\text{start}}/10$ `then`
20. accepted $\leftarrow$ false;
21. Remove $\{\mathbf{g}_j, \mathbf{R}_j\}_{j<k}$ from the history list;
22. $\alpha \leftarrow \alpha/2$;
23. `else`
24. accepted $\leftarrow$ true;
25. $E_{k+1} \leftarrow E(\mathbf{R}_{k+1})$;
26. Adjust α based on the gradient feedback described in Sec. 2.3.3;
27. `if` $k > m$ `then`
28. Remove $\mathbf{R}_{k-m}$ and $\mathbf{g}_{k-m}$ from storage;
29. `end if`
30. $k \leftarrow k+1$;
31. `end if`
32. `until convergence.`

Figure 2.7: Pseudocode of the SQNM algorithm. Reprinted with permission from Ref. [64]. Copyright 2015 by the American Institute of Physics.

SQNS

1. $\alpha' \leftarrow \alpha'_{\text{start}}$; $\alpha'_s \leftarrow \alpha'_{\text{s,start}}$;
2. $l \leftarrow 1$;
3. Initialize $\mathbf{R}_l$ with coordinates;
4. `repeat`
5. `if` recompute minimum mode `then`
6. Use algorithm of Sec. 2.3.5 and obtain a normalized minimum $\widehat{\mathbf{d}}_{\text{min}}$ of $C(\mathbf{d})$ at $\mathbf{R}_l$, use the previously computed minimum mode as input;
7. `end if`
8. `if` optimizing biomolecule `then`
9. Compute ∇E_{str} for $\mathbf{R}_l$, as outlined in Sec. 2.3.4;
10. Adjust α'_s based on the feedback described in Sec. 2.3.4;
11. $\mathbf{s} \leftarrow \alpha'_s \nabla E_{\text{str}}$;
12. $\mathbf{g}_l \leftarrow \nabla E(\mathbf{R}_l) - \nabla E_{\text{str}}$;
13. $\mathbf{R}_l \leftarrow \mathbf{R}_l - \mathbf{s} + 2\left(\mathbf{s} \cdot \widehat{\mathbf{d}}_{\text{min}}\right)\widehat{\mathbf{d}}_{\text{min}}$;
14. Check for trust radius condition as described in Sec. 2.3.6. Rescale, if needed;
15. `else`
16. $\mathbf{g}_l \leftarrow \nabla E(\mathbf{R}_l)$;
17. `end if`
18. Based on the $\{\mathbf{g}_j, \mathbf{R}_j\}_{j \le l}$ in the history list, compute the preconditioned gradient ∇E^{P} as outlined in Secs. 2.3.1 to 2.3.4;
19. $\mathbf{R}_{l+1} \leftarrow \mathbf{R}_l - \nabla E^{\text{P}} + 2\left(\nabla E^{\text{P}} \cdot \widehat{\mathbf{d}}_{\text{min}}\right)\widehat{\mathbf{d}}_{\text{min}}$;
20. Check for trust radius condition and for fragmentation as described in Sec. 2.3.6. Rescale and fix fragmentation, if needed;
21. Adjust α' based on the gradient feedback described in Sec. 2.3.6;
22. `if` $l > m'$ `then`
23. Remove $\mathbf{R}_{l-m'}$ and $\mathbf{g}_{l-m'}$ from the history list;
24. `end if`
25. $l \leftarrow l + 1$;
26. `until convergence.`

Figure 2.8: Pseudocode of the SQNS algorithm. Reprinted with permission from Ref. [64]. Copyright 2015 by the American Institute of Physics.

2.3.6 Finding Saddle Points – The SQNS Method

In this section a stabilized quasi-Newton saddle finding method (SQNS) is described, which is based on the same principles as the minimizer in the previous section. SQNS belongs to the class of the minimum mode following methods.[49,55,154]

In the SQNS scheme, the minimum mode of the Hessian is found by minimizing the curvature function $c : \mathbb{R}^{3N} \mapsto \mathbb{R}$

$$\begin{aligned} C(\mathbf{d}) &= \frac{\mathbf{d}^T H \mathbf{d}}{\mathbf{d}^T \mathbf{d}} \\ &\approx \frac{\Delta\mathbf{g} \cdot \Delta\mathbf{R}}{h^2}, \end{aligned} \tag{2.51}$$

where along with $h \ll 1$ the following definitions have been used: $\Delta\mathbf{R} := h\frac{\mathbf{d}}{|\mathbf{d}|}$ and $\Delta\mathbf{g} := \nabla E(\mathbf{R}+\Delta\mathbf{R}) - \nabla E(\mathbf{R})$. The vector $\mathbf{R}$ is the position at which the Hessian H is evaluated. For the minimization of $C(\mathbf{d})$, the algorithm described in Sec. 2.3.5 is used, where the energy as objective function is replaced by $C(\mathbf{d})$. In the pseudocode given in Fig. 2.8, the here discussed minimization is done at line 6. Under the constraint of normalization, the gradient $\nabla C(\mathbf{d})|_{|\mathbf{d}|=1}$ is given by

$$\begin{aligned} \nabla C(\mathbf{d})|_{|\mathbf{d}|=1} &= 2\left(H\mathbf{d} - C(\mathbf{d})\mathbf{d}\right) \\ &\approx 2\left(\frac{\Delta\mathbf{g}}{h} - \left(\frac{\Delta\mathbf{g}\cdot\Delta\mathbf{R}}{h^3}\right)\Delta\mathbf{R}\right). \end{aligned} \tag{2.52}$$

Blindly using the biomolecule preconditioner of Sec. 2.3.4 for the minimization of $C(\mathbf{d})$ would mean that the gradient of Eq. 2.52 was projected on the bond vectors of $\mathbf{d}$. The bond vector as defined in Sec. 2.3.4 has no meaning for $\mathbf{d}$. Therefore, Eq. 2.52 instead is projected onto the bond vectors of $\mathbf{R}+\Delta\mathbf{R}$.

At a stationary point, systems with free boundary conditions have six vanishing eigenvalues. The respective eigenvectors correspond to overall translations and rotations.[3] Instead of directly using Eq. 2.52 for the minimization of the curvature of those systems, it is advantageous to remove the translations and rotations from $\Delta\mathbf{R}$ and $\nabla C(\mathbf{d})|_{|\mathbf{d}|=1}$ in Eq. 2.52.[3,165,166] Different methods for this purpose are discussed in Appx. A.

The convergence criterion for the minimization of $C(\mathbf{d})$ has a large influence on the total number of energy and force evaluations needed to obtain overall convergence of the saddle point search. It, therefore, must be chosen carefully. The minimum mode is usually not computed at every iteration, but only if one of the following conditions is fulfilled:

1. at the first iteration of the optimization
2. if the integrated length of the optimization path connecting the current point in coordinate space and the point at which the minimum mode has been calculated previously exceeds a given threshold value r_{recomp}

3. if the curvature along the minimum mode is positive and the curvature has not been recomputed for at least n_{recomp} iterations

4. if the curvature along the minimum mode is positive and the norm of the gradient falls below the convergence criterion

5. at convergence (optional)

In the pseudocode, these conditions are checked in line 5. Among these conditions, condition no. 2 is, with respect to the performance, the most important one. The number of energy and gradient evaluations needed for converging to a saddle point can be strongly reduced if a good value for r_{recomp} is chosen. Condition 3 and 4 can be omitted for most cases. However, for some cases they can offer a slight reduction in the number of energy and gradient evaluations. For example for the alanine dipeptide system used in Sec. 2.3.7, these two conditions offered a performance gain of almost 10%. Although possible, n_{recomp} is usually not tuned, but typically $n_{\text{recomp}} = 10$ is used. Condition 5 can be made optional in an actual implementation. This condition is used if very accurate directions of the minimum mode at the saddle point are needed. In this case, this last minimum mode computation can also be performed at a tighter convergence criterion. Additional energy and gradient computations can be saved by using the previously computed minimum mode as the starting mode for a new curvature minimization.

As stated above, a saddle point is found by maximizing along the minimum mode and minimizing in all other directions. This is done by inverting the preconditioned gradient component that is parallel to the minimum mode. This is done at line 19 of the pseudocode in Fig. 2.8. For the case of biomolecules, the component of the bond-stretching gradient that is parallel to the minimum mode is also inverted (line 13). As already mentioned in Sec. 2.3.3, the feedback that adjusts the step size of $\nabla E_{\perp}$ is slightly different in case of the saddle finding method. Let $\widehat{\mathbf{d}}_{\text{min}}$ be the normalized direction of the minimum mode. Then, in contrast to minimizations, the step size that is used to scale $\nabla E_{\perp}$ is not based on the angle between the complete ∇E and ∇E^{P}, but only on the angle between $\nabla E - \left(\nabla E \cdot \widehat{\mathbf{d}}_{\text{min}}\right)\widehat{\mathbf{d}}_{\text{min}}$ and $\nabla E^{\text{P}} - \left(\nabla E^{\text{P}} \cdot \widehat{\mathbf{d}}_{\text{min}}\right)\widehat{\mathbf{d}}_{\text{min}}$. These are the components that are responsible for the minimization in directions that are not the minimum mode direction. Otherwise, the gradient feedback is absolutely identical to that described in Sec. 2.3.3.

A saddle point can be higher in energy than the configuration at which the optimization is started at. Therefore, in contrast to a minimization, it is not reasonable to discard the history, if the energy increases. As a replacement for this safeguard, a simple trust radius approach is used in which no atom must be moved by more than a predefined trust radius r_{trust}. A displacement exceeding this trust radius is simply rescaled. If the curvature is positive and the norm of the gradient is below the convergence criterion, the displacements that do not come from bond-stretchings are rescaled as well. The displacement is rescaled such that the displacement of the atom that moved furthest is finally given by r_{trust}. This avoids arbitrarily small steps close to minima.

On very rare occasions, it was observed for some cluster systems that over the course of several iterations a few atoms sometimes detach from the main cluster. To avoid this problem, the main fragment is identified and all neighboring fragments are moved towards the nearest atom of the main fragment.

In Fig. 2.8, the pseudocode for SQNS is given. It contains three parameters explicitly. α'_{start} and $\alpha'_{\text{s,start}}$ are initial step sizes that scale $\nabla E_{\perp}$ and ∇E_{str}, respectively. m' is the maximum length of the history list from which the significant subspace is constructed. The path-length threshold r_{thresh} that determines the recomputation frequency of the minimum mode is implicitly contained in line 5. Lines 14 and 21 imply the trust radius r_{trust}. Besides all the parameters that are needed for the minimizer of Sec. 2.3.5, line 6 additionally implies the finite difference step size h that is used to compute the curvature and its gradient. Line 18 implicitly contains the parameter ϵ, which is described in Sec. 2.3.1

The optimization is considered to be converged if the curvature along the minimum mode is negative and if the norm of the gradient is smaller than a certain threshold.

2.3.7 Benchmarks and Comparisons

Minimizers

In this section, the performance of the new SQNM method is compared to the performance of the FIRE and L-BFGS minimizers. The conjugate gradient method is not included in this benchmark, because FIRE has previously been shown to be significantly more efficient than CG.[132] Both FIRE and L-BFGS belong to the best optimizers in their class. With regard to the required number of energy and force evaluation, L-BFGS is one of the best minimizers available for the optimization of atomic systems. With respect to noise tolerance, the same is true for FIRE. Although more efficient than FIRE, L-BFGS tends to fail if there are inconsistent forces and energies due to computational noise.[132] Such inconsistencies are unavoidable in electronic structure calculations like for example DFT.

For Si_{20} clusters and the alanine dipeptide biomolecule, benchmarks were performed both at DFT and force field level. For L-BFGS the reference implementation of Nocedal[137,138] has been used, which is available from his website.[147] The author is not aware of any reference implementation of FIRE. However, FIRE is straightforward to implement and thus an own code was used. For the benchmarks of the minimizers at DFT level, all codes were coupled to the BigDFT electronic structure code.[148,149] For the benchmarks at force field level, the Assisted Model Building with Energy Refinement (AMBER) force field in the ff99SB variant as implemented in AMBER Tools[126] and the Lenosky Silicon force field were used.[127,128]

For alanine dipeptide and Si_{20}, test sets were generated by running MD simulations at force field level. At force field level each test set contains 1000 structures that were taken from the MD trajectories. Subsets containing 100 of these force field structures were used as benchmark

Table 2.1: Benchmark results for the minimizers. DFT test sets contain 100, force field test sets contain 1000 distinct structures. SQNM runs labeled with '(Bio)' indicate the usage of the preconditioner for biomolecules described in Sec. 2.3.4.

				To Same Minimum				To Arbitrary Minimum			
System	Level of Theory	Method	n_f[1]	N^2	$\langle n_{ef}\rangle^3$	$\langle r\rangle^4$	$\langle n_{woi}\rangle^5$	N^2	$\langle n_{ef}\rangle^3$	$\langle r\rangle^4$	$\langle n_{woi}\rangle^5$
Alanine Dipeptide	DFT	FIRE	0	93	454	14.01	7602	100	458	14.14	7662
↓	↓	L-BFGS	2	93	185	23.41	3876	98	188	24.02	3941
↓	↓	SQNM (Bio)	0	93	198	14.10	3711	100	207	14.29	3858
↓	Force Field	FIRE	0	954	414	12.21	–	1000	418	12.35	–
↓	↓	L-BFGS	1	954	156	19.69	–	999	158	20.39	–
↓	↓	SQNM (Bio)	0	954	188	12.38	–	1000	192	12.57	–
↓	↓	SQNM	0	954	356	12.27	–	1000	363	12.49	–
Si_{20}	DFT	FIRE	0	46	139	18.83	2458	100	143	19.52	2513
↓	↓	L-BFGS	30	46	73	27.19	1677	70	74	31.26	1714
↓	↓	SQNM	0	46	83	16.00	1740	100	86	16.50	1784
↓	Force Field	FIRE	0	486	147	13.26	–	1000	163	15.32	–
↓	↓	L-BFGS	0	486	57	25.49	–	1000	65	30.44	–
↓	↓	SQNM	0	486	72	10.82	–	1000	81	11.93	–

[1] Number of failed optimizations.

[2] Number of runs over which the averages are taken.

[3] Average number of energy and force calls (only successful runs).

[4] Average integrated path length of the optimization trajectory in units of Bohr.

[5] Average number of wave function optimization iterations.

systems at DFT level. For each method, the parameters were tuned at force field level for a subset of 100 configurations. Identical parameters were used both at force field and DFT level. The Si_{20} system was considered to be converged as soon as the norm of the force fell below 1.0×10^{-4} Hartree/Bohr. Even if far away from a stationary point, relatively small forces can arise in alanine dipeptide. Therefore, a much tighter convergence criterion of 1.0×10^{-5} Hartree/Bohr had to be chosen for this system.

Table 2.1 gives the benchmark results. In addition to the average number of energy and force calls $\langle n_{\mathrm{ef}} \rangle$, also the average integrated path length of the optimization path $\langle r \rangle$ is given. $\langle r \rangle$ is computed by summing all the distances between structures for which consecutive energy and force evaluations were performed.

There is no guarantee that minimizations that are started at the same configuration will converge to the identical minimum. Therefore, Table 2.1 gives averages for both, the subset of runs that all converged to identical minima and averages over all runs, regardless of whether the final minima were identical, or not. Identical configurations were identified by using the recently developed *s*-overlap fingerprints,[167] which are briefly recapitulated in Appx. B.

In all benchmarks, FIRE is clearly inferior to L-BFGS and SQNM. With respect to the average number of energy and force evaluations, the L-BFGS method is slightly more efficient than the new SQNM minimizer. However, $\langle r \rangle$ of L-BFGS is 1.6 to 2.6 times larger than the corresponding values of the SQNM method. On average, this means that L-BFGS displaces the atoms more violently than SQNM. In DFT calculations, the wave function of the previous optimization step can be used as input wave function for the current iteration. Roughly speaking, the less the positions of the atoms have changed, the better this input guess usually is. Therefore, less wave function optimizations are needed for convergence. To quantify this, the average number of wave function optimization iterations $\langle n_{\mathrm{woi}} \rangle$ needed for a minimization of the PES is given in Table 2.1. As a consequence of the smaller displacements in the SQNM method, the L-BFGS and the SQNM method roughly need the same number of wave function optimizations for converging to a minimum of the PES.

The L-BFGS minimizer needed less energy and force evaluations at force field level than at DFT level. It was verified that this is not due to the noise at DFT level, but a consequence of the different natures of both PESs. The force field PES is not a noiseless variant of the DFT PES, but a rather inaccurate approximation to it. In particular this means that the frequencies of the force field are different from the frequencies of the DFT energy surface and, therefore, the same is true for the condition numbers. For this reason, one cannot expect to obtain the same number of energy and force evaluations at force field and DFT level.

The L-BFGS minimizer proved to be unreliable at DFT level. For example, 30% of all Si_{20} minimizations failed to converge. In contrast to this, all SQNM runs successfully converged to a minimum. The convergence failures of the L-BFGS method are in general caused by failures of the line minimizations in the final part of the optimization where a large fraction of the forces consists of noise.

Saddle Finding Methods

The SQNS method was compared to an improved version[56–58] of the dimer method[55] as described in Ref. [58] and as implemented in the EON code.[168] A short review of the (improved) dimer method is also given in Sec. 2.2.1.

The same force fields as for the minimization benchmarks were used. For the DFT calculations, SQNS was coupled to the BigDFT code. The EON code offers an interface to the Vienna Ab-initio Simulation Package (VASP),[169–173] which consequently was used.

The same test sets as for the minimizer benchmarks were used. In particular, this means that the starting configurations are not close to a saddle point and, therefore, these test sets are comparatively difficult tests for saddle finding methods. Again, parameters were only tuned for a subset of 100 configurations at force field level. With exception to the finite difference step size that is needed to calculate the curvature and its gradient, the identical parameters were used at force field and DFT level. Because of noise, the finite difference step size must be chosen larger at DFT level. The same force norm convergence criteria as for the minimization benchmarks were used. In all SQNS optimizations the minimum mode was recalculated at convergence (condition 5 of Sec. 2.3.6).

The test results are given in Table 2.2. In contrast to the minimization benchmarks, no averages for the number of wave function optimization iterations are given, because the two saddle finding methods were coupled to two different electronic structure codes. Therefore, the number of wave function optimizations is not comparable.

In particular, in case of the Si_{20} system, both methods converged only seldom to the same saddle points and, therefore, the statistical significance of the corresponding numbers given in Table 2.2 is limited. However, averages over large sets could be made in the case of convergence to an arbitrary saddle point.

In the considered cases, the dimer method needed between 1.4 and 7.6 times more energy and force evaluations than the new SQNS method. In particular, for alanine dipeptide, the SQNS approach was far superior to the dimer method. Due to its inefficiency, it was impossible to obtain a significant number of saddle points for alanine dipeptide at DFT level when using the dimer method. For this reason, only benchmark results for the SQNS method are given for alanine dipeptide at DFT level.

2.3.8 Conclusion

Optimizations of atomic structures belong to the most important routine tasks in fields like computational physics, chemistry, or biology. Although the energies and forces given by computationally demanding methods like DFT are physically accurate, they are contaminated by noise. The computational noise comes from the underlying integration grids and from self-consistency cycles that are stopped at non-vanishing thresholds. The availability of

Table 2.2: Benchmark results for the saddle finding methods. DFT test sets contain 100, force field test sets contain 1000 distinct structures. SQNS runs labeled with '(Bio)' indicate the usage of the preconditioner for biomolecules as described in Sec. 2.3.4.

System	Level of Theory	Method	n_f[1]	To Same Saddle Point		To Arbitrary Saddle Point	
				N[2]	$\langle n_{ef} \rangle$[3]	N[2]	$\langle n_{ef} \rangle$[3]
Alanine Dipeptide	DFT	SQNS (Bio)	0	–	–	100	510
↓	Force Field	DIMER	0	87	1324	1000	3146
	↓	SQNS (Bio)	0	87	309	1000	415
↓	↓	SQNS	0	87	632	1000	757
Si_{20}	DFT	DIMER	0	8	234	100	444
↓	↓	SQNS	0	8	140	100	237
	Force Field	DIMER	0	20	264	1000	622
↓	↓	SQNS	0	20	189	1000	368

[1] Number of failed optimizations.

[2] Number of runs over which the averages are taken.

[3] Average number of energy and force calls (only successful runs).

optimization methods that are not only efficient but also noise-tolerant is, therefore, of great importance. In this section, a technique to extract significant curvature information from noisy PESs was presented. The technique was used to create a minimization algorithm (SQNM) and a saddle finding algorithm (SQNS). SQNM and SQNS were demonstrated to be superior to existing efficient and well established methods.

Until now, the SQNM and the SQNS optimizers have been used over a period of several months. During this time, they have performed tens of thousands of optimizations without failure at the DFT level. Because of their robustness with respect to computational noise and due to their efficiency, they have replaced the default optimizers that have previously been used in minima hopping[18,110] and minima hopping guided path search[62] runs.

Implementations of the minimizer and the saddle search method are made available via the BigDFT electronic structure package. The code is distributed under the GNU General Public License and can be downloaded free of charge from the BigDFT website.[65]

3 Finding Reaction Pathways*

The exploration of potential energy landscapes requires two important aspects to be considered. On the one hand, the geometries of low-energy states, including the global minimum, are of large interest. On the other hand, processes like protein folding, catalysis, chemical reactions in solutions and surfaces or the formation of stable phases in solids often force the reacting systems to undergo rarely occurring and complex transformations between long-lived states. Actively stabilizing or destabilizing long-lived states by inhibiting or promoting reaction pathways responsible for certain events allows the synthesization of new materials or substances with specifically tailored properties.[174–176] Unfortunately, the sole knowledge of the global minimum and a collection of local minima as obtained by global optimization methods provides not enough information for the target-oriented design of processes that influence reaction pathways in a desired way. Instead, an accurate knowledge of the atomistic details of reaction pathways is needed. For this reason, in addition to local minima, also transition states and the information which minima are connected by which transition states are of great importance. As soon as this data is available, various methods like the master equation approach, the discrete path formulation of discrete path sampling or kinetic monte carlo allow to compute dynamic properties.[3,38–40] Using graph-theoretic methods it is possible to extract reaction pathways from databases containing the just mentioned data. Since pathways with energetically high barriers have a vanishingly small contribution to properties like rate constants, it is important not to investigate just any pathways but to sample preferably those that have low overall barriers.

When exclusively used, methods like the nudged elastic band method or the splined saddle method,[177,178] are not suitable to systematically search for low-barrier reaction pathways, since they only find some minimum-energy pathway, but not necessarily a low-barrier pathway. Furthermore these methods often fail to find a connection between distant minima.[179] The problem of finding reaction pathways is similar to the problem of global minimization. Just

*Parts of this chapter have been published in B. Schaefer, S. Mohr, M. Amsler, and S. Goedecker, "Minima Hopping Guided Path Search: An Efficient Method for Finding Complex Chemical Reaction Pathways", The Journal of Chemical Physics **140**, 214102 (2014). Reprinted with permission. Copyright 2014 by the American Institute of Physics.

like a local minimization method is, in general, not sufficient for finding a global minimum, methods like nudged elastic band are not sufficient for the systematic search for low-barrier pathways. Nevertheless, reaction pathways can be partitioned into a sequence of stationary point crossings and, therefore, those "local" pathway or other local saddle point finding methods constitute important building blocks of algorithms that search for (energetically low) reaction pathways. In contrast, the present chapter focuses on the introduction of a novel reaction pathway finding scheme, named the minima hopping guided path search (MHGPS) algorithm (see Sec. 3.5). Based on the MH global optimization method, it allows the efficient sampling of low-barrier reaction pathways on complex potential energy surfaces (PESs), which is demonstrated in Sec. 3.5.1 with the help of benchmarks and comparisons to alternative approaches. Using MHGPS, the energy landscapes of LJ_{75} and LJ_{102} were mapped out. Despite numerous published investigations of the Lennard-Jones clusters, MHGPS was able to find many pathways that are significantly lower in energy and shorter with respect to the integrated path length and number of intermediate transition states than previously known pathways for LJ_{75}.[2] For LJ_{102} a third, previously unknown and energetically low-lying funnel was located. At the bottom of this funnel, a new structural motif is located. The pathways that were found between both lowest minima of LJ_{102} are also significantly shorter in terms of the number of intermediate transition states and in terms of the integrated path length when compared to previously presented pathways.[63]

The MHGPS method has been published in Ref. [62] and this chapter is a reorganized, partially rewritten, and extended version of this publication. Before introducing the MHGPS method itself, the chapter starts with shortly mentioning the basic terminology and then proceeds with a review of relevant alternative reaction pathway finding methods: First, the eigenvector following exploration (EFE) approach[2,35] is outlined, which is followed by transition path sampling (TPS)[26,27,180–184] and the very recently published stochastic surface walking based reaction sampling (SSW-RS).[185] EFE was explicitly used in the benchmark section for the comparison with MHGPS and, therefore, is recapitulated in this chapter. It was possible to straight forwardly compare TSP to MHGPS, since TSP previously had been applied to one of the test systems used in the benchmarks. SSW-RS is included in the review, because it uses ideas very similar to those exploited in the MHGPS approach. However, it should be noted that MHGPS was published several months before the SSW-RS method.

3.1 Terminology and Basic Methods

In this chapter, the usual definition of a transition state being a first order saddle point of the PES is used.[3] Steepest descent paths connect transition states to two stationary points, which in most cases are local minima. The "connectivity" of a PES is defined by these steepest descent path, though often a connectivity defined by more advanced energy minimization schemes, like quasi-Newton methods, is possible.[3] The terminology of Wales[3,38,39] is adapted and sequences of minima and transition states connected by steepest descent paths are denoted as "discrete paths". A collection of local minima, transition states and the information

which transition states connect which minima is called a "stationary point database".[3,38,39]

Building stationary point databases requires the identification or distinction of atomic configurations with or from each other. For this purpose the recently developed fingerprints which are based on the eigenvalues of a *s*-orbital overlap matrix were used.[167] A brief recapitulation of these fingerprints can be found in Appx. B. For the calculation of the fingerprints, the half of the LJ equilibrium distance ($2^{1/6}\frac{\sigma}{2}$) was used as the covalent radius of the LJ atoms. Two conformers were considered to be identical if their energy difference were smaller than $10^{-5}\epsilon$ and their fingerprint distance less than 2×10^{-4}.

Extracting from a stationary point database all lowest-barrier paths with the least number of intermediate transition states between two given minima poses a problem that is closely related to the so called shortest-widest[186] path problem. This can be solved by applying a modified Dijkstra's algorithm[187] twice.[186] In the first step the modified Dijkstra's algorithm searches for all paths that connect both minima with the lowest possible energy barrier $E_{\text{barr;lowest}}$. The stationary point database then is truncated by removing all transition states with energies higher than $E_{\text{barr;lowest}}$. Next, Dijkstra's algorithm passes through the truncated database and searches for the path with the smallest possible number of intermediate transition states.

It must be emphasized that, similar to most of the commonly used global optimization algorithms, the methods described below do not rigorously guarantee the finding of an optimal solution of the reaction path problem. That is, all presented structures and lowest-barrier pathways should be denoted as "putative lowest structures" or "putative lowest-barrier pathways". However, for convenience, sometimes the word "putative" is omitted.

3.2 The Eigenvector Following Exploration Method

In Ref. [35] Doye et al. presented an algorithm that allows mode following techniques to be used for the exploration of the potential energy landscape. Based on the method used for the transition state search, the potential energy landscape exploration method of Ref. [35] will be denoted as the eigenvector following exploration (EFE) method.

The walker of the EFE method starts at a local minimum M_{curr} and follows the Hessian eigenvector with the lowest non-vanishing eigenvalue until a transition state is found. Then, the connectivity of the transition state is established by computing the steepest descent pathways to the two neighboring minima. If the transition state is connected to at least one minimum that already is in the database, the transition state and the connected minima are added to the database. Of course, only new configurations that are not yet recorded are added. The connectivity computation results in one of the following cases.

1. If the current minimum M_{curr} is connected to the transition state and if the other minimum M_{other} is lower in energy than the current minimum, the move is accepted: $M_{\text{curr}} \leftarrow M_{\text{other}}$

2. If the current minimum M_{curr} is connected to the transition state and if the other minimum M_{other} is higher in energy than the current minimum, the move is rejected: M_{curr} is kept unchanged.

3. If the current minimum is not connected to the transition state, the move is rejected: M_{curr} is kept unchanged.

Next, a new transition state search is initiated from M_{curr}. This is done by either following the negative direction of the previously followed mode (at this minimum), or if this already has been done, by following the direction of the eigenvector belonging to the next higher Hessian eigenvalue. For each minimum, only a maximum number $n_{\max} \leq 6N-12$ of transition state searches is performed (N is the number of atoms, free boundary conditions are assumed). If $n_{\max}$ is exceeded for the current minimum, no new transition state searches are initiated from this minimum and the algorithm jumps to the lowest-energy known minimum for which the maximum number of transition state searches have not been accomplished yet. The whole procedure is repeated until a certain number of minima or transition states is found.

Conventional methods for computing Hessian eigenvectors that are based on an iterative minimization of the curvature tend to converge to the lowest Hessian eigenvector only. Therefore, deterministic methods using mode following approaches based on these conventional eigenvector computation methods run into the risk of being non-ergodic, because the number of available escape directions away from a local minimum is very limited. In Sec. 2.2.1, a stabilized mode following technique was discussed, which reliably allows the convergence to the nearest Hessian eigenvector. This overcomes the problem of converging only to the lowest eigenvector and, therefore, can be used to follow with a greater reliability the full number of $6N-12$ mode following search directions available in a N-atomic system (free boundary conditions assumed). Consequently, it was the stabilized mode following technique of Sec. 2.2.1 that was used for the current implementation of the EFE method.

As part of the MHGPS benchmarks in Sec. 3.5.1, the efficiency of the EFE method was re-examined. Although the stabilized mode following method alleviates the problem of preferentially escaping from a minimum along just the lowest Hessian eigenvector, the results are similar to those that were found by others in previous investigations.[2,15] In general the efficiency of the EFE method is far from being optimal and occasionally EFE fails to find lowest-barrier pathways, even for moderately sized systems like LJ_{38}. In retrospect, this result is not very surprising. For example, it has been shown for LJ_{13} that at least 911 structurally distinct transition states are connected to the global minimum – a number that is much larger than the 66 mode following search directions that are maximally available in this system.[2]

3.3 Stochastic Surface Walking Based Reaction Sampling

At the time this thesis was written, the SSW-RS[185] method introduced by Zhang and Liu probably was the latest development in the field of reaction pathway finding. It appeared

after the publication of the MHGPS approach, which was developed by the author of this thesis and which is discussed in Sec. 3.5. In fact, the SSW-RS method is very similar in spirit to the MHGPS approach. The basic idea of MHGPS is to use a global optimization scheme as a guide through the energy landscape and then to use a selection of the emerging minima as input for a saddle finding scheme that is able to connect two given minima by a discrete path. MHGPS is based on MH and the particular advantages of this approach are described in Sec. 3.5. In contrast to MHGPS, SSW-RS is based on the stochastic surface walking[188] (SSW) global optimization scheme.

First, the SSW method generates an unsoftened vector which is a mixture of a random vector and a vector consisting of the bond directions of two randomly chosen atoms. Then, using a constrained softening procedure,[189] the high frequency modes are removed from this unsoftened vector and a softened, normalized escape direction $\mathbf{N}^n$ is obtained. A similar softening of an initial velocity is done in the MH algorithm. The SSW method leaves a catchment basin of a minimum by displacing the current configuration by a distance ds along $\mathbf{N}^n$, followed by a relaxation on a biased PES V_b that, in the i-th escape step, is given by

$$V_\mathrm{b} = V_\mathrm{real} + \sum_{n=1}^{i} h_n \exp\left[-\frac{\left((\mathbf{R}-\mathbf{R}^{n-1})\cdot\mathbf{N}^n\right)^2}{2ds^2}\right], \tag{3.1}$$

where the h_n and ds are the height and width of the Gaussian bias. By going back to the biased softening of the escape direction, this escape procedure is repeated until a given number of escape iterations are performed. After this, the bias potential is removed and the structure is relaxed on the real potential V_real. The new structure is accepted with probability

$$P_\mathrm{r} = \begin{cases} 0 & \text{if reaction occurs,} \\ 1 & \text{otherwise.} \end{cases} \tag{3.2}$$

Zhang and Liu suggested to compare the bond matrix and the chirality of the new and current minimum for deciding whether a reaction has taken place.[188]

With probability $P = (1-P_\mathrm{r})P_\mathrm{mc}$, where P_mc is the usual Metropolis monte carlo acceptance criterion that is also used in MHGPS, an approximate input path is generated between the current and the new minimum. If the highest energy along this approximate path is below a predefined threshold value, the exact transition state and the minimum energy pathway is computed. For generating the approximate input path, SSW-RS uses the double-ended surface walking method[190] and for refining the initial guess for the transition state, the single-ended constrained Broyden dimer method is used.[191]

By virtue of Eq. 3.2, both a reactant and product region is defined. As a consequence of this acceptance criterion, the SSW-RS walker only explores the reactant region and discrete paths are only computed to configurations that are behind the border separating the reactant and the product regions.

In general, it seems difficult to find descriptors that accurately distinguish arbitrary reactant and product states. In fact, this is a problem very similar to defining order parameters, a problem that is known to be far from trivial. In many cases the definition of order parameters is ambiguous and probably even impossible if there is only limited experience for a given system. Furthermore for systems in which the reaction has to pass through a third phase to go from the reactant to the product state, it seems not to be possible to find a reaction pathway in a single SSW-RS run.

So far, to the knowledge of the author, the SSW-RS method has been applied to reactions containing only a single intermediate transition state.[185]

3.4 Transition Path Sampling

In contrast to the other methods described in this chapter, transition path sampling[26,27,180–184] (TPS) does not explicitly compute pathways that lead through minima and transition states, but it samples dynamical pathways by generalizing importance sampling to path space. By doing so, TPS focuses on reactive pathways, which, by definition, are paths that connect two given states A and B. By not computing stationary points on the PES, TPS is fundamentally different from the other algorithms described in this chapter. However, precisely because TPS does not require any, potentially expensive, optimizations of stationary points, it is particularly interesting to see if TPS can offer performance advantages over the other methods. The purpose of this section is not to give an in-depth discussion of TPS, but to explain the basic ideas and to briefly discuss its performance for a system that also was used in the benchmarks of Sec 3.5.1.

In TPS a pathway $Z(\tau) := \{z_0, z_{\Delta t}, z_{2\Delta t}, \ldots, z_\tau\}$ is defined as an ordered sequence of states in phase space. This sequence can be looked at as snapshots of a trajectory that is given by the time evolution of length τ of a physical system. The path space points $z_{i\Delta t}$ are called time slices and the Δt denote the length of the time separation of consecutive time slices. Time slices are elements of the phase space and thus represent space and momentum coordinates. Usually, it is assumed that the time slices constitute a Markov chain, that is, the probability to move from $z_{i\Delta t}$ to $z_{(i+1)\Delta t}$ only depends on $z_{i\Delta t}$, but not on earlier time slices. Thus, the probability $P[z(\tau)]$ to observe such a path $z(\tau)$ is given by[26,27,180–184]

$$P[z(\tau)] = \rho(z_o) \prod_{i=0}^{\tau/\Delta t-1} p(z_{i\Delta t} \mapsto z_{(i+1)\Delta t}), \tag{3.3}$$

where $\rho(z_o)$ is the probability distribution of the initial time slice, for example a Boltzmann distribution. The detailed form of the transition probability $p(z_{i\Delta t} \mapsto z_{(i+1)\Delta t})$ depends on the actual type of the dynamics that is used for propagating the system, like Brownian, Newtonian, Monte Carlo or Langevin dynamics.[181,192] TPS focuses on sampling the rare event transitions, for which reason the probability function in Eq. 3.3 has to be augmented by indicator functions $h_A(z_0)$ and $h_B(z_\tau)$ that confine both path ends to the regions A and B, respectively. The

indicator function for region A is defined as

$$h_A(z_0) = \begin{cases} 1 & \text{if } z_0 \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

A similar definition applies for $h_B(z_\tau)$. Given these indicator functions, the transition paths are sampled according to the following probability distribution

$$P_{AB}[z(\tau)] = \frac{h_A(z_0)P[z(\tau)]h_B(z_\tau)}{Q_{AB}(\tau)}, \tag{3.5}$$

where, $Q_{AB}(\tau)$ is the normalization factor of the distribution. Eq. 3.5 defines the transition path ensemble, which can be sampled by using the Metropolis algorithm.[181,192] In the Metropolis algorithm new pathways have to be generated and accepted or rejected according to the Metropolis rule.[193] In particular, by virtue of the above indicator functions, pathways that not start in A and not end in B are always rejected.

The efficiency of TPS strongly depends on the details of the generation of new pathways. A combination of so called shooting and shifting moves have been reported to belong to the most effective pathway generation schemes.[192,194] In a shooting move, a new pathway is generated by randomly drawing from a uniform probability distribution a time slice of the present pathway and perturbing its momentum by means of an isotropically distributed random variable. The equations of motions (e.g. Newtonian) are then integrated backwards and forward in time, until the time slots 0 and τ are reached. A forward shifting move consists of removing a certain length $n\Delta t$ from the beginning of the path and integrating the equations of motions for the same time $n\Delta t$, starting at the end of the path. The trajectory length $n\Delta t$ is drawn from a probability distribution. Backward shifting moves are done in an analogous way, by removing a certain path length from the path end and integrating a new path segment, starting at the beginning of the path and integrating backward in time. Mainly in the transition region new pathways generated by shifting moves are very similar to the old pathways. However, shifting moves are reported to improve convergence of averages computed from the path ensemble.[192] In contrast to the shifting move, the shooting moves ensure an ergodic sampling of the transition path ensemble.[192] The details of these moves and their acceptance or rejection can be found in several TPS review articles.[31,180,181,192,194,195]

It is important to note that free energies and reaction rates cannot straightforwardly be obtained from a simple TPS simulation.[31,192,196] TPS simulations are used for finding pathways between A and B,[192] whereas free energies and reaction rates need additional computation. For these additional calculations, however, the TPS framework in combination with umbrella sampling can be used. In these umbrella sampling simulations, the whole range of the order parameter $\lambda(x)$, which is used to distinguish the regions A and B, is divided into overlapping intervals. Then, for each interval, a separate transition path sampling run has to be performed, for which reason the reaction rate and free energy calculations are more expensive than straight forward TPS simulations.[26,27] The reason why free energies and reactions rates cannot

be obtained from straightforward TPS is the following. The points along the paths in the transition path ensemble are not distributed according to the equilibrium distribution ρ, because the indicator functions h_A and h_B introduce a bias of the path probability such that low probability configurations with respect to ρ are favored if they lie on pathways that must be traversed when going from A to B.[192,196]

Already in 1998, TPS was used to study rearrangement processes in small, two-dimensional Lennard Jones (LJ) clusters.[27] However, to the best of the knowledge of the author of this thesis, it was not before 2007 that results for more complicated three dimensional Lennard Jones cluster, like LJ_{38}, appeared in the literature.[30] It was reported by Miller and Predescu,[30] that TPS with shooting and shifting moves becomes trapped in the high-energy structures of LJ_{38} and cannot find the basins of stability. They thus developed a double-ended transition path sampling method, named sliding and sampling, which did find pathways between both funnels.[30] However, for LJ_{38} the main drawbacks of their method are the non-ergodicity of their simulation and the high computational cost of 10^5 CPU hours. Unfortunately, it was reported that even this improved technique could not identify the lowest known reaction pathways connecting the two energetically lowest structures of the LJ_{38} system, even though the sampled transition pathways were quenched to obtain minimum energy pathways.

It remains to be mentioned that it can be difficult to apply TPS to systems for which no previously obtained experience is available. The reason is that an order parameter is needed that must be able to discriminate between the A and B states. In particular, the order parameter must allow to define the A and B states such that they form disjoint sets.[31,180,194] Furthermore, an initial path connecting A and B is needed and it was reported in Ref. [192] that no universally applicable procedure is available for TPS that would be suitable for generating initial trajectories. Nevertheless, pathways generated by the other methods described in this chapter could be used for seeding TPS simulations. In particular qualitative connectivity databases, which are introduced in Sec. 5.2, may offer promising starting points for constructing initial pathways.

3.5 Minima Hopping Guided Path Search[†]

Searching for reaction pathways and the exploration of the connectivity of energy landscapes requires an algorithm that moves efficiently inside one funnel and between several funnels. An algorithm that has proven its efficiency in exploring the low energy regions of potential energy landscapes is the MH global optimization method.[9,18,19,110,197,198] The success of MH relies in large parts on the MD-type moves and on an energy feedback which satisfies the explosion condition.[18,199] The MD moves assure that only physically realizable structures are explored and, by means of energy conservation, only low-energy barriers are surmounted in

[†]The research presented in this section has been published in B. Schaefer, S. Mohr, M. Amsler, and S. Goedecker, "Minima Hopping Guided Path Search: An Efficient Method for Finding Complex Chemical Reaction Pathways", The Journal of Chemical Physics **140**, 214102 (2014). Reprinted with permission. Copyright 2014 by the American Institute of Physics.

unexplored regions of the PES. In well explored regions the explosion condition rigorously guarantees an exponential increase of the kinetic energy. Therefore, in contrast to many other PES exploration methods, MH is able to escape automatically from any funnel, irrespective of its depth. In general, the MD trajectories of MH are short and, therefore, one can expect consecutive minima along the MH trajectory to be structurally similar to and well aligned with each other. This significantly simplifies the process of finding intermediate transition states without the need of an explicit and computationally expensive optimization of the geometric and permutational structural alignment.[167] In conclusion, MH explores PESs efficiently, without the risk of becoming trapped. Moreover, MH generates consecutive minima that are particularly suitable for the input of methods that are intended to find transition states located between the input minima. It, therefore, seems natural to combine the capabilities of MH with a method that connects two given minima by a series of transition states to a minima hopping guided path search (MHGPS) technique.

MH, and with it MHGPS, starts at some local minimum and tries to escape from its catchment basin by following a short, random and soft mode biased MD trajectory at the end of which a local geometry optimization is performed. The softening procedure has been described previously[9] and it has been demonstrated that following softened escape directions preferentially leads over low energy barriers.[200] The escape trials are repeated until MHGPS successfully escapes from the catchment basin of the current minimum. In order to avoid becoming trapped in the current catchment basin, the kinetic energy is increased by a factor $\beta_s > 1$ after each failed escape trial. When MHGPS successfully escapes to a different minimum it either decreases the kinetic energy by a factor $\beta_n < 1$ or increases it by a factor $\beta_o > 1$, depending on whether the new minimum has been visited before or not. This introduces a feedback which promotes cooling down in unexplored regions and heating up in well explored regions of the potential energy landscape and thus ensures that the algorithm quickly samples the bottom of a funnel and at the same time does not become trapped in it.

Based on a Metropolis-like[193] criterion it is decided in the MHGPS scheme whether the current minimum M_{curr} and the new minimum M should be connected by a discrete path. If the energy of the new minimum E is lower than the energy E_{curr} of the current minimum, a connection attempt is always made. If its energy is higher than the energy of the current minimum, an attempt is made with a probability of

$$\exp\left(-\frac{E - E_{\text{curr}}}{E_{\text{diff}}}\right). \tag{3.6}$$

The parameter E_{diff} resembles the energy $k_B T$ of an ordinary Metropolis simulation. However, in contrast to an ordinary Metropolis simulation, E_{diff} constantly is adjusted. If the decision is made to connect M_{curr} and M, E_{diff} is decreased by a factor $\alpha_a < 1$, otherwise it is increased by a factor $\alpha_r > 1$. The connections are made by recursively applying bar-saddle and following approximate steepest descent paths from emerging intermediate transition states. Establishing the connection between the two bar-saddle input minima M_{curr} and M in a recursive or iterative fashion is essential, because there is no guarantee that the two minima M_{curr} and M

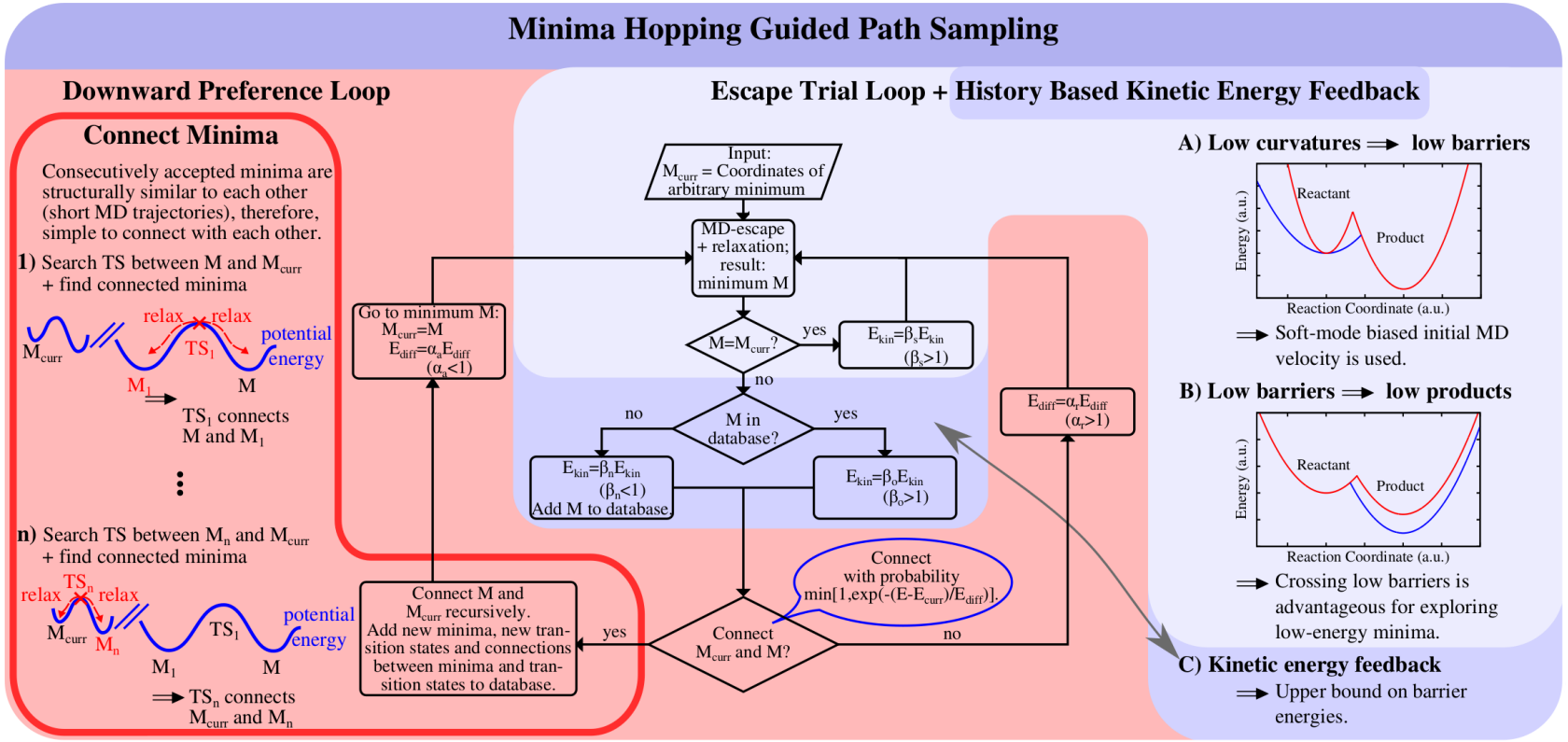

Figure 3.1: Illustrated is the MHGPS approach to reaction path sampling. The different parts of the algorithm are highlighted by different background colors. For the escape from a local minimum, a light blue is used. The kinetic energy feedback is highlighted by using dark blue and for the part that introduces a preference for the low-energy region of the PES a red background color is used. The essential part of the MHGPS method is contained in the red L-shaped box (left-hand side). Removing this box, which means going to minimum M (with probability $\min[1, \exp(-(E - E_{\text{cur}})/E_{\text{diff}})]$) without searching for any intermediate minima and transition states, results in the standard MH algorithm for finding the global minimum.[18] The MHGPS approach relies on the efficient MH exploration of the low-energy part of the PES (panel B). Furthermore, as illustrated in panels A and C, MH tends to cross low energy barriers, which is favorable in the context of reaction pathway search. The fact that MH uses short MD trajectories for moving on the PES results in a sequence of consecutive minima that are structurally similar to each other. This makes it simple to connect them with each other by a discrete path.

can be connected with each other by exactly one transition state. Hence, during a connection, intermediate transition states can appear that might not be connected to one or to both of the two input minima. In such a case the minima to which the intermediate transition states are connected also have to be connected to the corresponding bar-saddle input minima in order to obtain a discrete path that properly connects M_{curr} and M.

After connecting M_{curr} and M by a discrete path, the new minimum becomes the current one and the algorithm starts a new MD trajectory at this minimum. The whole procedure is stopped as soon as a given number of distinct minima or transition sates are identified.

Fig. 3.1 shows a flow chart of the MHGPS approach. The above explained procedure of iteratively connecting two minima is visualized in the left-hand side of this figure. The desirable consequences for the MHGPS approach of the softening procedure and of the energy conservation in the MD moves are described in panels (A) to (C) on the right-hand side of this figure.

In all MHGPS simulations presented in this chapter the standard MH parameters ($\beta_{\mathrm{s}} = \beta_{\mathrm{o}} = 1/\beta_{\mathrm{n}} = \alpha_{\mathrm{r}} = 1/\alpha_{\mathrm{a}} = 1.05$) were used.[9,199] However, if $\beta_{\mathrm{s}} > 1$ and as long as the remaining parameters fulfill the explosion condition

$$\frac{\log(\alpha_{\mathrm{r}})}{\log\left(\alpha_{\mathrm{a}}^{-1}\right)} \geq \frac{\log\left(\beta_{\mathrm{n}}^{-1}\right)}{\log\left(\beta_{\mathrm{o}}\right)}, \tag{3.7}$$

the efficiency of the MH based exploration of the PES is not very dependent on the detailed values of these parameters.[199] The explosion condition follows from requesting that the net-kinetic energy should rise if the algorithm begins to reject all new minima and only accepts previously visited minima. The rise in kinetic energy assures the escape from well explored regions of the PES. For the sake of being self-contained, a reiteration of the derivation of the explosion condition can be found in Appx. D.

MHGPS is not limited to using bar-saddle for the purpose of connecting minima. In principle any saddle search method that can find transition states between two given minima, like for example the Nudged Elastic Band method[51–54] or the Splined Saddle method[177,178] can be used. During the development of MHGPS, it was decided to use the bar-saddle method, because this was the most reliable implementation of a saddle finding method that was available to the author at this time. As of the writing of this thesis, bar-saddle has been replaced by a combination of the freezing string method[158] and the SQNS saddle finding method introduced in Sec. 2.3.6. By selecting the highest energy node along the freezing string path, a well educated guess for the transition state is obtained, which then is tightly converged to the exact transition state using the SQNS method. Bar-saddle was replaced to achieve a better efficiency and stability at ab-initio level. SQNS was not used for the initial version of MHGPS described in this chapter, because SQNS was developed after the advent of MHGPS.

3.5.1 Benchmarks and Applications

In all sampling approaches used below, the connectivity between transition states and minima was determined by stepping away from a transition state by adding to and subtracting from the transition state one-100th of the normalized Hessian eigenvector that corresponds to the negative curvature. Then, from the thus obtained points, Euler's method with a maximum step size of $10^{-2}\sigma$ was started, in order to approximate steepest descent paths. The Euler integrator was stopped, as soon as it entered the quadratic region surrounding a minimum. In this Euler integration scheme steps were rejected and the step size was decreased if either the angle between the gradients of two successive steps was larger than 60 degree or if the energy increased. Inside the quadratic region the Euler method was replaced by the fast inertial relaxation engine (FIRE)[132] in order to speed up the geometry optimization. For the FIRE integrator itself, it is not of any relevance whether it operates inside the quadratic region or not. However, compared to non-quadratic regions it seems less likely that inside the quadratic region the FIRE method will converge to a different minimum than Euler's method. Because dynamic properties computed from stationary point databases are unlikely to depend strongly on whether the connectivity of the potential energy landscape is established by using approximate steepest descent paths or paths from advanced minimization algorithms[3,40] like for example FIRE or BFGS,[133–136,201] the time used for relaxations to local minima could have shortened significantly when omitting the Euler integration and using advanced minimization algorithms throughout. However, because a new reaction pathway search method (MHGPS) is benchmarked and applied, the conservative Euler integration approach was used in order to sample connectivity information that is in accordance with the connectivity defined by the widely accepted intrinsic reaction coordinate.[77] Although no results based on FIRE-only minimization are reported, the differences of pathways obtained from FIRE-only and Euler integration plus FIRE optimization were compared. Only small changes in the number of intermediate transition states could be observed. In all cases the energetically lowest transition state between two states found by FIRE-only runs was identical to the lowest transition state found by connections established by approximate steepest descent paths.

Benchmarks

In contrast to global minimum searches, a performance analysis of stationary point database generation algorithms is not straightforward since there is no obvious stopping criterion. One possible stopping criterion can be defined by checking whether a putative lowest-barrier pathway between two minima has been found. Because of the computational cost of Dijkstra's algorithm, this check is not feasible if it has to be performed between every pair of minima for a given system. Therefore, a suitable test system should contain two outstanding and well defined minima for which pathways that connect them can be examined. The global minimum of LJ_{38} is located inside a small funnel containing fcc-like structures, the second-lowest minimum of LJ_{38} is contained inside a comparatively large icosahedral funnel. Both funnels are separated by a high energy barrier.[2,35] Furthermore, the number of atoms in LJ_{38}

Table 3.1: Results of performance test for LJ_{38}. Averages for $\langle n_{\mathrm{ts,diff}} \rangle$ and $\langle n_{\mathrm{ts}} \rangle$ are taken over $1000 - n_{\mathrm{f}}$ independent and successful runs.

Method	n_{ev} [1]	$\langle n_{\mathrm{ts,diff}} \rangle$ [2]	$\langle n_{\mathrm{ts}} \rangle$ [3]	n_{E} [4]	n_{f} [5]
MHGPS	n/a	9267	14580	3464	0
EFE	10	64611	168688	3384	24
EFE	25	72977	192097	3508	8
EFE	40	91313	268422	3492	1

[1] Number of lowest eigenvectors along which transition states were searched in positive and negative direction
[2] Average number of distinct transition states needed to be found before identifying a lowest-barrier pathway.
[3] Average number of transition states computations needed before identifying a lowest-barrier pathway.
[4] Number of totally performed energy evaluations divided by the number of totally performed transition state computations. The number of energy evaluations include the evaluations used for transition state searches, minimizations, softening and MD (if applicable).
[5] Number of runs in which lowest-barrier pathways could not be found before identifying 5×10^5 distinct minima.

is small enough to perform a sufficient number of runs within a feasible amount of time. Therefore, LJ_{38} possesses all properties of a suitable benchmark system.

Table 3.1 shows the results of a performance test based on 1000 independent runs for LJ_{38}. Each run was started using a random non-fcc structure as input geometry and, depending on what happened earlier, was either stopped as soon as the putative lowest-barrier pathway between the global minimum and the second lowest local minimum of LJ_{38} was identified, or if 5×10^5 distinct local minima were found. For all methods and all runs the same convergence criteria for the stationary points were used.

EFE needed roughly between a factor of 12 to 18 more transition state computations than the MHGPS method before encountering a lowest-barrier pathway of LJ_{38}. Because the number of energy evaluations per transition state computation n_{E} are similar for both methods, similar factors are obtained when measuring the computational cost in terms of energy evaluations.

For the EFE method, a small number of runs were observed that failed to find a lowest-barrier pathway at all. Since the number of failure runs decreased with increasing number of followed mode directions, these failures can be explained by the limited number of search directions available to the EFE method. Assuming free boundary conditions, the EFE method can follow at maximum $6N - 12$ directions per minimum for a N-atom system. However, as already mentioned in Sec. 3.2, the number of transition states connected to a minimum can exceed the number of $6N - 12$ directions by far. For example it is known that the global minimum of LJ_{13} is surrounded by 535 local minima which are connected to the global minimum by 911 transition states.[2] It is, therefore, possible to miss stationary points that potentially lie on the

lowest-barrier pathway. This general restriction of the EFE-method and similar deterministic mode following methods has been mentioned before by Malek and Mousseau.[15]

The average number of distinct transition states $\langle n_{\mathrm{ts,diff}} \rangle$ divided by the average number of computed transition states $\langle n_{\mathrm{ts}} \rangle$ was between 66% and 87% larger for the MHGPS method than corresponding ratios of the EFE method.

The average CPU time required before MHGPS identified the lowest-barrier pathways between both lowest structures of LJ_{38} was measured to be roughly 8 minutes (on a single core of an Intel Xeon E5-2665 CPU clocked at 2.40GHz). This timing should be compared to the 10^5 CPU hours that were required for the sliding and sampling computations reported in Ref. [30]. These timings differ by several orders of magnitude and, therefore, allow to give a rough idea on the performance differences between the different methods. They are particularly noteworthy when noting that Ref. [30] only presents pathways that are higher in energy than the known lowest-barrier pathway. As well as MHGPS, the EFE method is also several orders of magnitudes faster than sliding and sampling. On average, EFE needed just under 3 CPU hours to find the lowest barrier path for LJ_{38} ($n_{\mathrm{ev}} = 10$, average taken over successful runs). As the CPU time depends very strongly on the computer hardware and the implementation of an algorithm, one should compare methods that do not exhibit such a distinct timing difference by using more suitable quantities like, for example, those given in Table 3.1.

Fig. 3.2 shows the histories of all transition state energies of two typical MHGPS (panel (a)) and EFE (panel (b)) runs that were performed for the LJ_{38} system. Both runs were started at non-fcc structures and thus are residing inside the large icosahedral funnel during the first transition state computations. Fig. 3.2 illustrates the distinct transition state sampling behavior of both methods. In the very beginning, the EFE method is able to sample low-energy transition states. However, with an increasing number of totally sampled transition states, the energies of the lowest transition states that are being sampled rises as well. This means, the EFE-method explores the energy landscape in a bottom-up fashion. In conjunction with the limited number of search directions per minimum, this is a severe problem, in particular for multi-funnel systems. As can be seen from Fig. 3.2, in the beginning of the sampling procedure the bottom-up sampling forces a very detailed exploration of the icosahedral funnel. The EFE method is, therefore, not able to escape from the icosahedral to the fcc funnel until roughly 5000 transition states were computed. In very long runs, the same bottom-up sampling of the EFE method will lead at some point to the computation of transition states that almost entirely have energies above the highest barrier along the lowest-barrier pathway (energies above the bold, dashed and black line in Fig. 3.2). If the lowest-barrier pathway could not be found until this critical point, it is very unlikely that the EFE method will find it later. In contrast to the EFE method, the MHGPS method escapes from the icosahedral to the fcc funnel very quickly and regularly switches back and forth between both funnels. Because MHGPS does not strictly avoid previously visited low energy configurations, it does not suffer from the consequences of a strict bottom-up sampling. MHGPS is always able to go down to previously explored low energy configurations, however the history based energy feedback of MH takes care that well

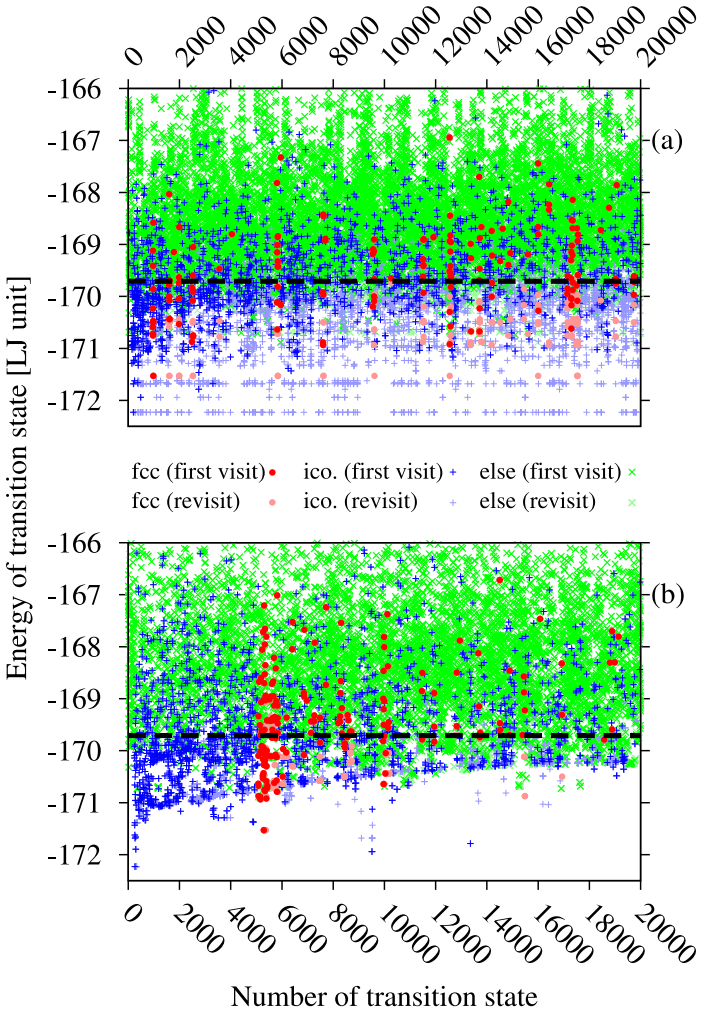

Figure 3.2: Scatter plots showing all computed transition state energies in chronological order. The shown data belongs to typical MHGPS and EFE runs for the LJ_{38} two-funnel system. Panel (a) shows MHGPS data, panel (B) displays EFE data. Transition states belonging to the fcc funnel are represented by red • and transition states belonging to the icosahedral funnel are represented by blue +. The green × represent all remaining transition states. If a transition state is visited for the first time, the respective data point is dark-colored, otherwise it is light-colored. The bold dashed line located at an energy of roughly -169.709ϵ represents the highest barrier along the lowest-barrier pathway connecting the two energetically lowest minima of LJ_{38}. An interpretation of this figure is given in the text of Sec. 3.5.1. Reprinted with permission from Ref. [62]. Copyright 2014 by the American Institute of Physics.

explored regions are left quickly. Therefore, as illustrated in Fig. 3.2, MHGPS is able to sample transition states from the whole energy range at any stage of the sampling procedure.

For the 75-atom Lennard Jones system a similar behavior as for LJ_{38} was found. Starting at the second lowest minimum of LJ_{75}, which is located in an icosahedral funnel, MHGPS and EFE test runs were preformed. The runs were stopped as soon as 275,000 transition states were computed. Within this amount of computed transition states, the present implementation of the EFE method showed not to be able to leave the icosahedral funnel, whereas the MHGPS method could switch between both LJ_{75} funnels multiple times.

For the LJ_{55} cluster, which is a strong structure seeker,[2] a short test runs was performed too. Despite its structure seeker character there exist two non-icosahedral minima which lie behind comparatively high barriers.[2,49,202] The test run of each method was started at the same arbitrarily chosen high energetic local minimum (-270.302962 ϵ) and was stopped as soon as 30,000 transition state computations were performed. The overall appearance of the disconnectivity graph containing the lowest 700 minima generated from EFE-sampling is equivalent to the graph presented in Ref. [2], however, in this test run, the present implementation of the EFE method could not identify the lower one of the two non-icosahedral minima. The other of the two mentioned non-icosahedral minima could be found by the EFE method, however the barrier connecting it to the global minimum funnel was significantly larger than the barrier found in Ref. [2]. In contrast, the disconnectivity graph containing the lowest 700 minima generated from the MHGPS run contained all important features of the LJ_{55} potential energy landscape, including both of the above mentioned non-icosahedral minima. The barriers connecting the two non-icosahedral minima to the global minimum funnel were also reproduced in accordance with the barriers of the disconnectivity graph presented in Ref. [2].

Application of MHGPS to LJ_{75} and LJ_{102}

Due to its advantages presented in Sec. 3.5.1, MHGPS was applied to the PESs of LJ_{75} and LJ_{102}. Concerning the task of sampling relevant stationary points, in particular LJ_{75} is known to be a very difficult system. This can be explained by the frustration of its PES and the large geometrical differences of both structures located at the bottoms of the two major funnels.[2]

For each system, 10 independent runs were started at the global minimum structures. For every run different random seeds were used. A run was stopped, as soon as 2×10^6 distinct local minima were found. For the analysis of the PESs, the stationary point databases resulting from all runs were merged into a single database for each system. For LJ_{75} this procedure resulted in a stationary point database containing roughly 12.0×10^6 distinct transition states connecting 7.0×10^6 distinct local minima. In case of LJ_{102}, a database containing roughly 10.9×10^6 distinct transition states, connecting 7.5×10^6 distinct local minima was obtained by this procedure. The disconnectivity graphs of both system are shown in Fig. 3.3. Fig. 3.4 and Fig. 3.5 show plots of the energy along the reaction pathways in dependence of the integrated path length S which is defined by the arc length of the steepest descent reaction path in the

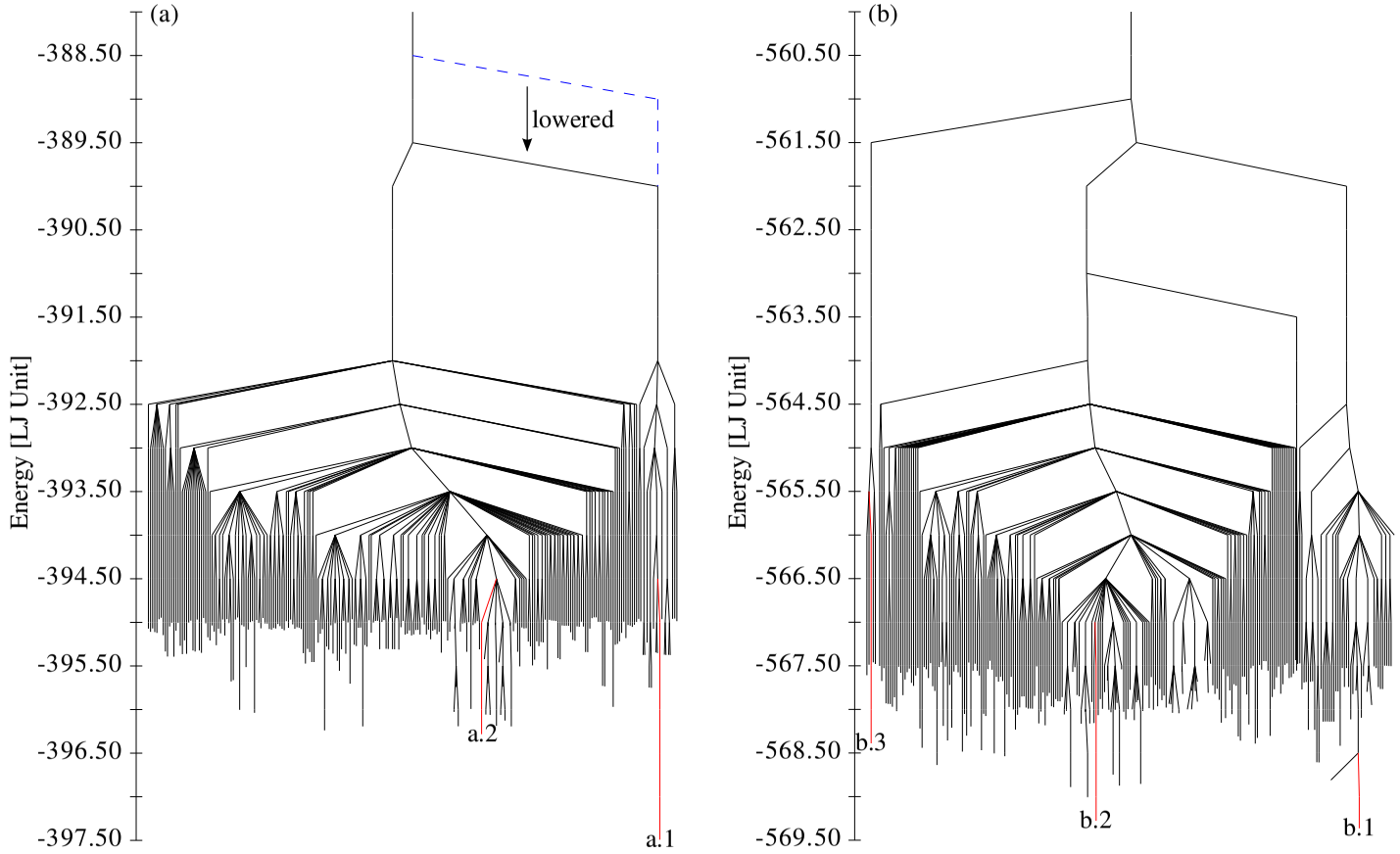

Figure 3.3: Disconnectivity graphs of LJ_{75} [panel (a)] and LJ_{102} [panel (b)]. Panel (a) shows the new putative lowest barrier between both funnels. The blue dashed line indicates the previously known lowest barrier connecting both funnels.[2] Panel (b) shows a third, previously unknown, funnel with an energetically low bottom structure (minimum b.3) and a high barrier connecting it to the other two funnels. Both graphs show the 250 lowest minima that were found for each system. The bottom structures of each major funnel are labeled and highlighted using red color.

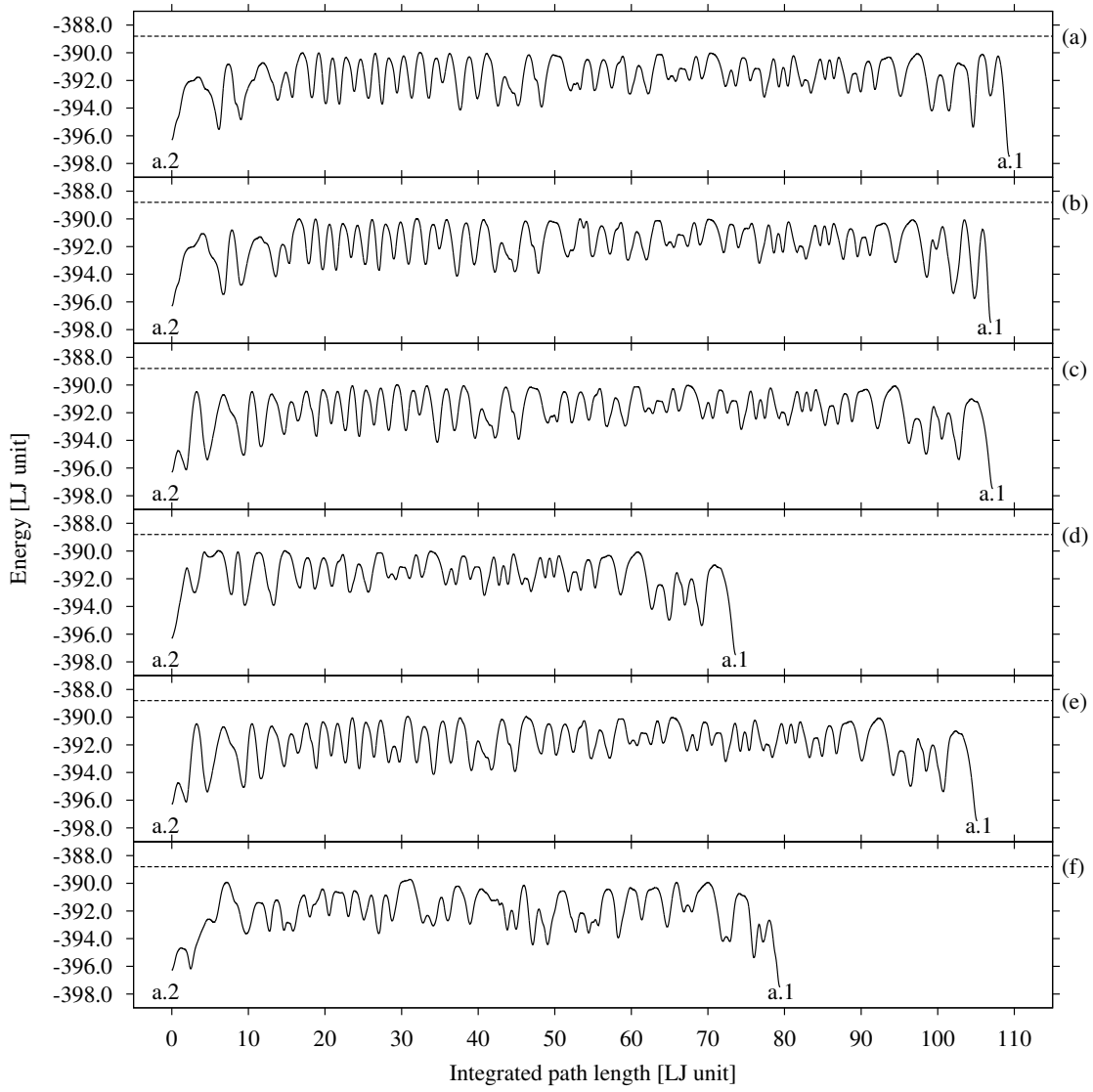

Figure 3.4: Pathways found by MHGPS that connect the bottom-structures of both LJ_{75} funnels (configurations a.2 and a.1 of Fig. 3.3). The dashed horizontal lines indicate the highest energy along the previously known lowest-barrier pathway.[2] Panels (a), (b) and (c) show three alternative putative lowest-barrier pathways. Panels (d), (e) and (f) show pathways that have been obtained by successively removing the highest energy transition state along the lowest-barrier pathway from the stationary point database [panels (d) and (e)] or from a preliminary test run [panel (f)]. They only have slightly higher barriers than the pathways of panels (a) to (c) and thus show that there exist a variety of pathways lying energetically between the lowest MHGPS results and the previously presented[2] lowest-barrier pathways for LJ_{75}. Reprinted with permission from Ref. [62]. Copyright 2014 by the American Institute of Physics.

$3N$-dimensional coordinate space.[50] Numerically the integrated path length is computed by summing up all the lengths $|\Delta\mathbf{R}|$ of all steepest descent steps:

$$S = \sum_{\text{steps}} |\Delta\mathbf{R}| \,. \tag{3.8}$$

In addition to the conservative combination of Euler's method and FIRE mentioned in Sec. 3.1, all new pathways explicitly reported (Figs. 3.4 and 3.5) were double-checked in a post-processing step. In order to obtain quasi-exact intrinsic reaction pathways, steepest descent paths were recomputed exclusively using Euler's method with a maximum displacement of $10^{-6}\sigma$ in each integration step. Before this steepest descent relaxation the structures were pushed away from the transition state one-10,000th of the normalized eigenvector belonging to the negative Hessian eigenvalue.

For LJ_{75} the highest barriers along the lowest-barrier pathways connecting the two major funnels that were found by MHGPS are significantly lower in energy than those of the previously known lowest-barrier pathways. Fig. 3.3a shows the MHGPS generated disconnectivity graph for LJ_{75}. Using Dijkstra's algorithm as outlined in Sec. 3.1, roughly 20,000 pathways, all having the same highest-barrier energies of 7.51ϵ and 6.30ϵ and the same number of 51 intermediate transition states, were identified. Compared to this, the previously known lowest-barrier pathway has significantly higher highest-barrier energies of 8.69ϵ and 7.48ϵ and possesses 65 intermediate transition states.[2] In order to illustrate typical differences between alternative lowest-barrier pathways, Fig. 3.4a, Fig. 3.4b and Fig. 3.4c explicitly show the steepest descent reaction paths of three lowest-barrier pathways. In order to check whether there might exist further pathways, which are energetically in-between the previously known lowest-barrier pathway and the putative lowest-barrier pathways found by MHGPS, the highest energy transition states along the lowest-barrier pathways were successively removed from the stationary point database, and Dijkstra's algorithm was applied. Pathways resulting from this removal are shown in Fig. 3.4d and Fig. 3.4e. For the pathway shown in Fig. 3.4d the barriers are 7.52ϵ and 6.31ϵ, for the pathway of Fig. 3.4e the barriers are 7.54ϵ and 6.33ϵ. They are only slightly higher in energy than the highest barriers along the putative lowest-barrier pathway. This suggests that there exists a whole range of pathways that are energetically between the putative lowest pathways presented in this study and the previously known lowest pathway. This conjecture seems to be substantiated by the pathway shown in Fig. 3.4f. This pathway was found in a preliminary single-run test in which only roughly 6×10^5 distinct local minima and roughly 9×10^5 distinct transition states were sampled. The highest barriers along this pathway are 7.78ϵ and 6.57ϵ.

As shown in Fig. 3.3b, MHGPS found a previously unknown funnel for LJ_{102}.[63] An illustration of the bottom structure of this funnel is given in Fig. 3.6. The new bottom structure possesses icosahedral elements and its surface is dominated by buckled hexagonal patches. Its has an energy of -568.388773ϵ.

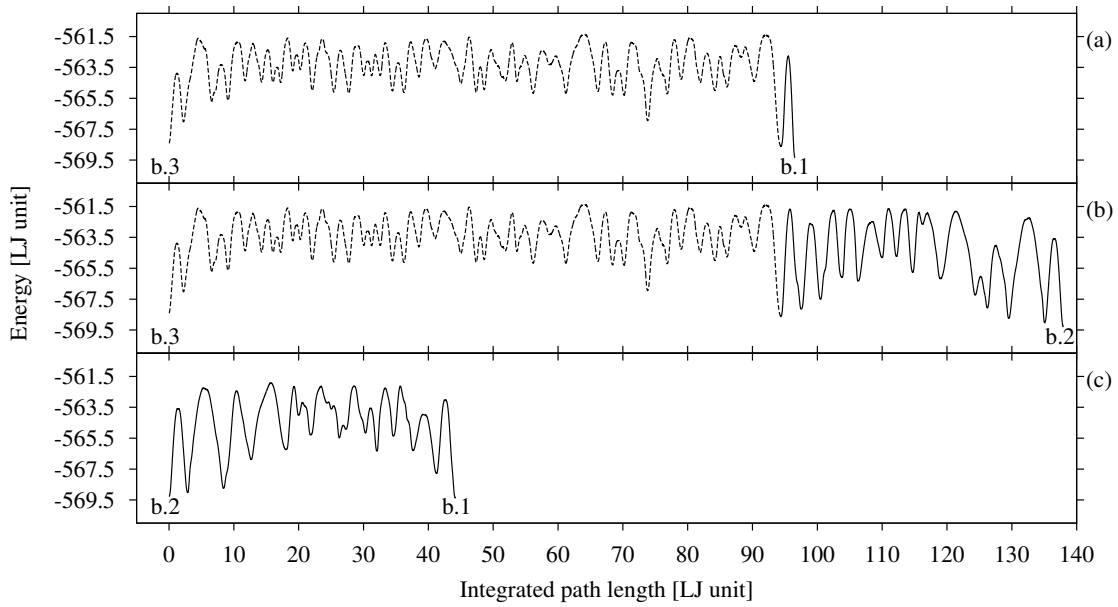

Figure 3.5: Putative lowest-barrier pathways that were found by MHGPS for LJ_{102}. Panel (a) shows a putative lowest-barrier pathway connecting the putative global minimum (configuration b.1 of Fig. 3.3) to structure b.3 of Fig. 3.3. A lowest-barrier pathway connecting the second-lowest minimum of LJ_{102} (configuration b.2 of Fig. 3.3) and configuration b.3 of Fig. 3.3 is shown in panel (b). The parts of the reaction pathways shown in panel (a) and (b) that coincide with each other are highlighted by using dashed lines. Panel (c) shows a putative lowest-barrier pathway connecting the second-lowest configuration of LJ_{102} (configuration b.2 of Fig. 3.3) to the putative global minimum (configuration b.1 of Fig. 3.3). Reprinted with permission from Ref. [62]. Copyright 2014 by the American Institute of Physics.

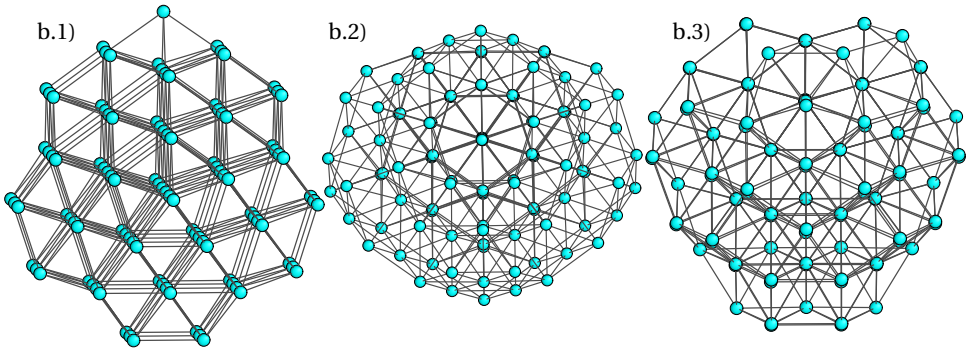

Figure 3.6: Bottom structures of the three major funnels of LJ_{102}. The labeling of the illustrations corresponds to the labeling of Fig. 3.3b. Reprinted with permission from Ref. [62]. Copyright 2014 by the American Institute of Physics.

Lowest-barrier pathways connecting the new structure to the global minimum and to the second lowest minimum are shown in Fig. 3.5a and Fig. 3.5b. The lowest-barrier pathways connecting this new structure and the global minimum contain 40 intermediate transition states and the highest barriers are 7.97σ and 7.89σ. The highest barriers of the lowest-barrier pathways that connect the second lowest minimum to the bottom of the new funnel are 7.97σ and 7.00σ. These pathways contain 53 intermediate transition states.

Furthermore, MHGPS could confirm the energy of the highest barrier along the putative lowest-barrier pathway connecting the global minimum to the second lowest minimum.[63] However, both in terms of the number of intermediate transition states and in terms of the integrated path length, the pathway found by MHGPS is significantly shorter than the previously known pathway. It contains only 16 intermediate transition states compared to 30 transition states contained in the pathway published earlier.[63] The integrated path length is roughly 11σ shorter (difference of paths length was estimated on the basis of the plot given in Ref. [63]).

3.5.2 Conclusion

MH is a practical guide for the search of low-barrier reaction pathways, because it uses short MD moves for the exploration of PESs and an energy feedback that satisfies the explosion condition[18,199]. As a consequence of the short MD moves, consecutive minima along the MH trajectory are structurally not too different from each other and thus are well suited as input structures for methods that can find transition states between two given input geometries. Furthermore, energy conservation assures that the maximum barrier energy between two consecutive minima is bounded from above. The explosion condition assures that the MH guide does not get stuck in deep funnels. As a consequence, MHGPS must perform computationally expensive transition states computations only between minima that are particularly promising for the purpose of finding energetically low barriers and between minima that are promising for the exploration of the potential energy landscape. MHGPS needs no human intuition and its MH based exploration of the PES is completely unbiased. It, therefore, does not fail to explore unforeseen and unexpected features of potential energy landscapes. In comparison to the EFE mode following approach, MHGPS detects a significantly larger number of distinct transition states when performing the same number of transition state computations. MHGPS reduces the cost of sampling stationary points and their connectivity information by over one order of magnitude compared to the EFE mode following approach. In contrast to other methods, MHGPS could successfully find the lowest-barrier pathways of LJ_{38} in all tests. The efficiency of the novel MHGPS scheme is also confirmed by new results that were found for LJ_{75} and LJ_{102}, systems that previously have been thoroughly examined for more than a decade.

4 Isomerism and Structural Fluxionality in the Au_{26} and Au_{26}^- Nanoclusters*

Since bulk gold is the most inert metal one could expect gold clusters to show no or only negligible chemical activity.[203,204] However, compared to bulk gold, the chemistry of gold nanoparticles is dramatically different, leading to promising and valuable properties for nanosciences like nanoelectronics, nanobiology and nanocatalysis.[205–210] In order to understand nanogold related processes, size-selected gold clusters have been the focus of both theoretical and experimental investigations.[198,204,205,211–240] In particular, joint photoelectron spectroscopy and theoretical studies have elucidated the structures of anionic gold clusters over a wide size-range.[204,221,225,235] For example, Au_{20}^- has been found to exhibit a tetrahedral pyramidal geometry[221] and molecular dynamics (MD) simulations of Au_{34}^- suggest this cluster to have a fluxional shell which could promise chemisorption.[204] Recently, Au_{36}^- to Au_{38}^- were shown to exhibit core-shell structures with a four-atom tetrahedral core.[240] Despite the successful previous work on gold clusters,[225,235,240] no theoretically predicted structure of Au_{26}^- has been experimentally confirmed so far. In order to close this gap the results of a joint theoretical and experimental photoelectron spectroscopy study on this missing cluster are outlined in this chapter.

Using the Minima Hopping[9,18,110] (MH) method, the author of this thesis carried out the structure prediction computer experiments and identified new global minimum candidates for both the neutral and anion systems. Members of the group of Prof. Xiao Cheng Zeng simulated the photoelectron spectra of energetically low-lying Au_{26}^- clusters. The photoelectron spectroscopy experiment and the interpretation of the experimental spectrum was performed by Prof. Lai-Sheng Wang's group. The comparison between the theoretical and the experimental results for Au_{26}^- is used to identify energetically low-lying nanostructures that most probably exist in the experiment.

*The research presented in this chapter has been published in B. Schaefer, R. Pal, N. S. Khetrapal, M. Amsler, A. Sadeghi, V. Blum, X. C. Zeng, S. Goedecker, and L. Wang, "Isomerism and Structural Fluxionality in the Au_{26} and Au_{26}^- Nanoclusters", ACS Nano **8**, 7413 (2014). Reprinted with permission. Copyright 2014 by the American Chemical Society.

For the anionic system also a disconnectivity graph[73] is presented, which is based on a database of connected minima and transition states that were computed at the density functional level of theory (DFT). The construction of this database was performed by the author of this thesis. This database is complete enough to allow the prediction of chemical activity at finite temperatures. To the knowledge of the author, this is the first time that such a local map of the energy landscape has been computed completely at the DFT level of theory. At the time of the research for this chapter, the MHGPS method of Sec. 3.5 had not existed yet. Therefore, the topology of the potential energy surface (PES) had to be established in a non-automatized approach that is outlined below. However, this work forms the starting point of the development of the MHGPS method, as well as of the remaining research that was conducted for this thesis.

4.1 Methods

4.1.1 Global Optimization of Au_{26} and Au_{26}^-

For global minimization of the PES the Minima Hopping (MH) method[9,18,110] coupled to the BigDFT code[148] was utilized. Within this thesis, the MH method already has been introduced in Sec. 3.5 as part of the MHGPS scheme and it is referred to this section for a description of MH. The global optimization was performed completely at the DFT level which has shown to be more efficient for the present system size than performing a global minimum search on a force field or other less accurate methods and post-relaxing an energetically low lying subset of configurations using DFT methods.[110] For all local geometry optimizations a combination of conjugate gradient[131] and modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[133–136] as implemented in the BigDFT code was used.

In order to predict the putative global minimum of the Au_{26} and Au_{26}^- potential energy landscapes, the computations were split into several steps. As the first step MH runs were performed on the energy landscape of the neutral Au_{26} cluster, using both the local density approximation (LDA)[86,96] and the Perdew-Burke-Ernzerhof (PBE)[105] functionals in combination with the corresponding relativistic and norm-conserving Hartwigsen-Goedecker-Hutter (HGH) semi-core pseudopotentials.[241,242] For each functional several separate runs with different starting configurations were performed. A selection of the initial configurations is shown in Fig. 4.1. The starting configurations were constructed manually or are from previous work, such as the tubular structure (Fig. 4.1c) which once was proposed to be the global minimum of Au_{26}.[243] The BigDFT code uses a systematic wavelet basis set. The corresponding grid spacing as well as the spatial extension of the basis function were chosen such that a rotation of a whole cluster in the $\mathbb{R}^3$-space changed the energy by less than 10^{-4} Hartree. For all MH runs, geometry optimizations were stopped as soon as 20% of the force consisted of computational noise. On average this happened when the largest force acting on any atom in a cluster was approximately 5×10^{-5} Hartree/Bohr in case of the LDA functional and 3×10^{-4} Hartree/Bohr for PBE.

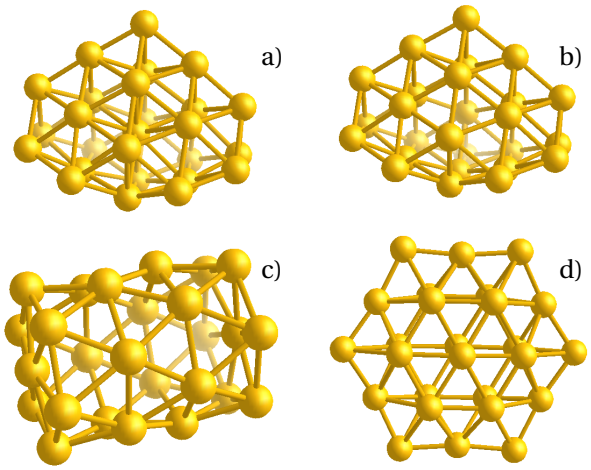

Figure 4.1: Representatives of the structural motifs. Isomer (a) represents the core-shell structure with a single internal atom, isomer (b) the empty cage, isomer (c) the high-symmetry tubular structure and isomer (d) the hexagonal motif. These isomers also constitute a selection of starting configurations used for different Minima Hopping runs. Members of each motif may not necessarily be element of the same point group as the sown representatives, but rather most configurations belong to the low-symmetry C_1 point group. Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

In the second step the minima of the LDA runs were post-relaxed while using the PBE functional and vice versa. Again parameters were used that result in an energy accuracy better than 10^{-4} Hartree and the largest force acting on any atom of approximately 5×10^{-5} Hartree/Bohr (LDA) and 3×10^{-4} Hartree/Bohr (PBE). The sets consisting of local minima from all the different runs were then merged. During this merge it was taken care to remove duplicate configurations. Two sets of local minima were formed: one from LDA and the other from PBE calculations. Both sets were relaxed a last time by treating the configurations as singly charged anions using both, LDA and PBE functionals. Again, configurational duplicates that emerged in this relaxation process were removed. The parameters for the relaxations of the anions were chosen such that the energy changed by less than 10^{-4} Hartree when rotating the configurations and the geometry optimizations were stopped as soon as 33% of the force consisted of noise. For both functionals this happened when the largest force on any atom was approximately 5×10^{-5} Hartree/Bohr. In order to assess the accuracy of the BigDFT pseudopotential calculations, the energetic ordering of several relaxed configurations were compared with results obtained by means of the FHI-aims code.[244] The basis set used in the FHI-aims calculations was the extremely well converged 'tier 2'[244] level with a large confinement radius (onset: 6 Å). Energies agreed within less than 3 meV per atom showing that the pseudopotentials in the BigDFT calculations were highly accurate.

In total, the above procedure resulted in roughly 900 distinct local minima for each of the four sets.

4.1.2 Computation of Transition States

If bond-breaking takes place, it is well known that the energies of transition states from DFT calculations using conventional exchange-correlation functionals are usually of poor quality.[245–247] For transitions involving only minor changes in the number of bonds, it has been argued by Zupan et al.[248] that there is little difference between GGA and LDA barriers and thus both GGA and LDA should perform equally well in this case. Ghasemi et al.[109] confirmed this by comparing the transition state energies of Si_8 resulting from LDA, PBE and B3LYP calculations to diffusion Monte Carlo calculations. They found that PBE and B3LYP are outperformed by the LDA functional. This result was explained by the fact that in contrast to the atoms of transition states of conventional chemical reactions those of Si_8 are in a similar environment as the atoms of local minima. As a consequence the transition states of Si_8 are difficult to distinguish from local minimum configurations by visual inspection. They concluded that DFT self interaction errors are expected to cancel to a large degree and highly inhomogeneous environments with large density gradients are not relevant for the calculation of the transitions states of the Si_8 cluster. For the current Au_{26}^- cluster, the situation is similar. In most cases only a local rearrangement of atoms takes place when going over a transition state from one minimum to another. On basis of visual inspection, it is almost impossible to distinguish a local minimum from a transition state. Additionally, in contrast to the PBE functional, the LDA functional is able to predict the experimentally observed Au_{26}^- structures as low-energy minima and the LDA energies correlate very well with the energies obtained from the highly accurate, but computationally more demanding, M06 meta-hybrid functional (see Table 4.1).[249] Furthermore LDA is believed to describe the gold metallic bonds better than the PBE functional.[250] For fcc gold it was shown that properties like the lattice constant, phonon dispersion and the equation of state are reproduced more accurately by the LDA functional than by PBE.[251–254] For these reasons, the LDA functional is expected to give reasonable transition state energies for the Au_{26}^- system. Therefore, this functional was used for the transition state search.

The transition state search was performed for Au_{26}^- using the bar-saddle method.[62] This method efficiently identifies transition states located in between two input configurations. The transition state search was started in the vicinity of the experimentally identified Au_{26}^- clusters. The initial input configurations were chosen by searching among all found local minima for structures that are close to the experimentally identified structures. As a distance measure the permutationally optimized root-mean-square displacement (RMSD) was used, which is briefly recapitulated in Appx. B.[167] The transition states found in this way do not need to be connected directly to the input minima, so if required, the transition state search had to be repeated recursively until the two input minima were connected. In order to decide whether two minima are directly connected to a transition state, a small step in the forward and backward direction of the negative mode at the transition state followed by a local geometry optimization was performed. The minima and transition states that emerged during the above connection attempts form a stationary point database which is visualized as a disconnectivity graph[73] using the disconnectionDPS[129] software. For minima which seemed

to lie behind high energy barriers, permutationally optimized RMSDs to all the other minima in the stationary point database were computed and connection attempts to the closest structures were performed. This procedure was iterated until 264 transition states connecting 118 minima were found. Due to the high computational cost of computing transition states, it was not possible to compute a significantly larger stationary point database. The size of the database is not large enough to ensure that the lowest lying paths among all of its minima have been found. However, for the purpose of finding upper bounds on the energy along transition paths, a fully converged disconnectivity graph is not necessary. Furthermore, during the addition of the last 50 transition states to the stationary point database, the disconnectivity graph did not show any significant changes.

On average, the relaxation of a transition state was stopped as soon as the largest component of the force acting on any atom was approximately 2×10^{-4} Hartree/Bohr.

4.1.3 Computation of Photoelectron Spectra

The electronic density of states (DOS) for several by the MH algorithm generated low-lying isomers of Au_{26}^- (typically within ~ 1.0eV of the lowest-energy isomer) were computed by members of the group of Prof. Xiao Cheng Zeng and were compared with the experimental photoelectron spectra. Single-point energy calculations of these low-lying structures were performed at the PBE0/CRENBL level of theory with the inclusion of spin-orbit (SO) effects as implemented in the NWChem 6.1.1 package.[255] Previous reports have shown that the inclusion of the SO-effects yields almost quantitative agreement between the experimental photoelectron spectra and computed DOS for gold clusters of various sizes and shapes.[235] The first vertical detachment energies (VDEs) of each isomer were calculated as the difference between the energies of the anionic and the corresponding neutral species at the anion geometry. The binding energies of the deeper occupied orbitals of the anion were added to the first VDE to approximate higher binding energy features. Each computed peak was fitted with Gaussian functions of 0.06 eV width to yield a computed photoelectron spectrum, which was used for the comparison with the experimental spectra of Au_{26}^-.

4.1.4 Experimental Methods

The photoelectron spectroscopy experiment described in this section was performed by the group of Prof. Lai-Sheng Wang. They used a magnetic-bottle apparatus equipped with a laser vaporization supersonic cluster source and a time-of-flight mass analyzer.[256] A pulsed laser beam was focused onto a pure gold disk target, generating a plasma containing gold atoms. Simultaneously, a pulse of high-pressure helium carrier gas was delivered to the nozzle, with the effect of cooling the plasma and initiating the nucleation. It was previously show by Akola et al.,[257] that by carefully controlling the residence time of the clusters in the nozzle, relatively cold and equilibrated clusters can be produced from the laser vaporization supersonic cluster source. The cooling effects have been confirmed recently by the observation of van der

Waals complexes of gold cluster anions with Ar or O_2.[229,230,258] In the present study, relatively cold Au_{26}^- clusters were produced using a helium carrier gas seeded with 5% Ar. In addition, Prof. Wang's group was able to produce even colder Au_{26}^- complexed with Ar atoms, $Ar_nAu_{26}^-$ ($n = 1,2$). The Au_{26}^- and $Ar_nAu_{26}^-$ clusters were selected by a mass gate and decelerated before being photo-detached by a 193 nm laser beam from an ArF excimer laser. Photoelectrons were collected with a magnetic bottle at nearly 100% efficiency in a 3.5-m-long electron flight tube for kinetic energy analyses. The photoelectron kinetic energies were calibrated by the known spectrum of Au^- and subtracted from the photon energy to obtain the here reported electron binding energy spectra. The electron energy resolution was $\Delta E/E \approx 2.5\%$ (i.e., 25 meV for 1 eV electrons).

4.2 Results and Discussions

4.2.1 Energy Landscape and Exchange-Correlation Functionals

Fig. 4.2 and Fig. 4.3 show the energy spectra of local minima that were found during the MH runs within 0.55 eV of the respective global minima. By means of visual inspection, four structural motifs were identified: empty cages, cages filled with a single atom, the tubular cage and hexagonal cages. The neutral and anionic systems possess the same motifs. Representatives for each motif can be found in Fig. 4.1. As can bee seen from the energy spectra, different functionals yield different energetic ordering for the structural motifs.

In the case of the anionic system a closer look at this circumstance was taken. Fig. 4.4a shows the energies of the five lowest-energy configurations of each motif as obtained by the PBE functional. Additionally, each configuration was relaxed using the LDA functional and the energies of corresponding configurations have been connected by lines. In case of the tubular motif only one representative could be identified and consequently only this single tube structure is shown. Fig. 4.4a shows a significant energetic reordering of the motifs using different functionals. The LDA functional favors filled cages, whereas according to the PBE functional the same motif possesses a much higher energy. Fig. 4.4b shows a more detailed plot of the energetic reordering of 25 configurations within the filled cage motif. Although both functionals produce a different energetic ordering, the overall ordering is conserved within the filled cage motif. This is in strong contrast to the energetic reordering of motifs.

In Fig. 4.5 the potential energies of the identified local minima are plotted versus the permutationally optimized RMSD.[167] Both quantities are measured with reference to the respective putative global minimum. In particular when compared to systems like for example C_{60} or $B_{16}N_{16}$, the here investigated Au_{26} and Au_{26}^- possess a vast number of structurally diverse minima within a small energy window above the putative global minimum.[66,260] This can also be seen from the energy spectra given in the Fig. 4.2 and Fig. 4.3. Using a Rosato-Guillopé-Legrand potential, Bao et al.[198] previously found similar results for large gold clusters. Considering the predictions of structurally different minima within a small energy range above the ground

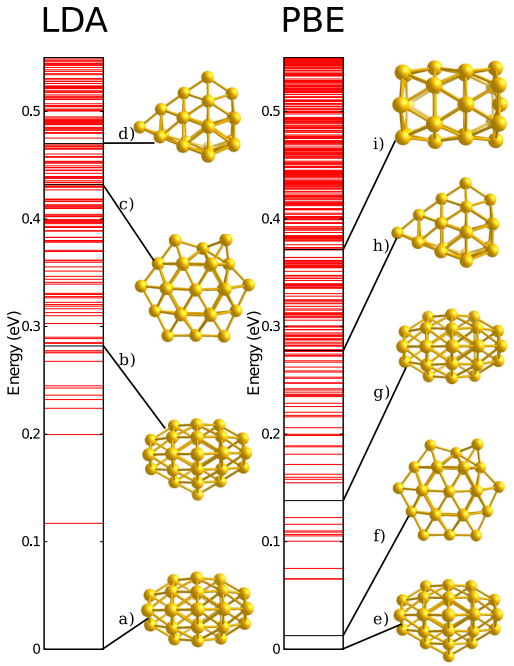

Figure 4.2: Energy spectra of the neutral Au_{26} based on the LDA and PBE functional. Shown are all minima with energy ≤ 0.55 eV. Depicted isomers are highlighted using black lines in the spectra and when neglecting minor changes in the bond length they are identical to their respective geometric counterparts of the other functional. Isomers (a), (b), (c) and (e), (f), (g) are the energetically lowest representatives that could be found for each of their structural motifs. Isomer (a) [(g)] is a filled cage. According to the LDA functional this filled cage is the putative global minimum. The PBE functional predicts isomer (e) [(b)], an empty cage, to be the putative global minimum. Isomers (d) [(h)] (pyramid) and (i) (tube) are previously claimed global minima[243,259] of the neutral Au_{26} cluster. The position of the tubular structure in the LDA spectrum is outside of the shown energy range. Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

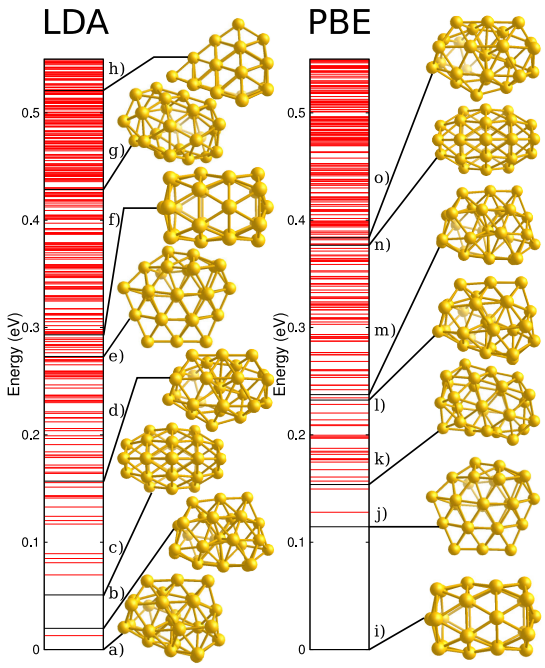

Figure 4.3: Same as Fig. 4.2 but for the Au_{26}^- cluster. The computed photoelectron spectra of the isomers (b), (c) and (d) [(m), (n) and (o)] can explain the experimentally measured photoelectron spectrum very well (see below). Isomers (a), (e), (f), (g) and (i), (j), (k), (l) are the energetically lowest representatives that could be found for each of their structural motifs. Isomers (a) and (e) [(l) and (j)] are the energetically lowest singly filled cage and hexagonal structures, respectively. Isomers (g) and (k) are the lowest empty cage structures; geometrically they are not identical. Isomers (f) [(i)] (tube) and (h) (pyramid) correspond to previously claimed global minima[243,259] of the neutral Au_{26} cluster. The position of the pyramidal structure in the PBE spectrum is outside of the shown energy range. Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

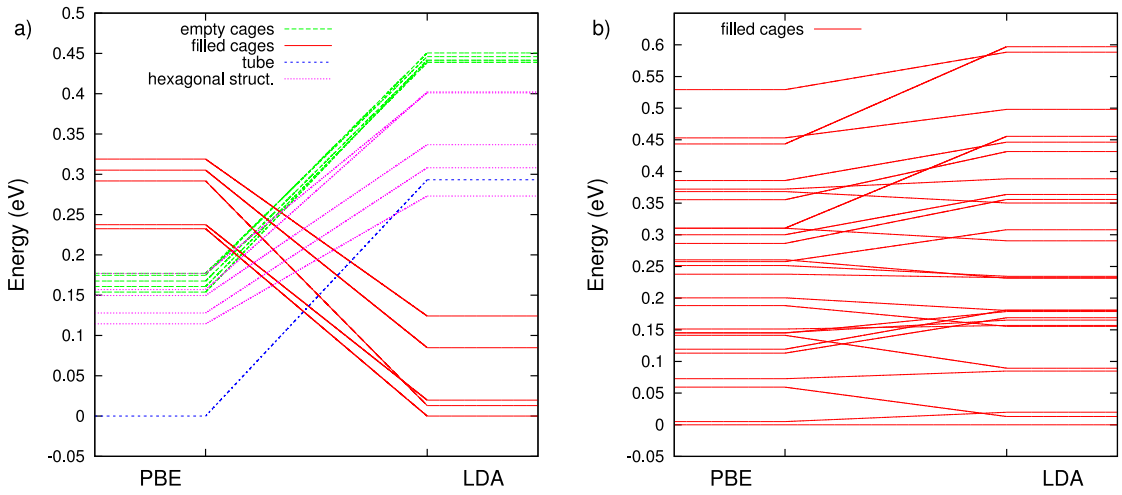

Figure 4.4: Energetic reordering of the Au_{26}^- minima using the PBE and LDA functionals. PBE and LDA energies of identical isomers are connected by lines. A significant reordering of motifs can be observed (a), whereas the overall energetic ordering within one motif is conserved (b). The energies in panel (a) are shifted with respect to the putative global minimum of each functional whereas in panel (b) the energies are shifted such that the energy of the lowest-energy filled cage isomer is zeroed. Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

state by the two different exchange-correlation functionals and the Rosato-Guillopé-Legrand potential, it is conceivable that multiple isomers can also be observed experimentally. Indeed, numerous isomeric forms have been observed for some small gold cluster anions previously.[233–236,258] In the case of Au_{10}^-, at least three low-lying isomers were observed experimentally beside the global minimum.[258]

The vast number of structurally diverse minima that can be found in a comparably small energy range above the putative global minimum in conjunction with an exchange-correlation-functional dependent energetic ordering of different structural motifs make Au_{26} and Au_{26}^- demanding systems for structure prediction: On the one hand one cannot be sure to use the right exchange-correlation functional, on the other hand, more than one minimum may contribute to experimental results. Thus, when trying to identify configurations observed in experiment, it is advisable to introduce different structural motifs, if possible. Attention then should be focused on the first few energetically lowest configurations of each motif, instead of only the lowest energy structures. The complexity of the Au_{26}^- cluster was the major reason why it was omitted in previous joint experimental and theoretical studies of medium-sized gold clusters.[225,235]

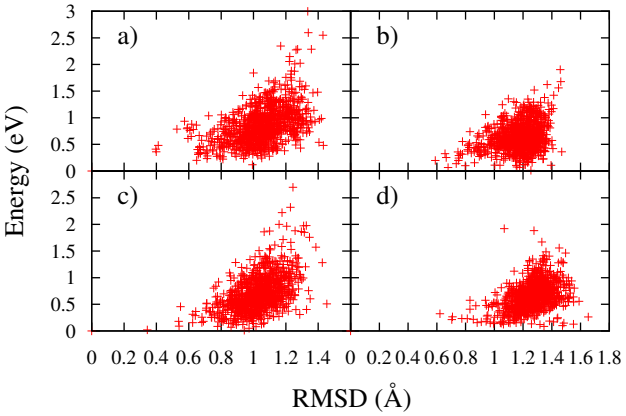

Figure 4.5: Potential energy versus permutationally optimized RMSD[167], both measured with respect to the corresponding calculated lowest energy configuration. The upper plots show data for the neutral Au_{26} cluster using LDA (a) and PBE (b) functionals. The lower plots show data for the Au_{26}^- anion using LDA (c) and PBE (d) functionals.Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

4.2.2 Computationally Predicted Low-Energy Configurations

Neutral Cluster

Next to the energy spectra of the neutral Au_{26} system (Fig. 4.2) several specific configurations are visualized. Configurations (d) [(h)] (pyramidal structure) and (i) (tubular structure) have previously been proposed to be the global minimum[243,259] of Au_{26}. According to the LDA functional the tubular structure is also a local minimum. However its energy ($\sim$ 904 meV) is outside the energy range of Fig. 4.2. The pyramidal structure can be obtained from the Au_{20} global minimum pyramid[221] by adding six atoms to one face of the Au_{20} pyramid.[259] As shown by both the LDA and the PBE exchange correlation functional, a large number of configurations that are significantly lower in energy than the pyramidal configuration were found, even though the LDA and PBE functionals predict different sets of low-energy configurations. Among the 12 energetically lowest LDA configurations ($\lesssim$ 278 meV) only filled cages consisting of a single core atom surrounded by a 25-atom shell can be found. In contrast to this, empty cages and hexagonal structures are found among the energetically lowest structures of the PBE calculations.

The putative global minimum based on LDA calculations (Fig. 4.2a) possesses C_{2v} symmetry and can be obtained from the pyramidal Au_{26} structure by removing the leftmost, uppermost and lowermost (with respect to the illustration given in Fig. 4.2d) corner atoms of the Au_{20} pyramidal part and attaching them to the right side of the illustration in Fig. 4.2d. Another illustration of this isomer is given by Fig. 4.1a. Only two further configurations (filled cages having C_1 and C_2 symmetry) were found in the energy region $\leq$ 200 meV.

The global minimum predicted by the PBE functional (Fig. 4.2e) is a C_{2v} empty cage and can be

constructed from configuration (a) [or (g)] of Fig. 4.2 by removing the core atom and attaching it to the lowermost part of Fig. 4.2a. Fig. 4.1b shows this configuration from a different angle. In the region below 200 meV the PBE spectrum is considerably denser compared to the LDA spectrum. The first low-lying isomer (Fig. 4.2f) is only 13 meV above the putative global PBE minimum. This structure is an isomer of the hexagonal motif with C_s symmetry. It is structurally considerably different from the putative global PBE minimum.

As mentioned above, not only the putative global minima (configurations (a) and (e) of Fig. 4.2) should be taken into consideration when trying to identify the isomers experimentally. Instead, the first few energetically lowest configurations of each of the motifs should be compared with future experimental results. Hence, the coordinates of all configurations that can be found within 150 meV above the energetically lowest representative of each motif (but not less than 5 isomers) are provided in Appx. E. The structures given in Appx. E are sorted in ascending order with respect to their energy. For the sake of completeness, also the pyramidal and the tubular structure are given in Appx. E.

Singly Charged Anion Cluster

Just as in the neutral case, the PBE and LDA functionals predict low-energy configurations that belong to different structural motifs for the anionic Au_{26}^- cluster (Fig. 4.3).

At the LDA level, only cages filled with a single atom are competing for the global minimum. The first non-filled cage (Fig. 4.3e) can be found at an energy of 273 meV. This isomer is of C_1 symmetry, but the hexagonal motif still can be recognized. The filled cage LDA global minimum (Fig. 4.3a) also possesses C_1 symmetry.

The energetically lowest configuration predicted by the PBE functional is the tubular structure (isomer Fig. 4.3i) with D_{6d} symmetry. The first low-lying isomer (Fig. 4.3j) is identical to isomer (e) at the LDA level with only minor changes in bond lengths. In the energy region above the two low-lying isomers and below 232 meV, only empty cages and hexagonal structures are found. The putative global minimum at the LDA level (Fig. 4.3a) is approximately 232 meV higher in energy at the PBE level (Fig. 4.3l). Nevertheless, it is still the energetically lowest filled cage that was found.

As will be discussed in detail in the subsequent sections, the computed photoelectron spectra of the structures (b), (c) and (d) [(m),(n) and (o)] match very well with the experimentally observed photoelectron spectra. According to the LDA energy ordering, structure (b) is the third isomer, (c) is the fourth isomer and (d) is the eighteenth isomer above the LDA global minimum. Henceforth these structures will be denoted as isomer 3, isomer 4 and isomer 18. The coordinates of the three identified structures are given in Appx. F.

In order to further assess the low-energy nature of isomers 3, 4 and 18 and the small energy window they can be found in, the energies of low-lying LDA isomers were re-evaluated at the SO-PBE0/CRENBL and M06/cc-pVDZ levels of theory as implemented in the NWChem

Iso.	A	B	C	Iso.	A	B	C
1	0.00	0.08	0.00	17	0.16	0.26	0.19
2	0.01	0.15	0.07	18	0.16	0.32	0.14
3	0.02	0.00	0.01	19	0.16	0.20	0.17
4	0.05	0.14	0.15	20	0.17	0.26	0.20
5	0.07	0.33	0.22	65	0.27	0.17	0.35
6	0.08	0.29	0.16	74	0.29	0.21	0.21
7	0.08	0.19	0.06	84	0.31	0.16	0.37
8	0.09	0.26	0.15	101	0.34	0.18	0.37
9	0.12	0.17	0.08	106	0.34	0.32	–
10	0.12	0.17	0.21	126	0.37	0.24	0.42
11	0.12	0.09	0.08	129	0.37	0.23	0.43
12	0.13	0.32	0.26	175	0.43	0.44	–
13	0.14	0.23	0.18	178	0.44	0.34	–
14	0.14	0.19	0.13	185	0.44	0.35	–
15	0.14	0.05	0.22	187	0.44	0.41	–
16	0.15	0.31	0.24				

Table 4.1: Relative energies (in eV) of Au_{26}^- isomers of all the motifs. The columns labeled with 'Iso.' give the isomer number, which follows the LDA ranking of Fig. 4.3. Column A shows LDA energies of geometries relaxed at the LDA level (same as in Fig. 4.3), column B shows SO-PBE0/CRENBL energies of geometries relaxed at the LDA level and column C shows M06/cc-pVDZ energies of geometries relaxed at the M06/cc-pVDZ level.

6.1.1 package.[255] The PBE0 functional (hybrid GGA[107]) and the M06 functional (meta-hybrid GGA[249]) are from higher rungs on 'Jacob's ladder'[261] than the LDA and PBE functionals and thus are expected to give a good energy ranking. In particular the M06 and M06-L functionals have previously been shown to be accurate for gold clusters.[262,263] Table 4.1 shows the PBE0 and M06 energies in eV of the low-lying isomers together with their LDA energies (isomers 1-20 being core-shell structures with a single internal atom, isomers 65, 84, 101, 106, 126 being hexagonal, isomer 74 being tubular, and isomers 129, 175, 178, 185, 187 being empty cages). The LDA energies are identical to those shown in Fig. 4.3. The relative energies of the core-shell isomers 3 and 4 are consistently found to be very low (within 0.15 eV) at all three levels of calculation. The energy of isomer 18 was found to be ~0.3 eV at PBE0, but ~0.14 eV using the M06 functional which corresponds well with the ~0.16 eV predicted by the LDA functional. The relative energy of the previously proposed tubular isomer 74 was consistently found to be more than 0.2 eV higher than the putative global minimum. Furthermore the predictions for the energetic ordering of the different motifs at the LDA, PBE0 and M06 levels of theory are found to be in good agreement with each other. It is worth to emphasize that the PBE functional, which usually gives better atomization energies than the LDA functional, is not able to identify the correct structural motif.

4.2.3 Experimental Photoelectron Spectra

The experimental photoelectron spectrum of Au_{26}^- is shown in Fig. 4.6a (magenta color). Numerous well-resolved photoelectron bands are observed below ~5.4 eV binding energies, which should come mainly from Au6s orbitals, whereas the more intense and almost continuous features above 5.4 eV should be due to the 5d band, according to previous photoelectron spectroscopy studies.[225,235,264] Neutral Au_{26} is expected to be closed shell with a gap between its highest occupied (HOMO) and lowest unoccupied (LUMO) orbitals. The X band with a VDE of 3.46 ± 0.03 eV should correspond to electron detachment from the extra electron in Au_{26}^- that occupies the LUMO. The X' band with a VDE of 3.75 ± 0.03 eV should correspond to electron detachment from the HOMO. However, the intensity of the X' band is comparable to that of band X and it seems too low to be from the HOMO. Instead, the X" band with a VDE of 4.04 ± 0.03 eV appears to be the HOMO, suggesting that the X' band should come from a different isomer of Au_{26}^- populated in the cluster beam.

Experimental evidence for this conjecture is provided by the photoelectron spectrum of the Ar-tagged van der Waals complex $ArAu_{26}^-$ (black curve in Fig. 4.6a), which should be in a colder condition than the bare Au_{26}^-. If normalized to the X band, the relative intensity of the X' band and in fact all the higher binding energy bands seem to decrease under cold conditions. Also the photoelectron spectrum of $Ar_2Au_{26}^-$ was measured (not shown), which is similar to that of $ArAu_{26}^-$. This change of relative intensity between the X and X' bands suggests that they come from two different isomers; and the X' should come from a slightly higher free energy isomer so that its relative intensity is reduced at lower temperatures. Hence, the higher binding energy bands must be a mixture of detachment features from the two isomers. As will be seen below in comparison with the simulated DOS, bands X", A", and B" also have contributions from a third isomer, whose first VDE contributes to band X". The presence of at least three isomers experimentally for Au_{26}^- is consistent with the high density of low-lying isomers predicted computationally, making it an extremely challenging system to interpret.

4.2.4 Simulated Photoelectron Spectra of Low-Energy Isomers and Comparison with Experimental Spectrum

For each motif of Au_{26}^-, the photoelectron spectra of several low-lying LDA geometries were simulated at the PBE0/CRENBL level of theory, including spin-orbit (SO) effects. At present there exists no mathematical rigorous and systematic method that would allow to quantify the similarity of two spectra and, therefore, the best-matching spectra had to be chosen among all possible matches based on visual inspection and experience. As described above, there is concrete evidence for the existence of multiple isomers in the experimental spectra. Indeed, it was found that no single isomer could fit all the observed features of the photoelectron spectra. As shown in Fig. 4.6b, the combination of isomer 3 (blue), isomer 4 (green), and isomer 18 (red) was found to reasonably reproduce the experimental spectrum. Isomer 4 gives a very low first VDE of 3.2 eV, which is in good agreement with band X, but lower by ~0.2 eV

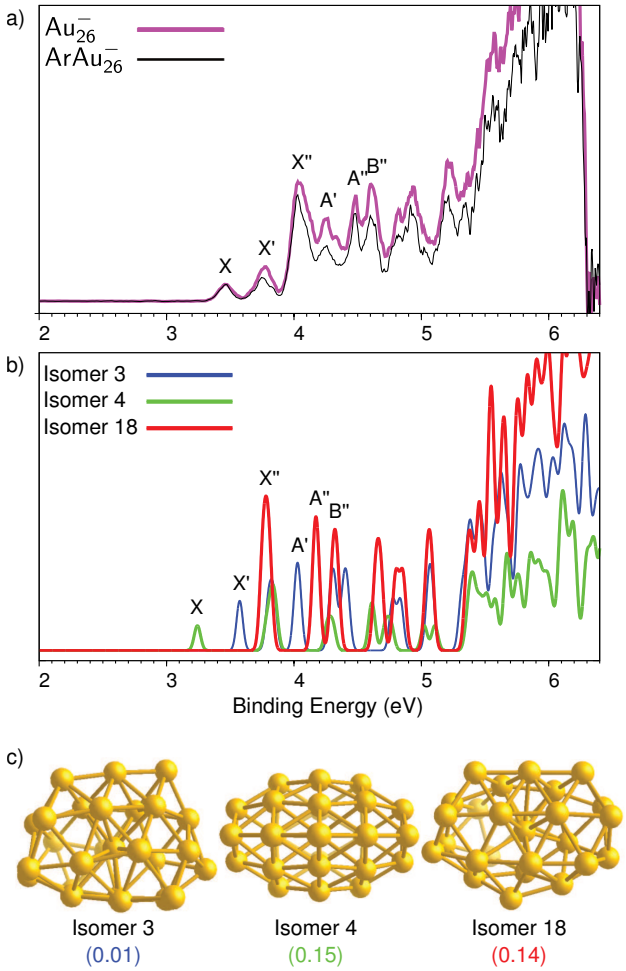

Figure 4.6: a) Experimental photoelectron spectra of Au_{26}^- (magenta) and $ArAu_{26}^-$ (black) at 193 nm (6.424 eV) photon energy; b) simulated photoelectroni spectra from three isomers (3,4 and 18), that best match the experimental spectra; c) structures of the three isomers (all having core-shell structures with a single internal atom) and their relative energies (in eV) calculated at the M06/cc-pVDZ level of theory. The energies are shifted with respect to the computationally lowest-energy isomer at the M06/cc-pVDZ level of theory (see Table 4.1). Isomer 3 is identical to Fig.4.3b, isomer 4 is identical to Fig. 4.3c and isomer 18 is identical to Fig. 4.3d. Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

relative to the observed VDE of 3.46 eV. This difference is within the error of the calculated VDEs that were observed in previous studies.[225,233–236] The second band of isomer 4 is around 3.8 eV, giving rise to a HOMO-LUMO gap of ~0.6 eV, similar to the gap between bands X" and X, suggesting isomer 4 contributes to band X". Higher binding energy features from isomer 4 are all overlapped with detachment features from other isomers. The first VDE of isomer 3 is at ~3.55 eV, in good agreement with that of band X', but again lower than the experimental VDE of band X' by ~0.2 eV. The third band of isomer 3 gives a VDE of ~4 eV, giving rise to

an energy difference of ~0.5 eV between this third and the first band. This energy difference is in close agreement with the gap between bands A' and X'. It is noticed that the X" band is quite strong and it cannot be fully accounted for by the second detachment band from isomer 3 and isomer 4 alone. Isomer 18 was found to give a high first VDE at ~3.8 eV, which is lower than the observed VDE of band X" by ~0.2 eV and could be a major contributor to this band. In fact, the second and third detachment transitions of isomer 18 are in good agreement with bands A" and B". Hence, only with the three isomers a good interpretation of the observed photoelectron spectra for Au_{26}^- can be obtained. The temperature dependence of the photoelectron spectra suggests that isomer 4, which corresponds to band X, should be the lowest in free energy, whereas isomers 3 and 18 should be slightly higher in free energy because the relative intensities of their bands decreased when the cluster was colder. Theoretically (see Table 4.1) isomer 3 is predicted to be slightly lower in potential energy than isomer 4. However due to low-temperature entropy effects and inaccuracies introduced by the exchange-correlation functional, a perfect one-to-one match between the experimentally observed (low temperature) free energy ordering and the theoretically computed (zero temperature) potential energy ordering cannot be expected.

4.2.5 Fluxional Character of Au_{26}^-

In order to estimate the transition rate out of a minimum across a transition state along a single reaction path, Eyring's transition state theory[130] can be used. In this theory the transition rate k_{mt} at temperature T out of a minimum m with energy E_m over a transition state t with energy E_t is given by

$$k_{mt} = \left(\frac{k_B T}{h}\right)\left(\frac{\overline{q_t}}{q_m}\right)\exp\left(-\frac{E_t - E_m}{k_B T}\right),$$

where $\overline{q_t}$ is the partition function of the transition state for coordinates normal to the reaction coordinate, q_m is the partition function of the minimum, k_B is the Boltzmann constant and h is Planck's constant. For the calculation of the following rates, it was assumed that the ratio of the partition functions in the above formula can be neglected since rough order-of-magnitude estimates are sufficient for the present purpose. At room temperature the transition rate across a barrier with an energy of 0.33 eV measured with respect to the corresponding minimum is roughly on the order of $10^7 s^{-1}$ and across a barrier of 0.13 eV roughly on the order of $10^{10} s^{-1}$. At a temperature of 200 K the transition rate across the same barriers are roughly on the order of $10^4 s^{-1}$ and $10^9 s^{-1}$, respectively.

Fig. 4.7 shows a disconnectivity graph of Au_{26}^-. Two of the experimentally matched structures (isomer 3 and 18) are located in one of two different funnels which merge in Fig. 4.7 at the low energy of 0.33 eV. Due to only limited data available for this graph, the existence of transition states which may merge these funnels at lower energies cannot be ruled out. Nevertheless, the transition states shown give an upper bound on the energetic height of existing transition pathways and thus allow to estimate whether transitions out of a minimum can be observed on experimental time scales.

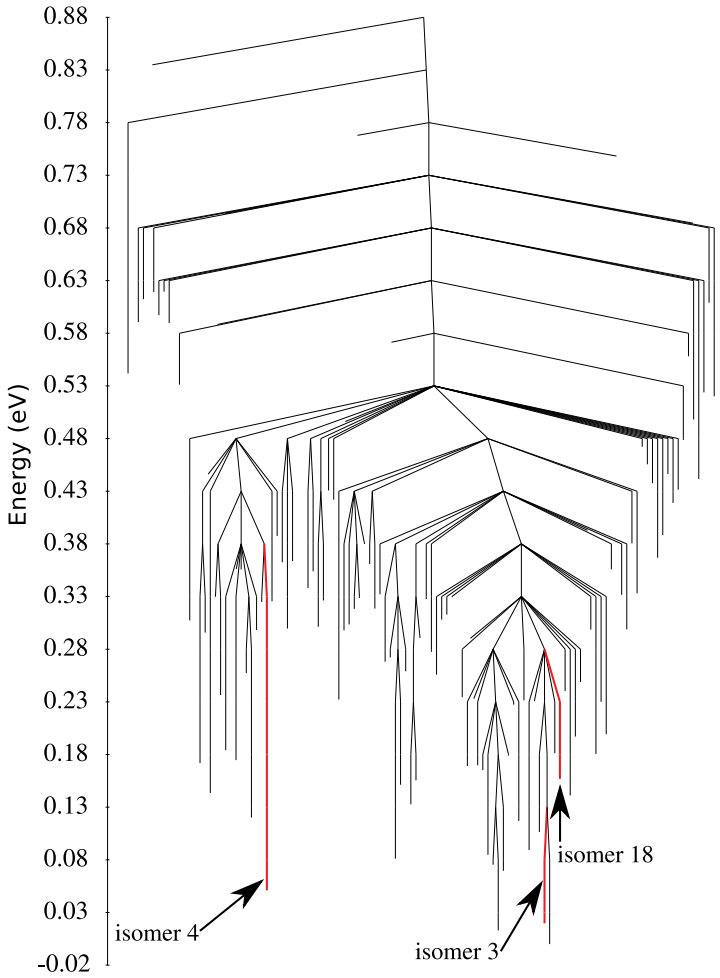

Figure 4.7: Disconnectivity graph of Au_{26}^- computed at the LDA level of theory. Energies are measured with respect to the putative global minimum (LDA). Reprinted with permission from Ref. [61]. Copyright 2014 by the American Chemical Society.

All the minima contained in the two funnels mentioned above can be interconverted into each other by crossing barriers not higher than 0.33 eV. For isomers 3 and 18 there even exist reaction paths to other minima with barriers lower than 0.13 eV. One, therefore, might expect that, in addition to isomers 3, 4 and 18, further isomers could exist in experiments. Indeed, this possibility cannot be excluded, as features of additional isomers might be buried under the strong peaks of the photoelectron spectra of isomers 3, 4 and 18.

All structures in the stationary point database used to generate Fig. 4.7 are core-shell structures with a single internal atom and thus mainly atoms located in the shells of these structures are taking part in the just mentioned transitions. In this sense the shell of Au_{26}^- flows around the core atom and the cluster can be considered to be fluxional. In a previous[204] study it has been reasoned by means of a MD simulation that Au_{34}^- is fluxional too. The fluxional property of clusters may promote catalytic, in particular chemisorptional, activities.[204]

4.3 Conclusion

Based on a thorough ab-initio exploration of the energy landscapes of Au_{26} and Au_{26}^-, it was found that these systems possess a variety of structurally different but energetically similar minima. Many of the found structures are significantly lower in energy than previously suggested global minimum candidates, showing the importance of an unbiased global minimum search. Based on the analysis of the energy landscape and energetic reordering between the LDA and PBE functional a set of new configurations for Au_{26} could be proposed. The configurations in this set are intended to be compared with future experimental results. Compared to systems with experimentally observable unique ground states,[260] both gold systems possess a large number of metastable structures within a small energy window above their computational putative global minima. Therefore, it is likely that a number of isomers can be found to co-exist experimentally. By comparing the simulated photoelectron spectra of a wide variety of isomers of Au_{26}^- with the experimental photoelectron spectra at different conditions, it was possible to identify three structures which can reasonably explain the experimental data. On the basis of a transition state search it was concluded that Au_{26}^- may be a fluxional cluster system.

5 Computationally Inexpensive Post-Processing of Minima Hopping Data for a Qualitative Characterization of Potential Energy Surfaces*

As already has been stated several times, thermodynamic and kinetic properties of multi-atomic systems are encoded in the topology of their potential energy surfaces (PES). For example, the folding of a protein into its native state seems to be impossible based on the sheer abundance of conformational possibilities (Levinthal's paradox).[265] However, assuming a steep, funnel-like shape of the PES introduces driving forces that necessarily lead the system towards its stable configuration, independent of its initial denaturated structure.[266] In contrast, multi funnel PES can explain why a certain system might be observed in a metastable state and glass formation can be identified with trapping in some disordered state.[3] Accurately assessing the shape of a PES usually requires not only the computation of local minima, but also the network of possible transitions and the corresponding energy barriers.

There exists various methods such as transition path sampling (TPS),[26,27,180–184] discrete path sampling (DPS),[38,39] stochastic surface walking based reaction sampling (SSW-RS),[185] or the minima hopping guided path sampling (MHGPS) approach,[18,62,64] that allow the rigorous sampling of reactive processes, some of which even at sophisticated levels of theory, like at the density functional level. Nevertheless, these methods are computationally very demanding, typically even more costly than the already challenging global optimization problem.

Therefore, computationally lightweight methods that allow to receive at least a qualitative impression of a PES are of high interest. To this end distance-energy (DE) plots that allow to distinguish glassy from non-glassy systems have been introduced recently.[66] In a DE plot

*The distance-energy plots that are briefly recapitulated in the introduction to this chapter have been published in S. De, B. Schaefer, A. Sadeghi, M. Sicher, D. G. Kanhere, and S. Goedecker, "Relation between the Dynamics of Glassy Clusters and Characteristic Features of their Energy Landscape", Physical Review Letters **112**, 083401 (2014)

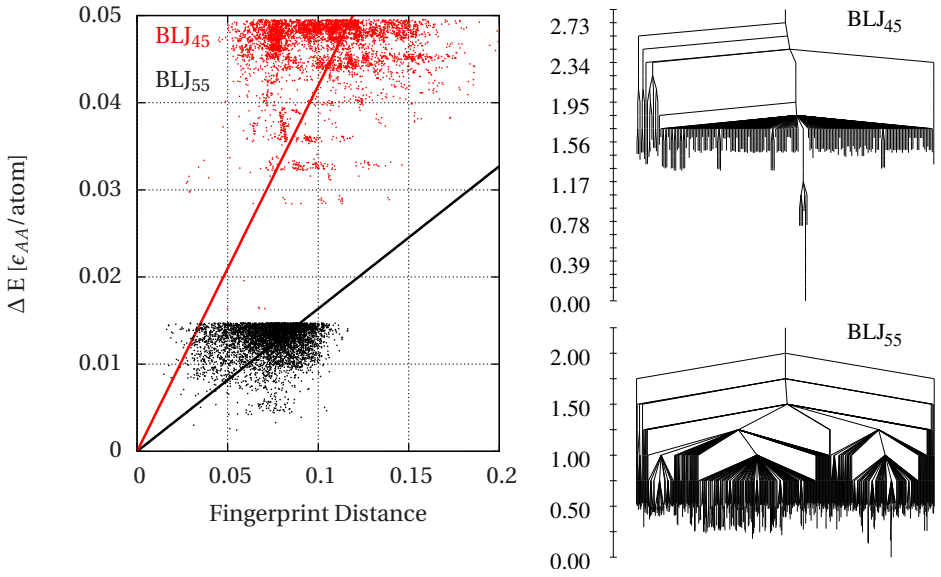

Figure 5.1: DE plots (left-hand side) and disconnectivity graphs (right-hand side) for BLJ clusters with 45 and 55 atoms. The BLJ-parameters for the BLJ_{45} and BLJ_{55} clusters are $\epsilon_{AB} = \epsilon_{AA} = \epsilon_{BB} = 0.25, \sigma_{AA} = 1.0, \sigma_{BB} = 1.3, \sigma_{AB} = 1.15$ and $\epsilon_{AA} = 0.25, \epsilon_{BB} = 0.125, \epsilon_{AB} = 0.275, \sigma_{AA} = 1.0, \sigma_{BB} = 0.88, \sigma_{AB} = 0.80$, respectively. For the disconnectivity graphs the energy is given in units of ϵ_{AA} and two sets, each of roughly 500,000 saddle points, along with the minima they connect were used for their respective construction. For the DE-plots only the 5,000 lowest minima were considered. For BLJ systems, a covalent radius is not well defined. Therefore, the following slightly modified overlap matrix had to be used for the calculation of the eigenvalues that define the structural fingerprints: $O_{ij} := \exp(-r_{i,j}^2/(2\sigma_{i,j}^2))$, where $r_{i,j}$ is the distance between atom i and j and $\sigma_{i,j}$ is the parameter of the Lennard-Jones potential. The solid lines in the DE plots show least-square fits of linear functions to the two data sets. Their slope is a measure for the average driving force towards the ground state. Reprinted with permission from Ref. [66]. Copyright 2014 by the American Physical Society.

the (atomization) energies per atom of metastable configurations are measured relatively to the global minimum and they are plotted versus their configurational distance to the global minimum. The recently proposed structural fingerprints, which are based on the overlap matrix of Gaussian type orbitals, can be used for measuring the configurational distances, if a covalent radius is well defined for all atomic species in a given system.[66,167] Those fingerprints are briefly recapitulated in Appx. B. Even on demanding levels of theory like DFT, it is computationally feasible to produce DE plots, because only the geometries and energies of a few hundred local minima, including the global minimum, are needed. The underlying idea of DE plots is the following. A strong driving force towards the global minimum can be expected if the global minimum is geometrically and energetically well defined. In other words, if both the energetic and structural distances from the global minimum to all other metastable minima are large, a system can be expected to exhibit a fast relaxation time towards its global minimum and, therefore, it can be considered to be a non-glassy system. On the other hand, glassy systems have a large number of energetically very similar minima

that are separated by barriers of various heights.[36] For this reason there is no strong driving force towards the global minimum. In DE plots, this small driving force is reflected by the existence of low energy minima, having both small and large structural distances to the global minimum. Fig. 5.1 displays DE plots and disconnectivity graphs for two binary Lennard-Jones (BLJ) systems having 45 atoms (13 of type *A* and 32 of type *B*) and 55 atoms (13 of type *A* and 42 of type *B*). Both from the DE plots and the disconnectivity graphs it is evident that BLJ_{45} is a non-glassy system and BLJ_{55} is a glassy system.

In contrast to the disconnectivity graphs of Becker and Karplus,[3,73] DE plots contain different and complementary information. DE plots focus on the relation of metastable configurations to the global minimum and display the density of the structures both with respect to energies and with respect to configurational distances. This allows the deduction of a measure for the driving force towards the global minimum, which is made clear by the least-square fits to the two data sets of the DE plot in Fig. 5.1. However, DE plots give no relation between two arbitrary minima and, therefore, cannot display topological information beyond the driving force towards the global minimum. This is a consequence of the very modest requirements of DE plots: only energies and geometries of the global minimum and a few hundred local minima are needed. There is no need for transition state energies or the information, which minima are connected with each other by only one intermediate transition state. However, in this chapter it will be demonstrated that, based only on the data obtained during conventional MH runs, an approximation to this connectivity information is available. Furthermore, it will be discussed that an empirical guess for the transition state energies can be obtained, which is based solely on fingerprint distances of local minima. The combination of the approximate connectivity information and the guess for the transition state energies allow to generate a new type of disconnectivity graph that shows a remarkable resemblance to disconnectivity graphs which are based on exact transition state energies and exact connectivity information. However, as will be discussed below, it is hoped to find a method for estimating transition state energies that is based on less empirical grounds. Nevertheless, already the empirical method presented in this chapter turned out to be astonishingly useful. The post-processing of the MH data for the generation of DE plots, for the extraction of the approximate connectivity information and for the computation of the transition state energy guess can conveniently be performed on a single core of a standard personal computer within a negligible amount of wall-clock time. Therefore, DE plots and the method presented below give a useful and very economical impression of the characteristics of a PES. They can serve as a valuable aid for making a decision if investing the computer time that is required for building a rigorous network of transitions and their corresponding barrier energies is worthwhile and expedient with respect to a certain research goal, or not. Furthermore, they provide a qualitative idea on the kinetics and thermodynamics of a system. Moreover, the method presented below will be demonstrated to be a promising tool for isolating physically reasonable intermediate metastable structures of complicated reactions, which, for example, might be used for generating initial pathways that are needed in methods like TPS or its discrete variant, DPS.

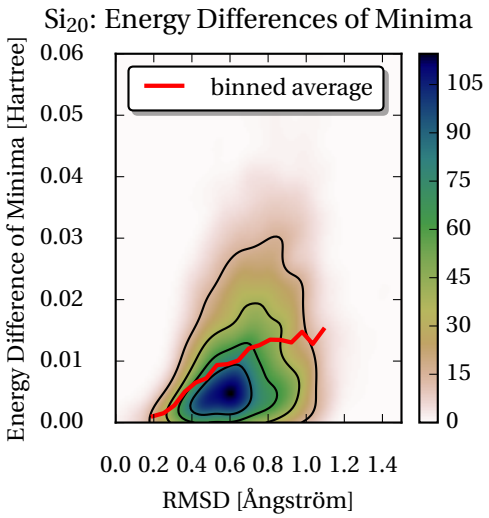

Figure 5.2: Density plot of the energy differences of pairs of minima versus their RMSD distance. The shown data sets consists of roughly 2900 minima pairs. Each pair of minima is connected by only one intermediate transition state. The structures, energies and the connectivity of the stationary points were determined at the DFT level of theory (PBE exchange correlation functional) by using the MHGPS method coupled to the BigDFT code.[62,64,148,149] The shown density was obtained from the corresponding scattered data by means of a Gaussian kernel density estimate as implemented in Python's scipy library. The red bold line shows the same data, but averaged within 25 bins along the RMSD axis. Only bins that contain at least 5% of the number of data points of the bin with the most data points are shown.

5.1 Correlating Transition State Energies with Structural Differences

Often the energies of two structurally similar minima are very close to each other, whereas the energy differences between structurally very different minima can be large. Nevertheless, it is clear that structurally very different minima can have very similar energies, as well. In other words, it is expected that for small structural differences the probability to find large energy differences is small, whereas for large structural differences, both, small and large energy differences between two adjacent minima are likely. Indeed, this expectation is supported by the data shown in Fig. 5.2. For the neutral silicon cluster consisting of 20 atoms, this figure shows the density of the distribution of energy differences of minima pairs plotted versus the corresponding chirally and permutationally optimized RMSD distance.[167] All minima pairs used for this plot are separated by only one intermediate transition state. It is seen from this plot, that for small RMSD values the density of the data points vanishes for large energy differences, whereas for larger RMSD values, there is a significant density, both for small and large energy differences. Because the variance in the energy differences is larger for increasing RMSD values, also the average values of the energy differences rises, as is shown by the binned average of the energy differences (red line).

Except for degenerate rearrangements, the barrier energy of every transition state can be

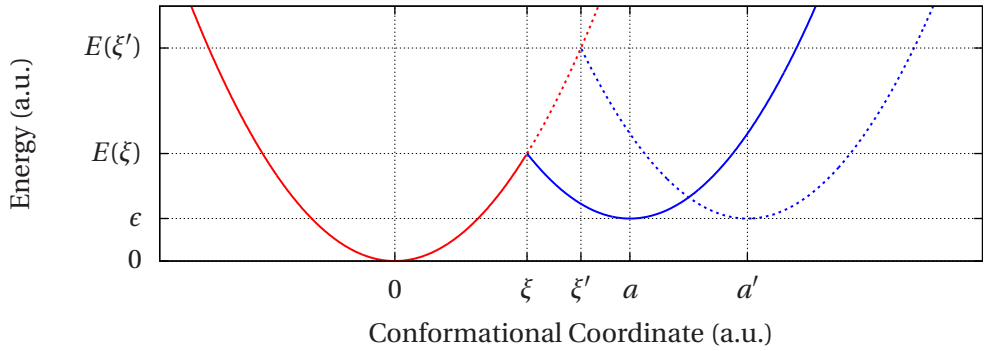

Figure 5.3: Parabola model for the transition state energy. For increasing structural differences of both minima the transition state energy is rising. Here this is illustrated by shifting the minimum of the solid blue parabola from a to a'. The sifted parabola is visualized by a blue dashed line.

measured with respect to the lower or the higher energy minimum to which the transition state is connected to. In contrast to the distribution of the energy differences of neighboring minima in an energy difference versus RMSD plot, it can be expected that there is a stronger correlation in a plot of the uphill (larger) barriers versus the RMSD. Intuitively, this partially should result from a combination of the fact that the absolute values of the energy differences of two neighboring minima are a lower bound for the uphill barriers and the assumption that the average downhill barrier energy should rise if the distance between the minima increases. Therefore, already due to this effect, the probability to find small uphill barriers between structurally very different minima should be expected to be small.

In order to analyze this idea more rigorously, a simple parabola model of the PES, as illustrated in Fig. 5.3, is used. In fact, similar parabola models can be used for the explanation of the Bell-Evans-Polanyi principle (a linear model is sufficient, though), the Marcus equation, Hammond's postulate and the relationship of low-curvature directions with low barrier energies.[19,161,200,267–270] The first and the latter of the just mentioned effects are used in the escape phase of MH and are illustrated in the right-hand side of Fig. 3.1 of Sec. 3.5. In such a parabola model, the transition state is given by the intercept $(\xi, E(\xi))$ of both parabolas. From Fig. 5.3 it is evident that the barrier energies should rise with increasing distances between the minima.

Here both parabolas are assumed to have the same curvature k ("force constant"), and their minimum values are shifted by an amount of ϵ. The structural distance of both minima is denoted as a. Consequently, the transition state ξ and its corresponding uphill barrier $E_\mathrm{u} = E(\xi)$ is given by

$$\xi = \frac{a}{2} + \frac{\epsilon}{2ak}, \tag{5.1}$$

$$E_\mathrm{u} = k\left(\frac{a}{2} + \frac{\epsilon}{2ak}\right)^2. \tag{5.2}$$

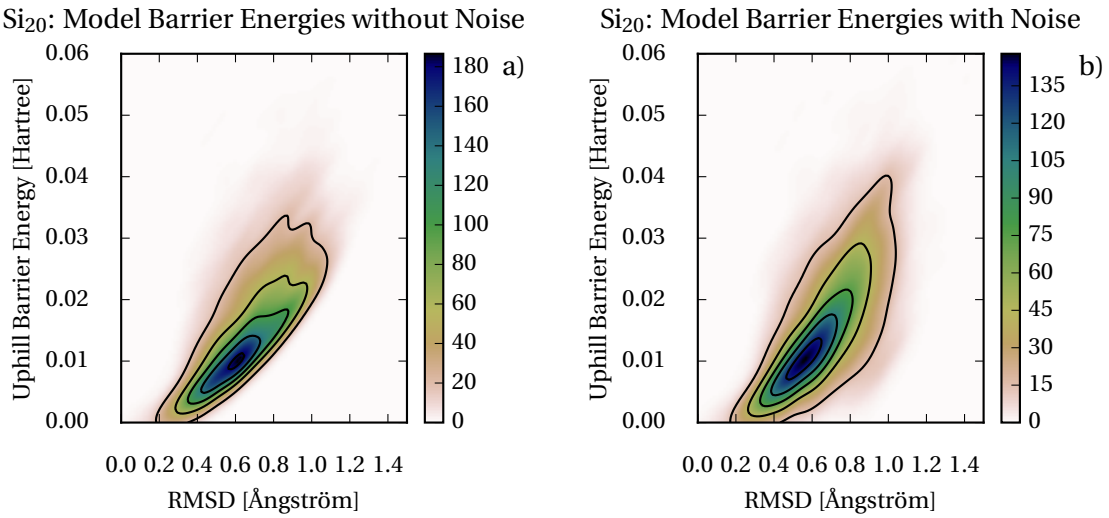

Figure 5.4: Same as Fig. 5.2 but for model uphill barrier energies instead of energy differences of minima. Panel a) shows the distribution of uphill barriers plotted versus the configurational distance of directly neighboring minima, as obtained by the deterministic model of Eq. 5.2. Panel b) shows the same distribution when using the stochastic model defined by the probability density of Eq. 5.4. Both plots use the same pairs of minima that also were used for Fig. 5.2.

For each pair of minima, this model is applied to the data of Fig. 5.2 and the result is visualized in Fig. 5.4a ($k = 0.08\,\mathrm{Ha/Å^2}$). In contrast to the energy differences of the minima in Fig. 5.2, this model predicts a clear correlation between the structural difference (RMSD) of two directly neighboring minima and the energy of the corresponding uphill barrier.

However, in real systems, the strict validity of Eq. 5.2 should not be expected. Similar to the Bell-Evans-Polanyi principle, the present relationship is expected to be fulfilled more likely in an average sense.[19] In an attempt to obtain a plot of the barrier distributions that might be closer to reality, the model of Eq. 5.2 is modified in the following. Given the energy difference ϵ of a pair of minima, the uphill barrier is written as

$$E_\mathrm{u} = \epsilon + E_\mathrm{d}, \tag{5.3}$$

where E_d is the downhill barrier. The energy difference ϵ of the minima can be regarded as the exactly known part of the uphill barrier. Using Eq. 5.2 and a Gaussian distribution, the downhill barrier E_d is drawn from the distribution

$$\rho(E_\mathrm{d}) \propto \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(E_\mathrm{d}-\mu)^2}{2\sigma^2}\right) & \text{if } E_\mathrm{d} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

Here the mean is given by $\mu := k\left(\frac{a}{2} + \frac{\epsilon}{2am}\right)^2 - \epsilon$ (cf. Eq. 5.2) and the standard deviation of the Gaussian part is assumed to be $\sigma := \frac{1}{2}\mu$. Clearly, there is no evidence that the uphill barrier

energies actually follow this distribution, but, in the spirit of a thought experiment, the present attempt is used to speculate on how the real uphill barrier distribution might look like. Setting $k = 0.08\,\mathrm{Ha/Å^2}$ and drawing for every pair of minima of the data shown in Fig. 5.2 an energy from the distribution given by Eq. 5.4 results into a data set that is visualized in Fig. 5.4b. As is evident from this figure, despite the stochastic noise introduced by Eq. 5.4, there still is a good correlation between the RMSD and the uphill barrier.

It remains to be verified if the energies of real (computed) uphill barriers between structurally very different minima also tend to be larger than the energies of the uphill barriers between structurally similar minima. If there is a breakdown in this hypothesis, it is expected that no correlation of the type shown in Fig. 5.4 is seen. For this verification, transition states and their directly connected minima were computed for Si_{20} and Au_{26}^- at the DFT level of theory as implemented in the BigDFT code and for $(NaCl)_{32}$ and $(NaCl)_{29}$ using the Born-Mayer-Huggins-Tosi-Fumi[118–122] (BMHTF) force field. For Si_{20} the PBE[105] functional was used, whereas for Au_{26}^- the LDA[86,96] functional was used and in case of the BMHTF force field the parameters of Ref. [123] were chosen. Furthermore, transition states and the directly connected neighbors were computed for the Lennard-Jones[111,112] clusters of sizes 19, 38 and 55. Except for Au_{26}^-, the geometries and energies of the minima, as well as their connectivity, were established using the MHGPS method as implemented in the BigDFT suite. In the case of Au_{26}^- the data was taken from Chap. 4 and it is referred to this chapter for a description of its computation. The density plots of the uphill barrier energies versus the RMSD are given in Fig. 5.5. As can be seen from this figure, there is indeed a good correlation between the structural difference (RMSD) and the uphill barrier.

Though the permutationally optimized RMSD is a very natural measure for structural differences, it is very time consuming to compute, which often makes it impracticable to use. For example, the computation of the roughly 58,000 RMSDs for the LJ_{55} plot in Fig. 5.5 took about 14 hours (wall clock time), despite using 150 cores in parallel. Of course, actual wall clock times depend very strongly on the underlying hardware. Nevertheless, this example illustrates that computing large numbers of RMSDs can be problematic in practice. Therefore, the plots of Fig. 5.5 have been repeated using *s*- and *p*-orbital fingerprint distances instead of RMSDs and are shown in Fig. 5.6. Again, a correlation between the structural difference measured by the *s*- and *p*-orbital fingerprint distance and the uphill barrier energy can be observed. Using *s*- and *p*-orbital based fingerprint distances as a measure for structural differences, the LJ_{55} plot in Fig. 5.6 took on the order of minutes on a single core, which is a striking advantage over the RMSD and makes it much more useful in practice. Plots from fingerprint distances using only *s*-type orbitals have a very similar appearance and are given in Fig. 5.7.

Finally, a short comment seems to be appropriate on why in the present chapter it is almost exclusively focused on the uphill barriers. After all, as can be seen from Eqs. 5.2 and 5.3, the same dependence of the downhill barriers on the structural distance as for the uphill barriers is predicted, except for a constant energy shift that is given by the energy difference of both minima. This, however, does not imply that necessarily a similar correlation as for the uphill

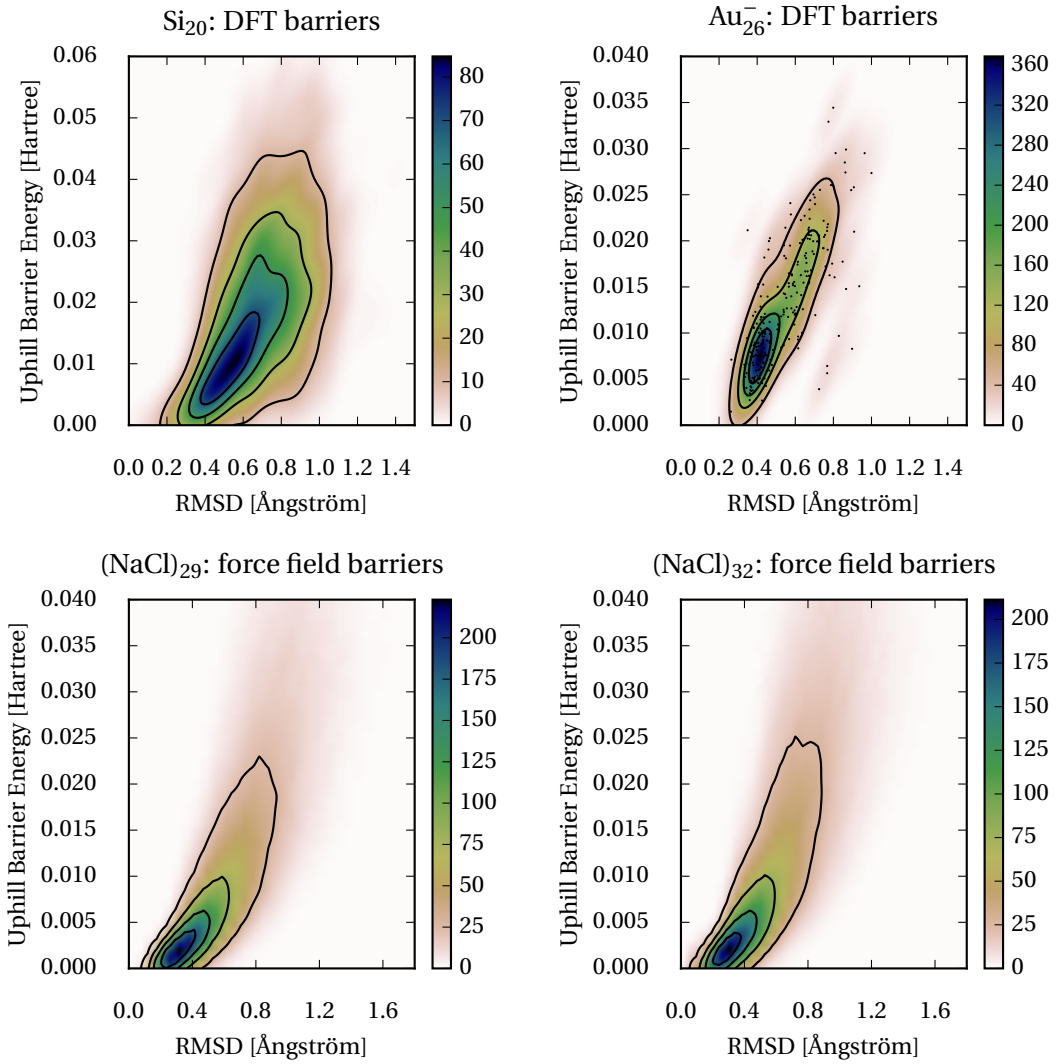

Figure 5.5: Gaussian kernel density estimates of the uphill barrier energies versus the (permutationally and chirally optimized) RMSD distance of minima pairs that are separated by only one intermediate transition state. If two minima are connected by more than one intermediate transition state, only the transition state with the lowest energy was included in the data sets used for these plots. The plot for Au_{26}^- was obtained from only 259 transition states. It, therefore, is possible to show every single data point for Au_{26}^-, which allows to demonstrate the soundness of the Gaussian kernel density estimate. The plot for Si_{20} was generated from roughly 3,000 transition states and the plots for the systems described by force fields were obtained from roughly 50,000 to 70,000 transition states.

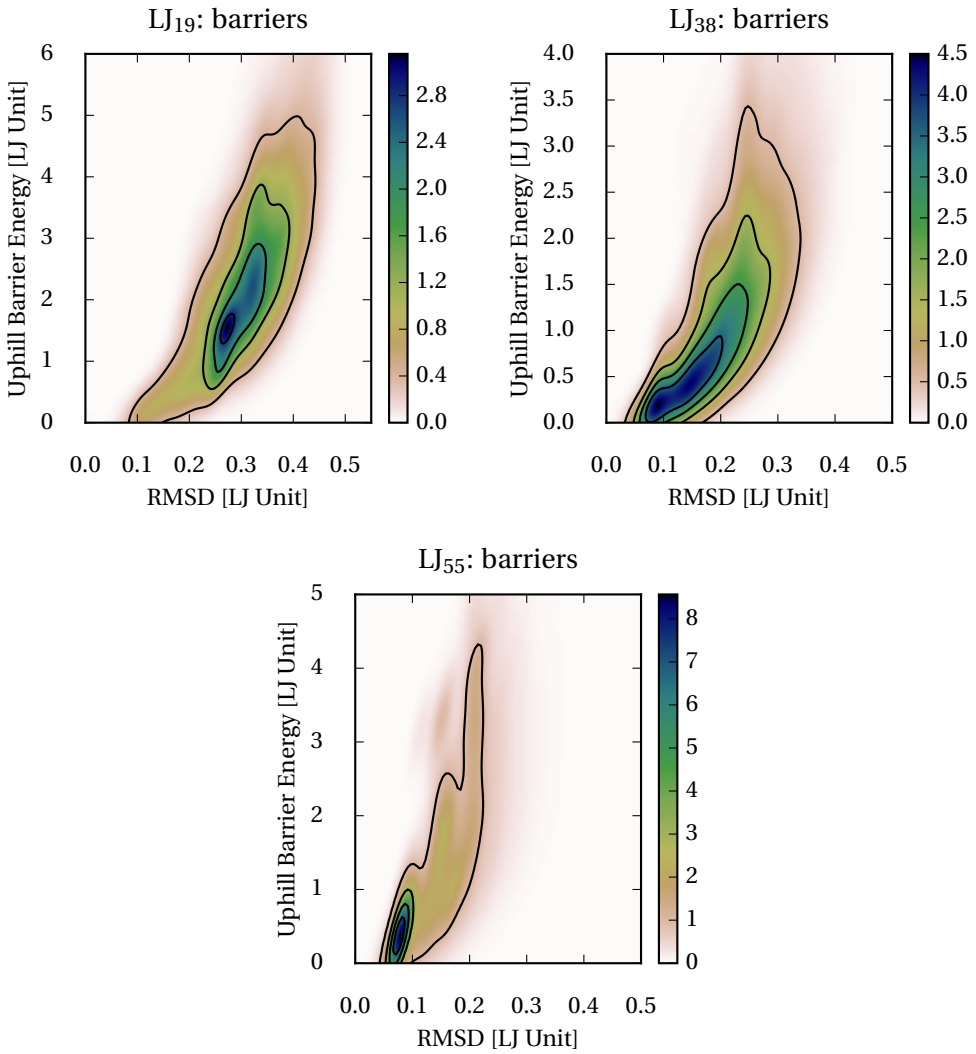

Figure 5.5 (*Continued.*)

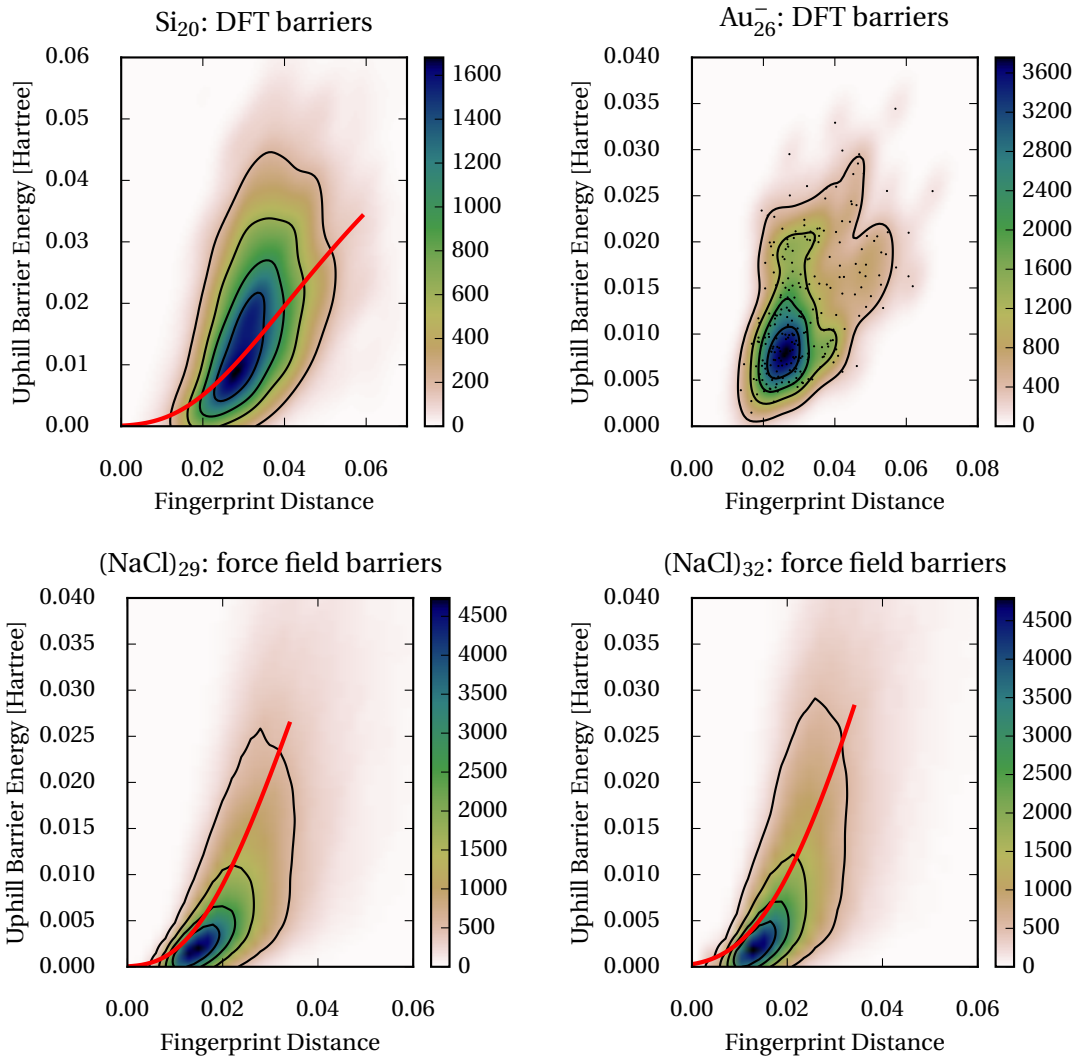

Figure 5.6: Same as Fig. 5.5, but using *s*- and *p*-orbital fingerprint distances instead of the permutationally optimized RMSD. Plots from fingerprint distances using only *s*-type orbitals have a very similar appearance and are given in Fig. 5.7. The red lines are graphs of Eq. 5.6 and are discussed in Sec. 5.2.

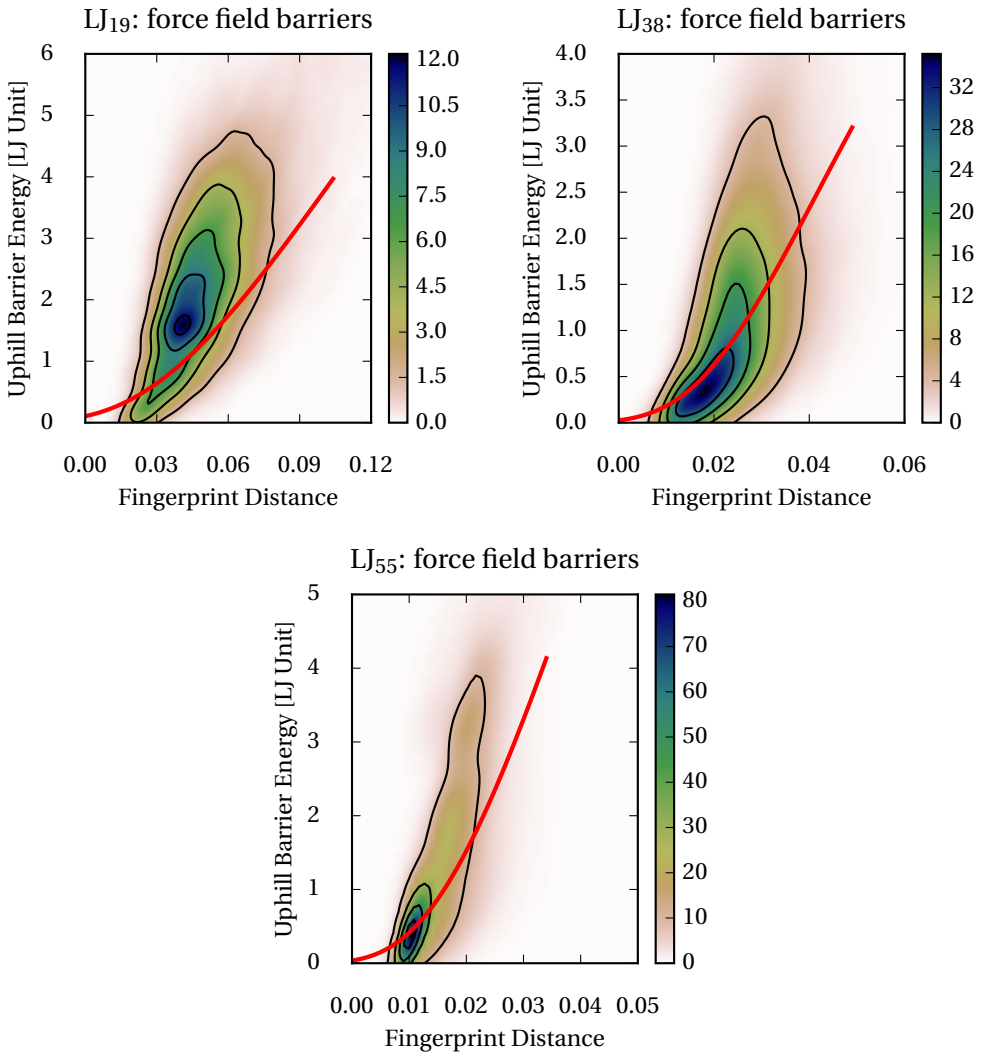

Figure 5.6 (*Continued.*)

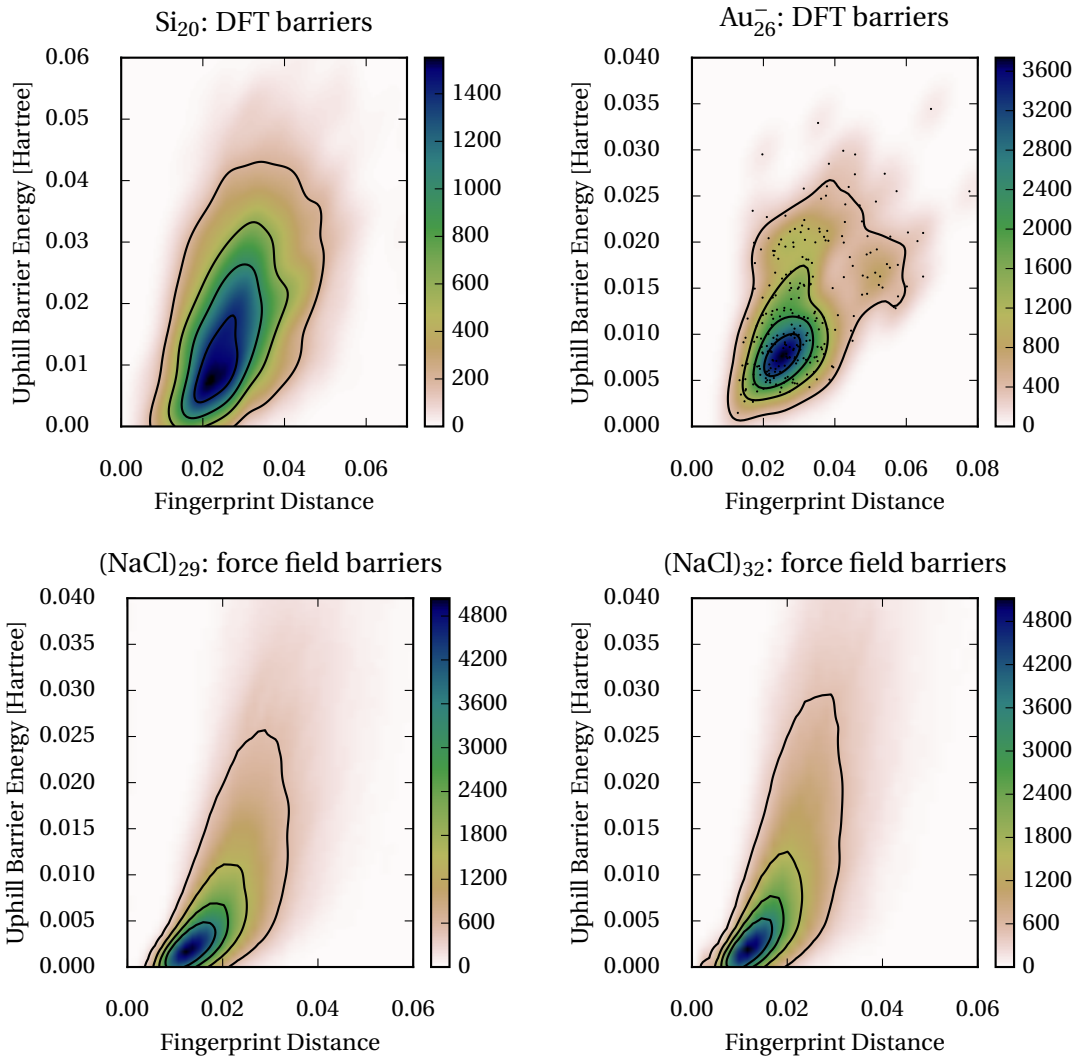

Figure 5.7: Same as Fig. 5.6 but using only *s*-orbital based fingerprint distances.

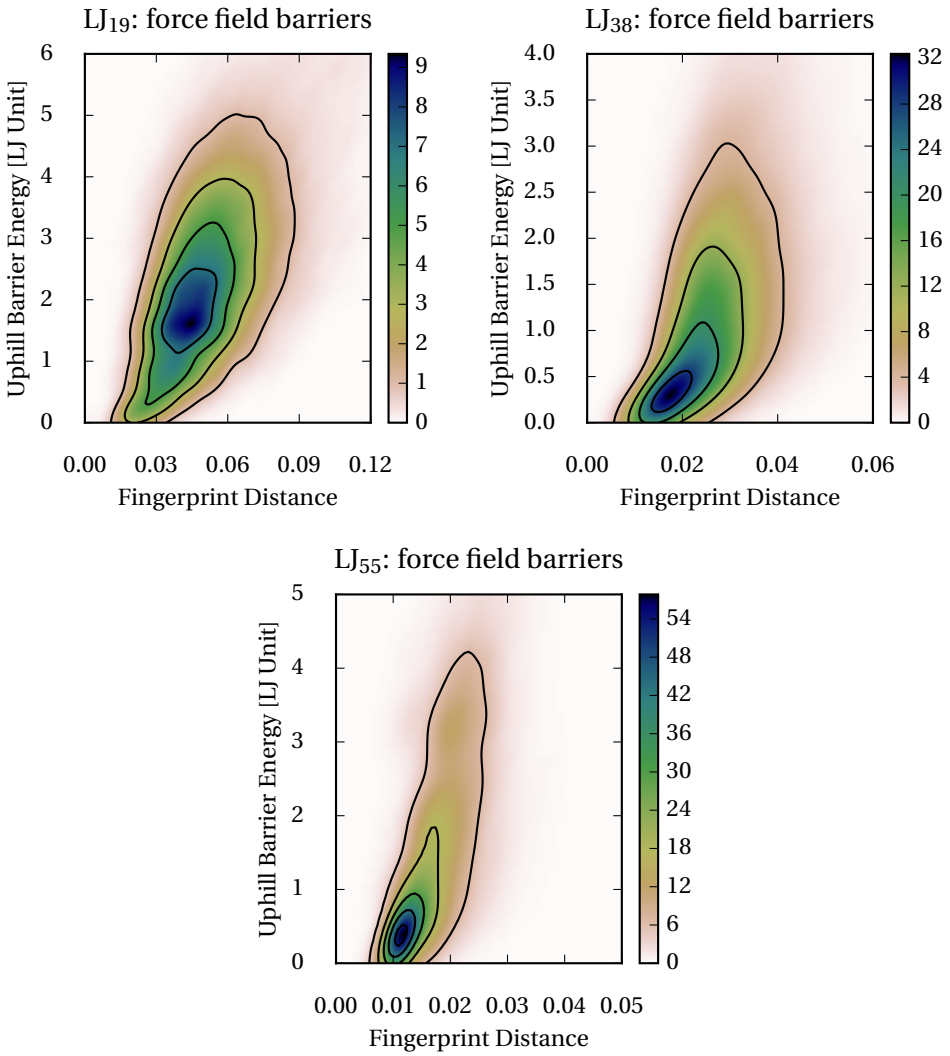

Figure 5.7 (*Continued.*)

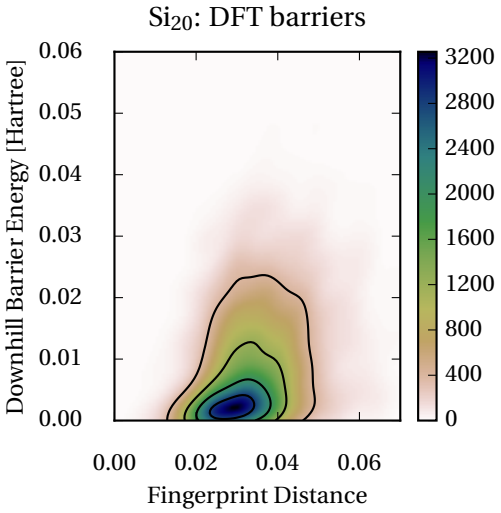

Figure 5.8: Gaussian kernel density estimates and binned averages of the downhill barrier energies of Si_{20} versus the structural distance measured by *s*- and *p*-orbital based fingerprints of minima pairs that are separated by only one intermediate transition state.

barriers must be observed for the downhill barriers. The reason is, that even though two minima might be far apart from each other, the downhill barrier can be vanishingly small if, in return, the energy difference between the two minima is comparatively large. Indeed, plotting the downhill barrier versus the structural difference results in a distribution that looks very similar to the distribution of the energy differences of the minima. As an example, such a plot is given in Fig. 5.8 for the Si_{20} system.

5.2 Generating Rough Overviews of Potential Energy Surfaces

In this section, a preliminary and empirical method suitable to generate qualitative connectivity databases is presented. This method is based on post-processing data obtained from one or several MH runs. Once MH runs are done, the computational cost of this method is independent of the underlying level of theory that was used for the MH runs. On a single core of a standard office computer, this method allows the generation of qualitative connectivity databases within a negligible amount of wall clock time, even if the qualitative connectivity databases shall describe PESs that are defined by computationally demanding methods, like for example DFT. To introduce this novel method, first the term “qualitative connectivity database” is defined. A qualitative connectivity database is understood to contain three types of information. First, it contains all local minima visited during a certain number of MH runs. Second, it contains the information which minima were visited consecutively by the MH walkers and finally, also a qualitative measure for the energy needed to interconvert the consecutively visited minima is part of a qualitative connectivity database. Furthermore, a pair of minima visited consecutively by the MH walker will be denoted as “transition pair”.

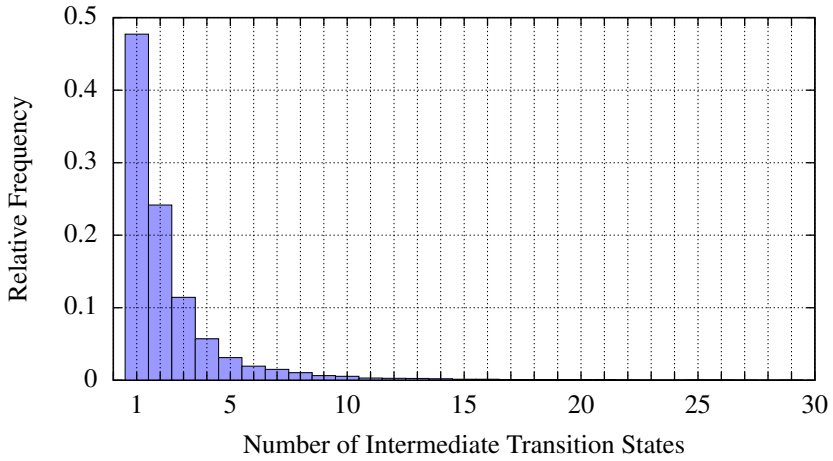

Figure 5.9: Shown for the LJ_{55} system is the relative frequency of the number of intermediate transition states needed by the MHGPS approach as implemented in the BigDFT-suite to connect pairs of consecutively accepted minima. The data set consists of more than 20,000 connection attempts that were stopped if the connection could not be established within 30 transition state computations.

In contrast to such a qualitative connectivity database, the stationary point database outlined in Chap. 3 and previously defined by Wales[3,38,39] contain minima, transition states and the information to which minima the transition states are connected by minimum energy or energy minimized pathways. Thus, a qualitative connectivity database can be seen as an approximation to a stationary point database. The connectivity information is approximated by the information which minima were visited consecutively by the MH walker. This is a reasonable approximation, because the MH walkers explore the PES by means of short MD trajectories that, at most times, have relatively moderate initial kinetic energies. As a consequence, the geometries of transition pair members typically are very similar to each other, a circumstance that is also used in the MHGPS scheme discussed in Sec. 3.5. Quantitative evidence for the validity of this connectivity approximation is given in Fig. 5.9. In this figure, the relative frequency of the number of intermediate transition states needed by the MHGPS method to connect pairs of consecutively accepted minima is given. These numbers constitute an upper bound to the minimum number of intermediate transition states located in between two consecutively accepted minima. It can be seen from this figure that the majority of consecutively accepted minima can be connected with each other by no more than two intermediate transition states.

What remains to be discussed is, how an educated guess for the energy, which is needed to interconvert the minima of a transition pair, can be obtained. Before describing the actual method for obtaining such a guess, a different approach is discussed. From a theoretical point of view, it would be very satisfying if Eq. 5.2 could be used to obtain a guess for the transition state energy. Indeed, using a suitable value for the force constant k, it turned out to be possible to generate disconnectivity graphs of similar quality as those based on the method

that is presented below. However, so far, it was only possible to choose good values for k, if the correct appearance of the disconnectivity graph was known. Unfortunately, a procedure that is able to reliably determine the force constant and that is able to give disconnectivity graphs of similar quality as those based on the method outlined below has yet to be found. In fact, using inappropriate values for k can produce completely misleading disconnectivity graphs. In contrast to this, in all tested cases, the approach discussed below produced qualitatively very reasonable disconnectivity graphs. The following empirical method only uses the geometries and energies of transition pairs to obtain a guess for the transition state energy. Nevertheless, MH is gathering much more information on the PES, for example by means of the softening procedure, the MD trajectories or the relaxation trajectories. Therefore, the possibility to develop a method that can extract suitable quantities from these data such that Eq. 5.2, or a similar model, could reliably be used for obtaining a measure for transition states energies seems conceivable. For this reason, it is hoped that further research effort will render this information accessible for the usage of generating rough characterizations of PESs.

The remainder of this section focuses on describing the empirical method that, so far, was able to produce the best qualitative guess for the transition state energies. In this approach the energy difference of the two minima of a transition pair is compared to the average energy difference of minima of transition pairs that are separated by a similar structural fingerprint distance. If the energy difference is larger than the average value at this fingerprint distance, the uphill barrier of a transition pair is estimated as the absolute value of the energy difference of the two transition pair members. Otherwise, the uphill barrier is estimated as the average absolute value of energy differences at this fingerprint distance. In practice, this is done by plotting the absolute values of the energy differences of the minima of each transition pair versus their fingerprint distance and computing binned averages of this data. A continuous function describing this binned average is obtained by means of a fitting procedure. Of course, this approach does not give a quantitative estimate of the energy of each single barrier, but it is intended to reproduce the energy scale and roughly the average trend in uphill barrier energies that was discussed in the previous section. More explicitly, assuming the minima energies of a transition pair to be E_1 and E_2 with $E_1 \leq E_2$, the absolute energy E_t needed to interconvert the two minima is estimated as

$$E_\mathrm{t} := \max\left(E_1 + E_\mathrm{u}(a), E_2\right), \tag{5.5}$$

where the max-function returns the larger of its two arguments and the uphill barrier energy E_u is a function of the s-only or s- and p-overlap fingerprint distance a. E_u is defined as

$$E_\mathrm{u}(a) := \alpha \exp(-\beta|a+\gamma|^\delta), \tag{5.6}$$

where the parameters α, β, γ and δ are obtained by a fit to the binned averages of the energy differences of the minima of transition pairs. The fitting function given in Eq. 5.6 is a heuristic and pragmatic choice that turned out to work well in all tested cases. Of course, other functions can be chosen, if they reasonably reproduce the binned averages and, thereby, reproduce

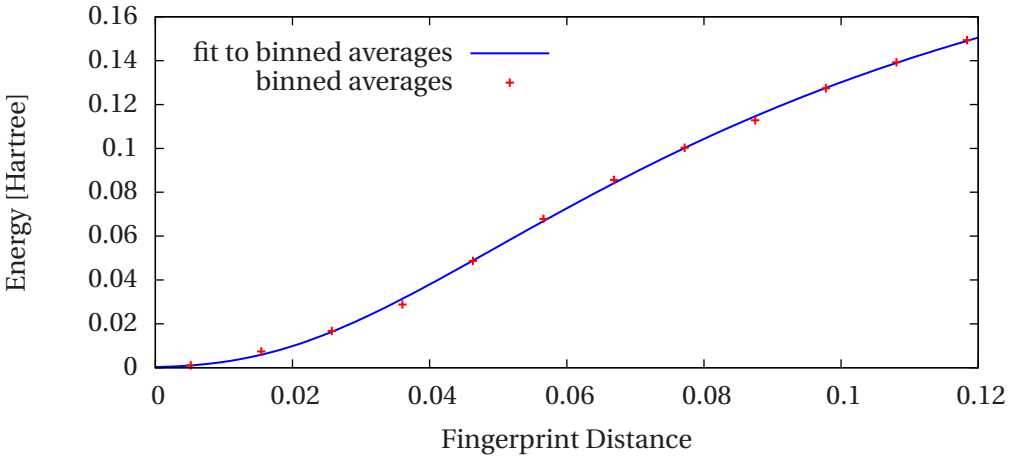

Figure 5.10: Fit of E_u as defined in Eq. 5.6 to the binned averages of the energy differences of $(NaCl)_{32}$ transition pairs, as modeled by the BMHTF force field, versus their structural difference measured by the overlap matrix fingerprint distance using *s*- and *p*-type orbitals. 25 bins were used for grouping the roughly 28.000 data points. Of those 25 bins, only those that contain at least 5% of the data points of the bin with the most data points are shown and were used for the fit. The values of the fitting parameters are $\alpha = 0.2449$ Ha, $\beta = 0.0128$, $\gamma = 0.0445$ and $\delta = -2.0159$.

the energy scale and the average tendency of increasing barrier heights for increasing structural differences. The fitting itself is performed with the help of the nonlinear least-squares Marquardt-Levenberg algorithm as implemented in the gnuplot code.[271–273] Of course, other fitting methods can be used, because E_u is only required to provide a continuous function of the qualitative trends for the uphill barrier energies. A plot exemplifying such a fit is given in Fig. 5.10 for the case of $(NaCl)_{32}$.

It turned out that by using Eq. 5.5 for obtaining transition state energy guesses, it is possible to produce disconnectivity graphs that reasonably reflect the characteristics of a PES. Before presenting these disconnectivity graphs, it is appropriate to discuss why the reasonable performance of Eq. 5.5 should present no mystery. To see this, first it is realized that Eq. 5.5 splits up the transition pairs into two sets.

In the first set, the uphill barrier of a transition pair is guessed by means of Eq. 5.6. In Fig. 5.6, the fitting function Eq. 5.6 is plotted on top of the uphill barrier distributions of Si_{20}, $(NaCl)_{29}$, $(NaCl)_{23}$, LJ_{19}, LJ_{38} and LJ_{55}. From these plots it is evident that the binned average of the absolute values of the energy differences of transition pair minima is a reasonable guess for the uphill barrier energy. Eq. 5.6 prevents the assignment of low transition state energies to transition pairs with structurally very different minima and, therefore, is in agreement with the results of Sec. 5.1. This agreement is essential for an acceptable reproduction of the characteristics of a PES. Otherwise, as will be seen from the disconnectivity graphs that are presented below, superbasins are likely to be merged, which can result into a completely misleading appearance of a PES. Furthermore it can be seen from Fig. 5.6 that the uphill barrier

energy which is assigned to a transition pair corresponds in most cases to a not completely unlikely uphill barrier energy at a given structural distance. As was demonstrated by Fig. 5.9, the minima of many transition pairs are separated by only one intermediate transition state and it is clear that the trend of increasing uphill barrier energies with increasing structural distances that was described in Sec. 5.1 can be applied to these transition pairs. However, there is no strict guarantee for the minima of a transition pair to be in a close neighborhood to each other. Despite this fact, it is still the same trend that is used to obtain a guess for the barrier energies of those transition pairs. At a first glance, this might be surprising since two structurally very different minima, which only can be interconverted into each other by crossing many intermediate transition states, might very well be separated by a low overall barrier. For example, this can be the case if the pairwise structural distances of all intermediate minima are small. Using a measure for the transition state energies that is based on the correlation discussed in Sec. 5.1, a high barrier energy will be assigned to the direct transition between such minima. However, this is not a disadvantage, but rather a desirable effect. Typically, the analysis of a qualitative connectivity database will focus on low energy pathways. In such an analysis, the direct interconversion of those far apart minima is disfavored due to the high energy that is assigned to their direct interconversion. In contrast, low barrier energies are properly assigned to the pathway that leads over the large number of pairwise structurally similar minima, which allows for its identification.

In the second set, the uphill barriers of transition pairs are approximated by the energy of the energetically higher minimum. For transitions with downhill barriers that are small compared to the uphill barrier, this is a sufficient approximation. However, if the energy difference between two minima is small and their structural difference large, this approximation is not only quantitatively, but also qualitatively very inaccurate. Fortunately, Eq. 5.6 rigorously prevents the latter transition pairs from being included into this second set. This second set only contains transition pairs with above-average energy differences with respect to a given structural distance. Therefore, for those transition pairs for which a significant underestimation of the transition state energy endangers a reasonable reproduction of the overall PES characteristics in a disconnectivity graph, the uphill barriers are not estimated by the energy difference of the involved minima.

Fig. 5.11 displays disconnectivity graphs for Si_{20}, $(NaCl)_{29}$, $(NaCl)_{32}$, LJ_{19}, LJ_{38} and LJ_{55}. As above, the PES of Si_{20} was computed at the DFT level of theory as implemented in the BigDFT code (PBE exchange correlation functional). For the sodium chloride clusters, again the BMHTF force field was used. No disconnectivity graphs are presented for Au_{26}^- because only the local minima, but not the complete minima hopping history, were archived from the MH runs that were performed in the course of the work presented in Chap. 4. The panel labels of Fig. 5.11 follow the scheme (x.n), where "x" is one of a, b, c, d, e or f and represents the system (a=Si_{20}, b=$(NaCl)_{29}$, c=$(NaCl)_{32}$, d=LJ_{19}, e=LJ_{38} and f=LJ_{55}) and n runs from one to three. Disconnectivity graphs in the panels (x.1) and (x.2) (the left and middle column of Fig. 5.11) are based on qualitative connectivity databases, where for the (x.1) panels the barrier energies were set to the energy of the higher minimum and for the (x.2) panels the barrier energies were

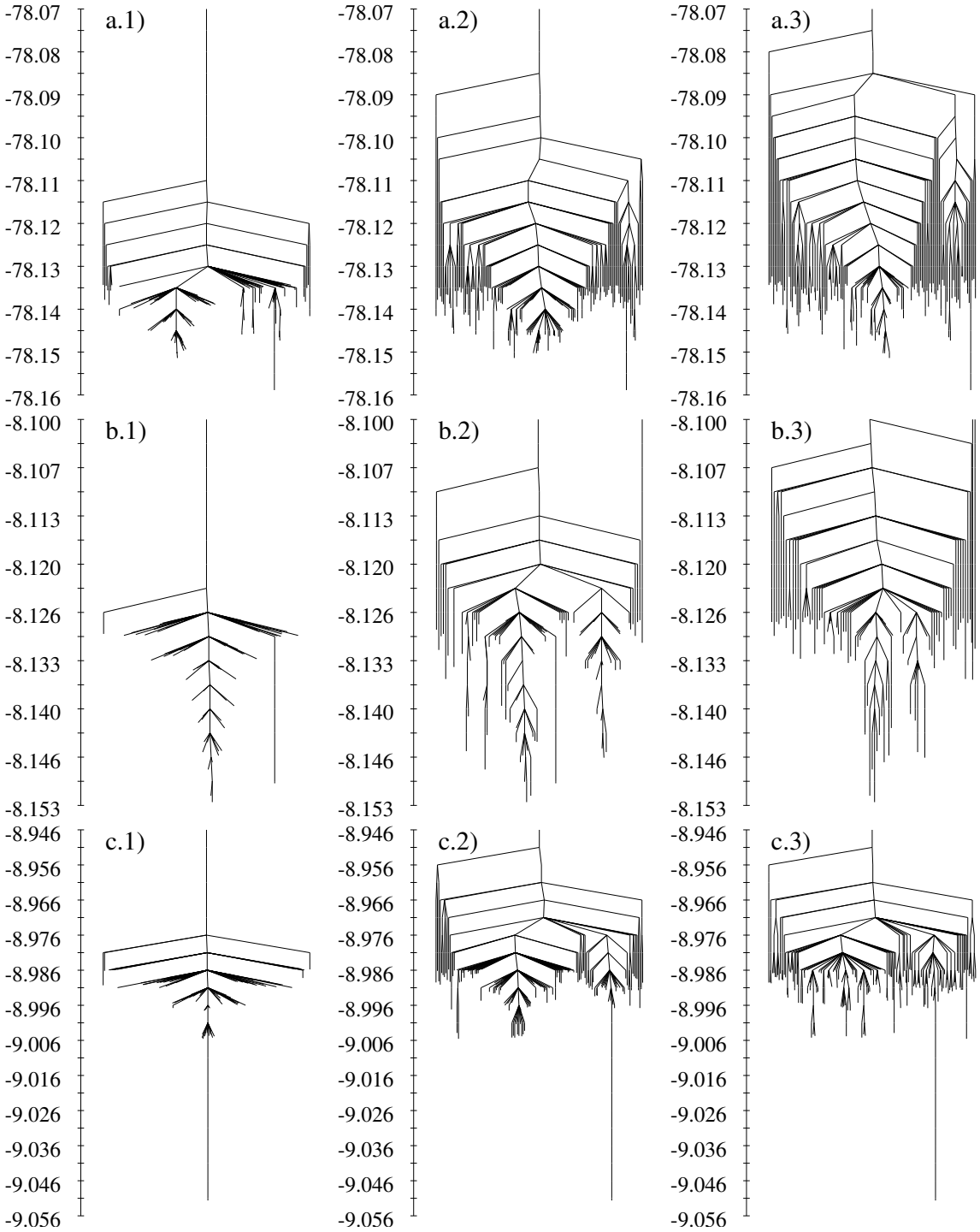

Figure 5.11: Disconnectivity graphs for Si_{20} (panels (a.n)), $(NaCl)_{29}$ (panels (b.n)), $(NaCl)_{32}$ (panels (c.n)), LJ_{19} (panels (d.n)), LJ_{38} (panels (e.n)) and LJ_{55} (panels (f.n)). The graphs in panels (x.1) and (x.2) are based on qualitative connectivity databases. For the (x.1) panels, the barriers were eliminated, whereas the approximations to the barrier energies described in Sec. 5.2 were used for the (x.2) panels. Reference graphs based on stationary point databases that were sampled by the MHGPS approach are shown in the rightmost column (panels (x.3)). The energy scale is in Hartree (Si_{20}, $(NaCl)_{29}$, $(NaCl)_{32}$) and in Lennard-Jones units (LJ_{19}, LJ_{38}, LJ_{55}).

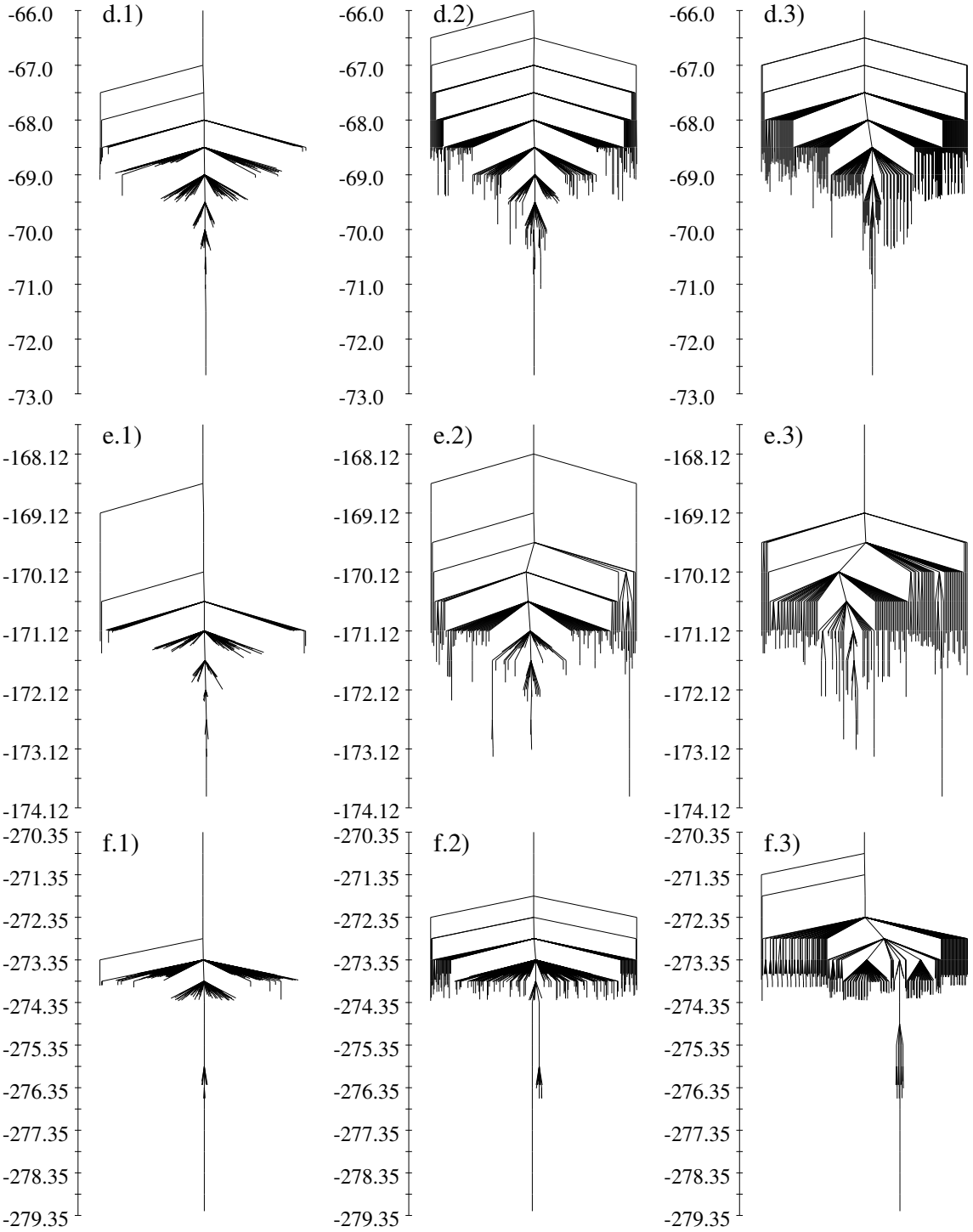

Figure 5.11 (*Continued.*)

approximated by Eq. 5.5 and the above described fitting procedure. The (x.2) disconnectivity graphs will also be denoted as "fingerprint disconnectivity graphs". For the center column of Fig. 5.11, fingerprint distances based on *s*- and *p*-orbitals were used. Disconnectivity graphs in the rightmost column of Fig. 5.11 (panels (x.3)) are based on stationary point databases that were generated by means of the MHGPS approach of Sec. 3.5. These standard disconnectivity graphs are considered as the reference for the present purpose. For each system, all three disconnectivity graphs show the same number of minima, however, not necessarily the identical minima. This is, because the stationary point databases are usually much more detailed, because they were thoroughly sampled by the MHGPS approach in order to generate exact reference disconnectivity graphs. In the following, rough sizes of the underlying databases are given in the format $[n;m]_{\text{sys}}$, where n indicates the number of transition pairs in case of qualitative connectivity databases and the number of transition states in case of the stationary point databases that were used for the standard graphs. The number of distinct minima is indicated by m and the system is indicated by the subscript. Qualitative connectivity databases: $[7,000;5,000]_{\text{Si}_{20}}$, $[82,000;71,000]_{(\text{NaCl})_{29}}$, $[28,000;25,000]_{(\text{NaCl})_{32}}$, $[1,800;1,100]_{\text{LJ}_{19}}$, $[87,000;64,000]_{\text{LJ}_{38}}$, and $[35,000;33,000]_{\text{LJ}_{55}}$. Stationary point databases: $[3,400;2,000]_{\text{Si}_{20}}$, $[200,000;171,000]_{(\text{NaCl})_{29}}$, $[68,000;61,000]_{(\text{NaCl})_{32}}$, $[65,000;14,000]_{\text{LJ}_{19}}$, $[68,000;45,000]_{\text{LJ}_{38}}$, and $[59,000;49,000]_{\text{LJ}_{55}}$.

Even if only using the connectivity as provided by the qualitative connectivity database, but eliminating all barriers, the double-funnel landscape of Si_{20} is clearly visible (Fig. 5.11a.1), nevertheless, the appearance of the disconnectivity graph is improved by using the fitting procedure for approximating transition state energies (Fig. 5.11a.2). It should be pointed out that MHGPS (Fig. 5.11a.3) found the energy landscape of Si_{20} to have a distinct double-funnel character. This finding is not essential for demonstrating the viability of qualitative connectivity databases, but to the best of the knowledge of the author this is, in itself, a previously unreported result. Though, for Si_{20}, the most important feature of the system is already visible in the (a.1) panel, the same is not true for the remaining systems. Except for Si_{20}, completely eliminating the barriers results in disconnectivity graphs that correspond to extreme structure seekers and the true topology of the PESs is not visible in the (x.1) panels. In contrast to this, the fingerprint disconnectivity graphs in the (x.2) panels exhibit a remarkable resemblance to the standard disconnectivity graphs shown in the (x.3) panels of Fig. 5.11.

The fingerprint disconnectivity graphs based on *s*- and *p*-orbital fingerprints are slightly more similar to the standard disconnectivity graphs than those based only on *s*-orbitals and shown in Fig. 5.12. Nevertheless, also the fingerprint disconnectivity graphs based on the *s*-only fingerprints provide a striking resemblance to the standard disconnectivity graphs, in particular if taken into account that generating fingerprint based disconnectivity graphs is a quasi-free lunch post-processing of MH data.

Besides for generating disconnectivity graphs and qualitatively judging the kinetics and thermodynamics of PESs, qualitative connectivity databases can also be used to extract well aligned sequences of minima that can be hoped to lie on a low-energy pathway between

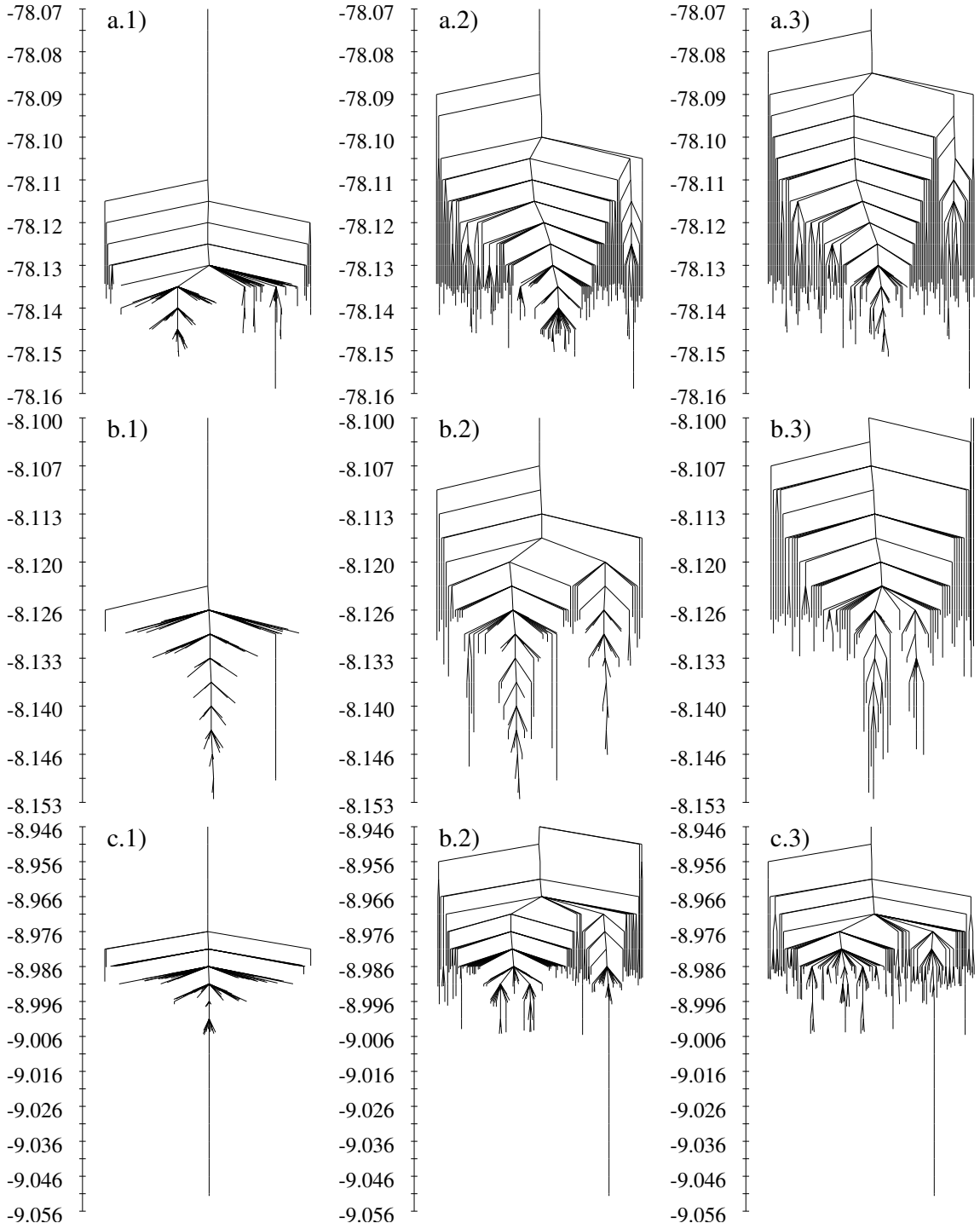

Figure 5.12: Same as Fig. 5.11 but using *s*-overlap fingerprints for the disconnectivity graphs in the center column.

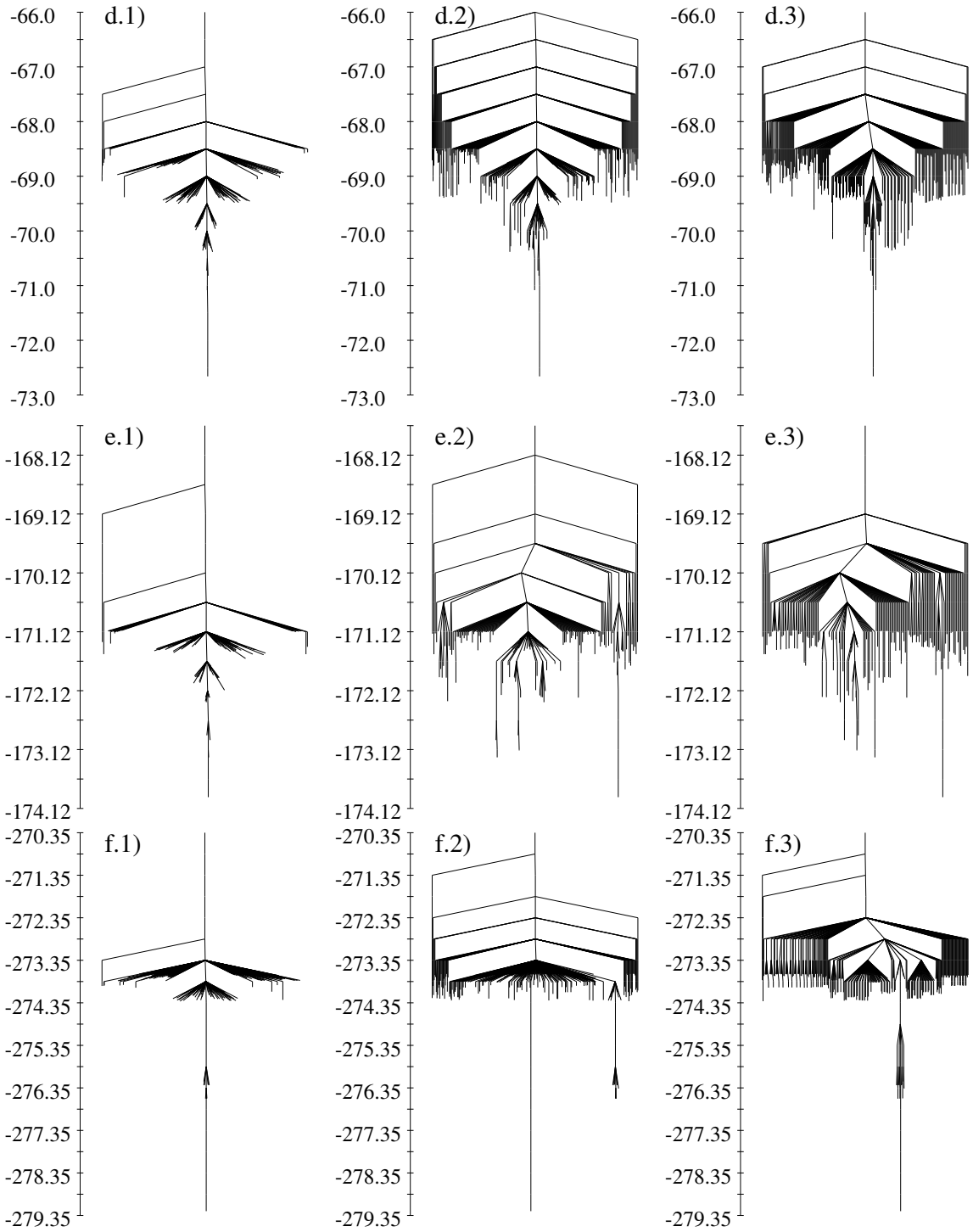

Figure 5.12 (*Continued.*)

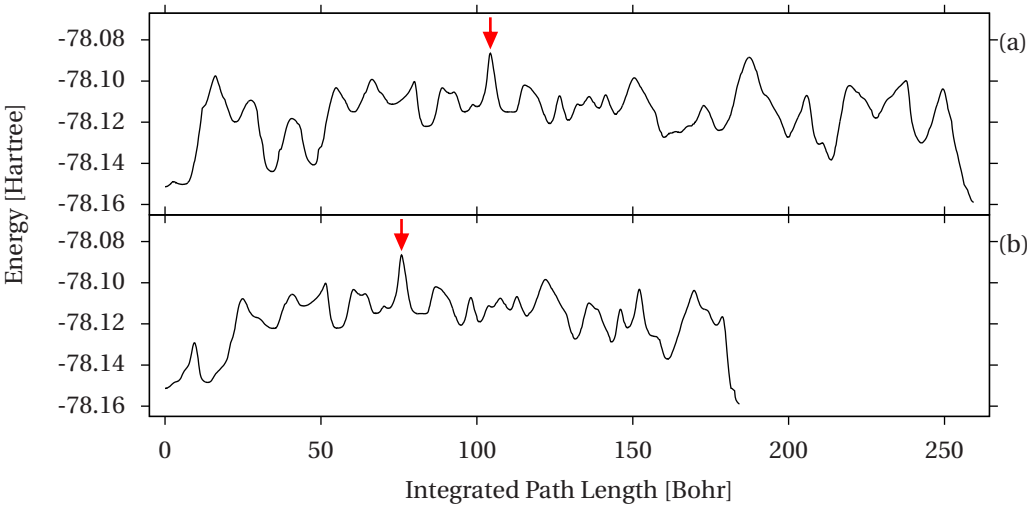

Figure 5.13: Two energy minimized pathways connecting the two lowest minima of Si_{20} (DFT, PBE). The pathway in panel (a) was obtained by extracting a sequence of minima from the qualitative connectivity database and using this sequence of minima as input for the MHGPS[62] code. Panel (b) shows a pathway that was extracted from a stationary point database sampled by entirely unbiased MHGPS runs (as described in Sec. 3.5). The shown pathways are SQNM[64] trajectories obtained by relaxations from the transition states after stepping away a small distance in positive and negative direction of the negative eigenmode. The transition states in the MHGPS runs were tightly converged by means of the SQNS[64] method. The red arrows indicate the highest energy transition states along the pathways. In both pathways, the highest energy transition states are identical.

two given states. Such minima sequences are of great importance, because they provide promising starting points for generating initial pathways that are needed for methods like TPS or its discrete variant, DPS.[26,27,38,39,180–184] For non-trivial reactions involving large structural changes such a generation of initial pathways is in itself a very difficult task and no generally applicable solution seems to exist, so far.[192] Isolating a suitable sequence of minima from a qualitative connectivity database can be done by applying a modified Dijkstra's algorithm which in a first round searches for a path that minimizes the maximum barrier at any of its transitions and in a second round minimizes with respect to the number of intermediate transitions (already mentioned and used in Sec. 3.1). Of course, the thus isolated pathways are not necessarily complete in the sense that it might not be possible to connect the two minima of a transition pair by only one single intermediate transition state. However, the isolated sequence of minima represents minima that were visited in consecutive order by an MH walker. Therefore, they are suitable for getting connected by the MHGPS code (instead of the usual sequence of accepted MH configurations).

For the Si_{20} system a sequence of minima between the putative global minimum and the putative second lowest minimum was extracted from the qualitative connectivity database. For this sequence of minima, all intermediate transition states and further emerging intermediate minima were determined by means of the MHGPS code as implemented in the BigDFT suite.

A pathway given by the trajectories of the SQNM energy minimizer[64] is shown in Fig. 5.13a. This pathway consists of 27 intermediate transition states. Fig. 5.13b shows a lowest barrier pathway that was extracted from the stationary point database which was sampled by means of unbiased MHGPS runs and already used for the standard disconnectivity graphs in Fig. 5.11a.3. The pathway in Fig. 5.13b consists of 20 intermediate transition states. Remarkably, both paths exhibit the same highest energy transition state which is highlighted by the red arrows in Fig. 5.13. Still, the path extracted from the stationary point database (Fig. 5.13b) is shorter than the path in Fig. 5.13a, both in terms of the integrated path length and in terms of the number of intermediate transition states.

Of course, there is no guarantee that extracting a sequence of minima from a qualitative connectivity database and connecting these minima by searching intermediate transition states will result in a pathway that has the same highest barrier as the pathway with the lowest highest barrier that is contained in a thoroughly sampled stationary point database. However, computer experiments performed for the LJ_{38} cluster indicate that physically reasonable pathways can be extracted from qualitative connectivity databases. Using the modified Dijkstra's algorithm, a sequence of minima was extracted from the complete qualitative connectivity database for LJ_{38}. By successively removing the highest energy transition along the lowest barrier pathway from the qualitative connectivity database, this process was repeated four more times. In this way, five different sequences of minima were obtained. Again, for each sequence, missing intermediate minima and transition states were added by means of the MHGPS code. This procedure resulted in four pathways with non-identical highest barriers, which are shown in Fig. 5.14. The dashed line at an energy of -169.708 LJ units indicates the highest barrier along the lowest-known barrier pathway.[2,35] The highest barriers along the pathways in Fig. 5.14 are not much higher than this lowest-known barrier.[†]

[†] For example, in the case of argon 1 LJ energy unit corresponds to roughly 10 meV.[114–116]

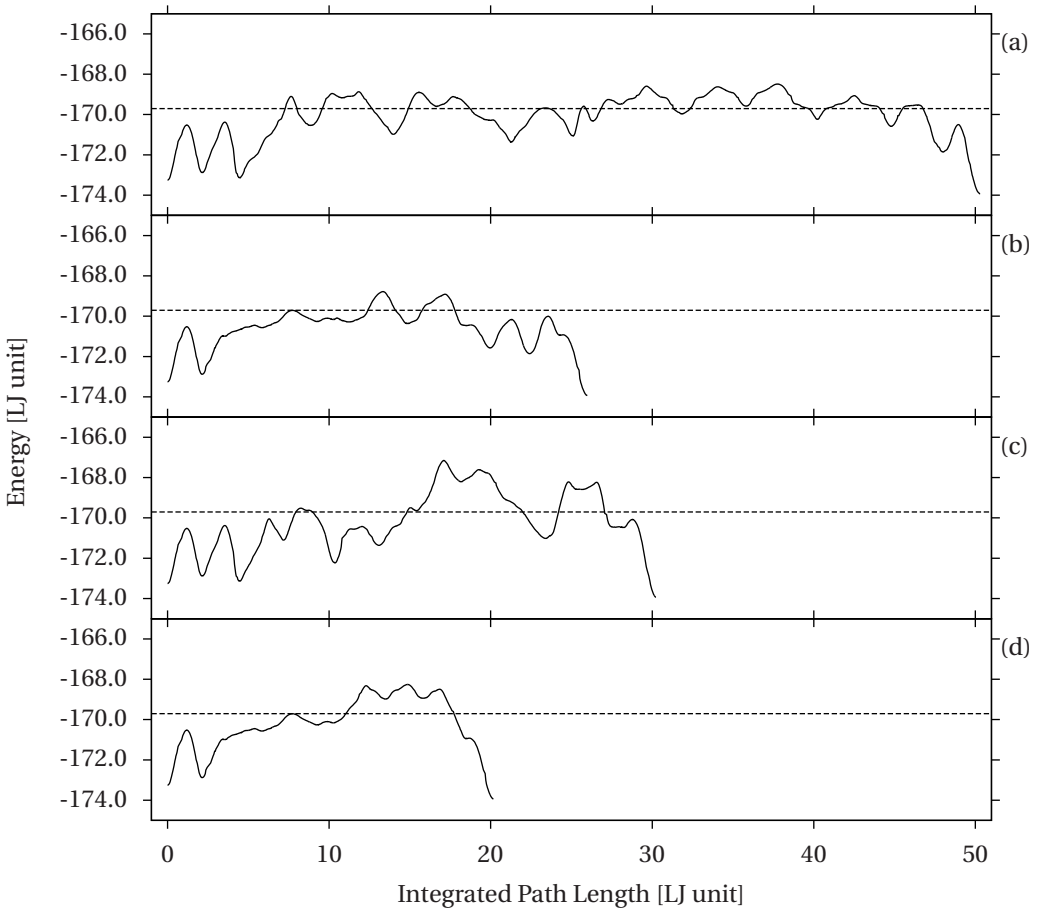

Figure 5.14: Energy minimized pathways connecting the two lowest minima of LJ_{38}. The pathway in panel (a) was obtained by extracting a sequence of minima from the complete qualitative connectivity database. Panels (b), (c) and (d) show pathways that were obtained by successively removing the highest energy transition along the lowest-barrier pathway from the qualitative connectivity database. Using the sequences of the extracted minima as input for the MHGPS[62] method, complete pathways were reconstructed. The SQNS[64] and SQNM[64] methods were used for converging to transition states and relaxing to the connected minima.

5.3 Conclusion

Based on a set of qualitatively different systems that exhibit covalent, metallic or ionic bonds, it was found that uphill barrier energies of transition states between directly connected minima tend to increase with increasing structural differences of the two minima. Based on this insight it also turned out that post-processing MH data at a negligible computational cost allows to obtain qualitative topological information on PESs that is stored in so called qualitative connectivity databases. These qualitative connectivity databases can be used for generating fingerprint disconnectivity graphs that allow to obtain a qualitative idea on thermodynamic and kinetic properties of a system of interest. Besides allowing to asses system properties without the need of a computational expensive explicit sampling of transition states and the assessment of the PES's connectivity based on minimum energy or energy minimized pathways, this method also serves as a valuable tool in terms of deciding if a certain multi-atomic system may exhibit desired properties in advance of investing significant resources for assessing theses properties more rigorously.

6 Conclusion and Outlook

Using the minima hopping global optimization method at the density functional (DFT) level of theory, new low-energy structures for neutral Au_{26} and its anion were found. The local-density and a generalized gradient approximation of the exchange-correlation functional predicted different structural motifs. A vast number of isomers within a small energy range above the respective putative global minima were observed with each exchange-correlation functional. Photoelectron spectroscopy of Au_{26}^- under different experimental conditions revealed definitive evidence of the presence of multiple isomers, consistent with the theoretical predictions. Comparison between the experimental and simulated photoelectron spectra suggested that the photoelectron spectra of Au_{26}^- contain a mixture of three isomers, all of which are low-symmetry core-shell-type clusters with a single internal gold atom. A disconnectivity graph for Au_{26}^- was presented that was computed completely at the DFT level. The transition states used to build this disconnectivity graph were complete enough to predict Au_{26}^- to have a possible fluxional shell, which can facilitate the understanding of its catalytic activity.

Motivated by the work on the disconnectivity graph for Au_{26}^-, the minima hopping guided path search (MHGPS) method was developed. Based on minima hopping (MH), MHGPS uses physically realizable molecular dynamics (MD) moves in combination with an energy feedback that guarantees the escape from any potential energy funnel. The energy conservation in the MD moves limits the heights of crossed potential energy barriers. Furthermore, the MD moves are short and, as a consequence, the consecutively accepted minima are structurally similar to each other. Therefore, consecutively accepted minima along the MH trajectory are particularly suitable as input structures for methods capable of finding transition states between two minima. Within the MHGPS code, the required iterative search for all the intermediate transition states between two consecutively accepted minima is fully automatized. The MHGPS approach does not rely on human intuition and PESs are probed in a completely unbiased fashion. MHGPS, therefore, does not fail to explore unforeseen and unexpected features of PESs. For Lennard-Jones benchmark systems, MHGPS was compared to transition path sampling (TPS) and a further previously known approach for the exploration of potential energy landscapes that is based on deterministic eigenvector following (EFE). Compared to these methods, MHGPS reduces the cost of finding complex reaction pathways by over one

order of magnitude. Furthermore, in contrast to TPS and EFE, MHGPS could successfully find the lowest-barrier pathways of LJ_{38} in all tests. When performing the same number of transition state computations, the MHGPS method was observed to detect a significantly larger number of distinct transition states than the EFE method. In a first application, MHGPS was used to study the 75-atom and 102-atom Lennard-Jones systems. For the 75-atom system new pathways were found with highest barrier energies that are significantly lower than the highest energy along any previously published lowest-barrier pathway. Furthermore, many of these pathways contain a smaller number of intermediate transition states than the previously published lowest-barrier pathway. In case of the 102-atom system, MHGPS found a previously unknown and energetically low-lying funnel.

In its core, MHGPS relies on the efficient and reliable computation of stationary points. For this purpose, a novel stabilized quasi-Newton minimization (SQNM) method and a stabilized quasi-Newton saddle finding approach (SQNS) were developed. Both optimizers are based on a technique that allows to obtain significant curvature information from noisy potential energy surfaces (PESs). These new optimizers replaced their initially used counterparts in the MHGPS code. The minimizer and the saddle finding method were compared to well established alternative methods, both at force field and DFT level of theory. In these benchmarks, the dimer saddle finding method[55–58] required between 1.4 and 7.6 times more energy and force evaluations for converging to a saddle point than the novel SQNS method. With respect to the number of wave function optimization iterations needed in DFT computations, the novel minimizer has demonstrated to be comparable in efficiency to the L-BFGS[137,138] method, however, without suffering from instabilities on noisy PESs – an issue the L-BFGS method is known to be prone to.[132]

Using binary Lennard-Jones clusters, it was argued that the relation between structural distances as given by fingerprint distances and energy differences, both measured between local minima and the global minimum of a system, can be used to discriminate glass-like from non-glassy PESs. Furthermore, it was found that uphill barrier energies of transition states between directly connected minima tend to increase with increasing structural differences of the two minima. At force field and DFT level of theory, this finding was demonstrated for different systems exhibiting covalent, metallic or ionic bonds. Based on an empirical post-processing approach of MH data, this insight can be exploited to obtain qualitative topological information on PESs that is consolidated in so called qualitative connectivity databases. From these databases, novel fingerprint disconnectivity graphs can be generated, which give a first qualitative insight into the thermodynamic and kinetic properties of a system. In the context of validating this empirical approach, the MHGPS method was applied to the PES of Si_{20} at the DFT level of theory to generate a reference database of minima, transition states and the information, which transition states are connected to which minima.

In retrospect, the MHGPS method and the stabilized quasi-Newton optimizers are possibly the most important achievments of this thesis. Currently, the MHGPS method is used at the DFT level of theory for probing the effects of system size and charge on the character of PESs

of cluster systems. With the presentation of the DFT disconnectivity graph of Si_{20} in Sec. 5.2, a first, partial result of this MHGPS study was already mentioned within the scope of this thesis, albeit only as a side note in the context of demonstrating the viability of the qualitative connectivity databases. Nevertheless, already the PES of Si_{20} is in itself an interesting and novel result. To the knowledge of the author, such an extreme double-funnel landscape has not been reported before at the DFT level of theory.

Hitherto MHGPS turned out to work well for the purpose of finding low-energy reaction pathways on PESs. It, therefore, is planned to extend its application to the investigation of reactive processes on surfaces. The MHGPS approach seems to be mainly limited by the available computer resources. For this reason, methods akin to those in Chap. 5 that allow the qualitative prediction of the characteristics of a PES are in the focus of further research. Such methods can allow to use the MHGPS code in a more target-oriented fashion.

Appendices

A Eliminating Translational and Rotational Modes

In the following, techniques to eliminate overall translations and rotations from a given vector $\mathbf{v}$ are shortly outlined. In favor of the usage of a common physical terminology, $\mathbf{v} \in \mathbb{R}^{3N}$ is considered to be a velocity vector of an N-atom system. The presented techniques, however, are not limited to velocities only, but can be applied to any vector that is used to displace atomic coordinates. Furthermore, all atoms are assumed to have unit mass.

The normalized translations $\hat{\mathbf{t}}_i := \mathbf{t}_i / |\mathbf{t}_i|$, $i = x, y, z$ (see Eq. 1.6) are orthonormal and a vector $\mathbf{v}'$ without any translational contributions can be obtained from $\mathbf{v}$ by projecting out the translational components. Under the assumption of unit masses, this is equivalent to subtracting the center of mass velocity from every atomic velocity:

$$\mathbf{v}' = \mathbf{v} - \sum_{i=\{x,y,z\}} (\mathbf{v} \cdot \hat{\mathbf{t}}_i) \hat{\mathbf{t}}_i. \tag{A.1}$$

The elimination of the rotational components with respect to the center of mass can also be achieved by simple projections, however, the vectors corresponding to overall rotations of Eq. 1.7 first have to be orthonormalized by a suitable orthonormalization scheme. The positions $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ of all N atoms, and with it the rotation vectors $\boldsymbol{\rho}_i$, $i = x, y, z$, are assumed to be expressed with respect to the center of mass. Gram-Schmidt orthonormalization[274,275]

yields three orthonormal vectors $\hat{\boldsymbol{\rho}}_i$, $i = x, y, z$ that span the rotational subspace:

$$\hat{\boldsymbol{\rho}}_x := \frac{\boldsymbol{\rho}_x}{|\boldsymbol{\rho}_x|} \tag{A.2}$$

$$\tilde{\boldsymbol{\rho}}_y := \boldsymbol{\rho}_y - \left(\boldsymbol{\rho}_y \cdot \hat{\boldsymbol{\rho}}_x\right)\hat{\boldsymbol{\rho}}_x$$

$$\hat{\boldsymbol{\rho}}_y := \frac{\tilde{\boldsymbol{\rho}}_y}{|\tilde{\boldsymbol{\rho}}_y|} \tag{A.3}$$

$$\tilde{\boldsymbol{\rho}}_z := \boldsymbol{\rho}_z - \left(\boldsymbol{\rho}_z \cdot \hat{\boldsymbol{\rho}}_x\right)\hat{\boldsymbol{\rho}}_x - \left(\boldsymbol{\rho}_z \cdot \hat{\boldsymbol{\rho}}_y\right)\hat{\boldsymbol{\rho}}_y$$

$$\hat{\boldsymbol{\rho}}_z := \frac{\tilde{\boldsymbol{\rho}}_z}{|\tilde{\boldsymbol{\rho}}_z|}. \tag{A.4}$$

A velocity vector $\mathbf{v}''$ without any rotational components can be obtained from $\mathbf{v}'$ by a simple projection:

$$\mathbf{v}'' := \mathbf{v}' - \sum_{i=\{x,y,z\}} (\mathbf{v}' \cdot \hat{\rho}_i)\hat{\rho}_i. \tag{A.5}$$

For linear molecules, care must be taken during the normalization of above equations as some rotational vectors can vanish and consequently must not be normalized.

Another technique for the elimination of rotational components not depending on the explicit knowledge of the rotation vectors $\boldsymbol{\rho}_i$ is inspired by the classical mechanics of rigid bodies. Again assuming unit masses for all atoms, the angular momentum is given by

$$\mathbf{L} = \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{v}_i \tag{A.6}$$

$$= I\boldsymbol{\omega}, \tag{A.7}$$

where I is the real symmetric inertia tensor and $\boldsymbol{\omega}$ the angular velocity. The tangential velocity $\mathbf{v}_{t,i}$ of atom i is given by $\mathbf{v}_{t,i} = \boldsymbol{\omega} \times \mathbf{r}_i = (I^{-1}\mathbf{L}) \times \mathbf{r}_i$. With that, one obtains the velocity $\mathbf{v}_i''$ of atom i, from which the rotational (tangential) components are eliminated:

$$\mathbf{v}_i'' = \mathbf{v}_i' - \left(I^{-1}\mathbf{L}\right) \times \mathbf{r}_i \tag{A.8}$$

For linear molecules, the inertia tensor is not invertible due to vanishing principal moments of inertia (eigenvalues of the inertia tensor). Therefore, in practice, this idea can be implemented by going to the principal axes frame which is defined by the eigenvectors $\mathbf{l}_j$, $j = 1, \ldots, 3$ of the inertia tensor. In the principal axes frame, the tangential velocity of the $i-$th atom can be

written as

$$\begin{aligned}\mathbf{v}_{t,i} &= \sum_{j=1}^{3} \omega_j \mathbf{l}_j \times \mathbf{r}_i \\ &= \sum_{j=1}^{3} \omega_j \tilde{\boldsymbol{\omega}}_{j,i}, \end{aligned} \tag{A.9}$$

where the ω_j are expansion coefficients of the angular velocity with respect to the principal axes of inertia and $\tilde{\boldsymbol{\omega}}_{j,i} := \mathbf{l}_j \times \mathbf{r}_i$. The rotational components of $\mathbf{v}'$ are now given by the projection of $\mathbf{v}'$ onto the subspace spanned by the $\tilde{\boldsymbol{\omega}}_j := \left(\tilde{\boldsymbol{\omega}}_{j,1}, \ldots, \tilde{\boldsymbol{\omega}}_{j,N}\right)$ that correspond to the non-vanishing principal moments of inertia.

B Measuring Structural Differences

Probably the most natural way to asses the structural difference of two N-atom configurations $\mathbf{R}^A = (\mathbf{R}^A_1, \ldots, \mathbf{R}^A_N)$ and $\mathbf{R}^B = (\mathbf{R}^B_1, \ldots, \mathbf{R}^B_N)$ is to shift and rotate the two geometries and to permute the indices of their atoms such that their root-mean-square displacement (RMSD) is minimal. In more rigorous mathematical terms this means:[167]

$$\mathrm{RMSD}(\mathbf{R}^A, \mathbf{R}^B) := \frac{1}{\sqrt{N}} \min_{P,U} \|\mathbf{R}_1 - U\mathbf{R}_2 P\|, \tag{B.1}$$

where it was implied that the $\{\mathbf{R}^A_i\}_{i=1,\ldots,N}$ and $\{\mathbf{R}^A_i\}_{i=1,\ldots,N}$ are measured with respect to their respective centroids. In the above formula, P and U are a $N \times N$ permutation and rotation matrix, respectively. There exist efficient methods to solve both the rotational and the permutational problem separately from each other. For example, Kabsch's algorithm or a more recent quaternions based approach can be used to find the optimal rotation.[276,277] The permutational assignment problem can be solved in polynomial time by means of the Hungarian algorithm.[278,279] However, the rotational and the permutational problem are not independent of each other. In order to find the optimal rotation, the optimal permutation has to be known and vice versa. In particular, this is a problem if $\mathbf{R}^A$ and $\mathbf{R}^B$ represent geometrically distinct atomic configurations. Recently, Sadeghi et al. approached this problem in a Monte Carlo fashion.[167] Unfortunately, the practicality of this Monte Carlo method is limited, because the time to find the optimal RMSD scales exponentially with the number of permutable atoms.[167] Motivated by this, Sadeghi introduced[167] configurational fingerprints which are given by the eigenvalues of an overlap matrix

$$O_{ij} := \int \Phi^{\mathbf{l}}_i(\mathbf{r}) \Phi^{\mathbf{l}'}_j(\mathbf{r}) \, d\mathbf{r}. \tag{B.2}$$

The Φ_i are Gaussian type orbitals centered on the atom at position $\mathbf{r}_i$

$$\Phi^{\mathbf{l}}_i(\mathbf{r}) \propto (x - x_i)^{l_x} (y - y_i)^{l_y} (z - z_i)^{l_z} \exp(-\alpha_i \|\mathbf{r} - \mathbf{r}_i\|^2), \tag{B.3}$$

where $\mathbf{l} = (l_x, l_y, l_z)$ is a multi-index indicating the angular momentum $L = l_x + l_y + l_z$. Depending on the value of L, the orbitals are classified as s-type orbitals ($L = 0$), p-type orbitals ($L = 1$),

d-type orbitals ($L = 2$), and so on. The orbital width α_i is usually chosen inversely proportional to the covalent radius of the atom species on which the orbital is centered on. The set of sorted overlap matrix eigenvalues for a given cluster can be considered to form a vector which defines the fingerprint of the configuration. The structural difference between two clusters is given by the root mean square of the difference vector between the two fingerprint vectors and throughout this thesis, this distance measure is denoted as "fingerprint distance". Besides being invariant under translations, rotations and reflections of the configuration, these fingerprints are also invariant under permutation of the atomic indices. In contrast to the RMSD, the overlap-matrix based fingerprints are computationally cheap and turned out to be reliable for distinguishing distinct geometrical configurations.[167] For this reason, in addition to comparing energies, these fingerprints are used in the minima hopping guided path sampling approach (see Sec. 3.5) for identifying identical configurations. Furthermore, in Chap. 5 it is demonstrated that these fingerprints can be used to obtain an educated empirical guess for the amount of energy that is needed for the interconversion of two atomic configurations.

C Stability of Hessian Eigenvectors*

The curvature along an arbitrary vector $\mathbf{d}$, evaluated at the position $\mathbf{x}_0$, is defined as

$$C_{\mathbf{x}_0}(\mathbf{d}) = \frac{\mathbf{d}^T \mathbf{H}_{\mathbf{x}_0} \mathbf{d}}{\mathbf{d}^T \mathbf{d}}, \tag{C.1}$$

where $\mathbf{H}_{\mathbf{x}_0}$ is the Hessian at $\mathbf{x}_0$. If $\mathbf{d}$ was an eigenvector $\mathbf{v}_i$ of $\mathbf{H}_{\mathbf{x}_0}$, $C_{\mathbf{x}_0}(\mathbf{d})$ would give the corresponding eigenvalue λ_i. Under the constraint of normalization, the gradient $\nabla C(\mathbf{d})|_{|\mathbf{d}|=1}$ is given by

$$\nabla C(\mathbf{d})|_{|\mathbf{d}|=1} = 2\left(H\mathbf{d} - C(\mathbf{d})\mathbf{d}\right). \tag{C.2}$$

For the case that $\mathbf{d}$ is an eigenvector, Eq. C.2 vanishes, which shows that the eigenmodes are stationary points of $C_{\mathbf{x}_0}(\mathbf{d})$.

What remains to be shown is that among all these stationary directions, only the lowest mode is a local minimum and, with it, also is the global minimum. As a consequence, rotating a slightly misaligned dimer according to its rotational force will lead back to this mode. The eigenvectors $\mathbf{v}_i$ of the real symmetric Hessian $\mathbf{H}_{\mathbf{x}_0}$ constitute an orthonormal basis, which allows to expand an arbitrary direction $\mathbf{d}$ in terms of the eigenvectors. That is, $\mathbf{d} = \sum_i c_i \mathbf{v}_i$, where the coefficients c_i are assumed to fulfill the normalization condition $\sum_i c_i^2 = 1$. Inserting this into Eq. (C.1) and using the orthonormality of the eigenvectors gives

$$C_{\mathbf{x}_0}(\mathbf{d}) = \sum_i c_i^2 \lambda_i = c_l^2 \lambda_l + c_m^2 \lambda_m + c_n^2 \lambda_n + \sum_{i \notin \{l,m,n\}} c_i^2 \lambda_i. \tag{C.3}$$

There are three cases to consider:

- m corresponds to the lowest eigenvalue: Eq.(C.3) is minimal for the set $\{c_l = 0, c_m =$

*The proof presented in this appendix has been published in B. Schaefer, S. Mohr, M. Amsler, and S. Goedecker, "Minima Hopping Guided Path Search: An Efficient Method for Finding Complex Chemical Reaction Pathways", The Journal of Chemical Physics **140**, 214102 (2014). Reprinted with permission. Copyright 2014 by the American Institute of Physics.

$1, c_n = 0, c_i = 0\}$, proving that the lowest mode corresponds to a minimum.

- m corresponds to the highest eigenvalue: Eq.(C.3) is maximal for the set $\{c_l = 0, c_m = 1, c_n = 0, c_i = 0\}$, proving that the highest mode corresponds to a maximum.

- m corresponds neither to the lowest nor to the highest eigenvalue: Without loss of generality, it is assumed that $\lambda_l < \lambda_m < \lambda_n$. Then, for some ϵ in $(0, 1]$, the coefficients $\{c_l = \epsilon, c_m = 1-\epsilon, c_n = 0, c_i = 0\}$ result in $C < \lambda_m$, whereas $\{c_l = 0, c_m = 1-\epsilon, c_n = \epsilon, c_i = 0\}$ results in $C > \lambda_m$. Therefore, all the modes that do not correspond to the lowest or highest Hessian eigenvalue are saddle points of $C(\mathbf{d})$.

D The Explosion Condition of Minima Hopping

Suitably choosing the MH parameters guarantees increasing (exploding) kinetic energies if the MH walker is stuck in some region of a PES. The explosion condition has previously been described in Ref. [199]. For the sake of being self-contained, the explosion condition is restated here.

At some time during a MH run, there are $N_{\mathrm{o,a}}$ old (previously visited) minima that were accepted and $N_{\mathrm{o,r}}$ old minima which were rejected. In total, the number of old (revisited) and new minima is given by $N_{\mathrm{o}} = N_{\mathrm{o,a}} + N_{\mathrm{o,r}}$ and $N_{\mathrm{n}} = N_{\mathrm{n,a}} + N_{\mathrm{n,r}}$, respectively. Similar, the number of accepted and rejected minima can be counted as $N_{\mathrm{a}} = N_{\mathrm{o,a}} + N_{\mathrm{n,a}}$ and $N_{\mathrm{r}} = N_{\mathrm{o,r}} + N_{\mathrm{n,r}}$. The number of same minima is denoted as N_{s}. This count is increased if MH did not escape from a local minimum. For the defintion of the parameters that are used in the following, it is referred to Sec. 3.5.

MH is stuck in some region of the PES if $N_{\mathrm{n,a}} = N_{\mathrm{o,r}} = 0$. In order to escape from such a region, the kinetic energy is required to increase on average. That is

$$\beta_{\mathrm{s}}^{N_s} \beta_{\mathrm{o}}^{N_{\mathrm{o}}} \beta_{\mathrm{n}}^{N_{\mathrm{n}}} = \beta_{\mathrm{s}}^{N_s} \beta_{\mathrm{o}}^{N_{\mathrm{o,a}}} \beta_{\mathrm{n}}^{N_{\mathrm{n,r}}} \overset{!}{>} 1. \tag{D.1}$$

Because $\beta_{\mathrm{s}} > 1$, this is equivalent to

$$\beta_{\mathrm{o}}^{N_{\mathrm{o,a}}} \beta_{\mathrm{n}}^{N_{\mathrm{n,r}}} \overset{!}{\geq} 1. \tag{D.2}$$

Taking the logarithm of this expression, and rearranging it, it can be seen that the requirement for an increasing kinetic energy is fulfilled, if

$$\frac{N_{\mathrm{o,a}}}{N_{\mathrm{n,r}}} \geq \frac{\log\left(\beta_{\mathrm{n}}^{-1}\right)}{\log\left(\beta_{\mathrm{o}}\right)} \tag{D.3}$$

An expression for $N_{\mathrm{o,a}}/N_{\mathrm{n,r}}$ is obtained by requiring the E_{diff} parameter to be unchanged on

average, such that

$$\alpha^{N_a}\alpha^{N_r} = \alpha^{N_{o,a}}\alpha^{N_{n,r}} \stackrel{!}{=} 1. \tag{D.4}$$

Taking the logarithm and rearranging gives

$$\frac{N_{\mathrm{o,a}}}{N_{\mathrm{n,r}}} = \frac{\log(\alpha_{\mathrm{r}})}{\log\left(\alpha_{\mathrm{a}}^{-1}\right)}. \tag{D.5}$$

Inserting Eq. D.5 into Eq. D.3 results in the explosion condition[18,199]

$$\frac{\log(\alpha_{\mathrm{r}})}{\log\left(\alpha_{\mathrm{a}}^{-1}\right)} \geq \frac{\log\left(\beta_{\mathrm{n}}^{-1}\right)}{\log\left(\beta_{\mathrm{o}}\right)}. \tag{D.6}$$

E Coordinates of Au_{26}

In this section the coordinates of neutral Au_{26} isomers are given in the xyz file format. All configurations that were found within 150 meV above the energetically lowest representative of each motif are provided (but not less than 5 isomers). All energies are given with respect to the putative global minimum of the respective exchange-correlation functional.

E.1 Empty Cage Motif

```
26
Distances in Angstroem, Energy =  0.0000000E+00 eV (PBE)
Au  5.43416485940463900E+00 -1.30820851574064645E-01  3.07031566513054432E+00
Au  1.19601847768324276E+00  1.98360205681964152E+00  4.22337019757839549E+00
Au  1.40986642993954714E+00  4.73942283165761058E+00  3.40582930167778386E+00
Au  8.15524709731595543E+00  3.40668890768290211E+00  6.33133345802549563E+00
Au  4.59197853661036337E+00  9.31104370501192768E+00  3.75138474184709203E+00
Au  6.07281420512988390E+00  6.92197918168196580E+00  3.78221421253049872E+00
Au  7.15067912291870655E+00  4.35413077739716670E+00  3.85550488328099261E+00
Au  5.39118737035853002E+00  3.58001897777619593E+00  6.13884882312034375E+00
Au -8.50922049298030586E-02  3.95870346496983938E+00  5.68259797611678774E+00
Au  1.36831770695561472E+00  6.31409049066508210E+00  5.81488463647088683E+00
Au  4.17905398327460542E+00  6.08205702753458688E+00  6.23297893998823582E+00
Au  5.76314305994313614E+00  8.38434548271423807E+00  6.19258828449657894E+00
Au  7.02784024239244420E+00  5.93596400535223534E+00  6.25978154341935866E+00
Au  2.94306590309192018E+00  8.57424837825023545E+00  5.97308728025667968E+00
Au  6.86482110390249822E+00  1.60583984373257826E+00  4.66795029314177601E+00
Au  4.01061009299372984E+00  1.70007757883290411E+00  4.62242245363340132E+00
Au  3.10818319701098433E+00  5.44145611914982918E+00  1.31849485097409413E+00
Au  2.64960321250207187E+00  2.73927113102565745E+00  1.91583030982887004E+00
Au  2.61079227674886694E+00  5.83263550332723396E-02  2.84710260874503884E+00
Au  4.26013255606487462E+00  8.24583367105470866E-01  6.40099442825281861E-01
Au  5.90064549972512076E+00  2.51892526227417113E+00  2.16936226510935848E+00
Au  5.91082181072272927E+00  5.25137973695907334E+00  1.53477427644385034E+00
Au  4.66870491871206816E+00  7.73120413635495041E+00  1.45212705031705069E+00
```

```
Au  2.82488793515485614E+00  7.13801011017448683E+00  3.53105012558937004E+00
Au  2.67410211977667389E+00  3.76059446137904141E+00  5.92584914448715949E+00
Au  4.50141818659673820E+00  3.50581565203903178E+00 -9.52520350349348593E-02
26
Distances in Angstroem, Energy =  6.5276168E-02 eV (PBE)
Au  5.27520961918036324E+00  9.02078163866275456E-02  2.74737822171102719E+00
Au  1.03426276792996785E+00  2.07638601378783472E+00  4.04286529274401563E+00
Au  1.28117938326206859E+00  4.94135821656471741E+00  3.45732379304528115E+00
Au  5.17297423365279840E+00  3.50165236480569275E+00  6.18041314404905151E+00
Au  4.57966454798976041E+00  8.14320346969584641E+00  1.92948010097338152E+00
Au  5.73576505307295381E+00  7.09003365715735256E+00  4.19927084294479691E+00
Au  6.97388376707249602E+00  4.48386682511379853E+00  4.01119383839647092E+00
Au  7.94126649897080927E+00  3.30399811652228248E+00  6.35158160170398745E+00
Au -2.86920049949901557E-01  3.97046679035574135E+00  5.55123410885771307E+00
Au  1.17857997284418636E+00  6.30582834515779211E+00  5.97311962621155157E+00
Au  3.96792850309718270E+00  5.99530506193701296E+00  6.46796296327830600E+00
Au  5.53965302415805816E+00  8.30975725882239757E+00  6.66368453836374108E+00
Au  6.81498507728282732E+00  5.84938363414489348E+00  6.52112397550045042E+00
Au  2.78226477619275414E+00  8.53469478995902442E+00  6.38799487479000039E+00
Au  6.65035394935837232E+00  1.62544608558926340E+00  4.58679131540571294E+00
Au  3.81679862235614076E+00  1.74849369420668355E+00  4.48103412974860404E+00
Au  2.93481606379375926E+00  5.83597806969939370E+00  1.50601107604389561E+00
Au  2.58014260235815129E+00  3.07146687339183133E+00  1.85159354773325746E+00
Au  2.46866039232921208E+00  3.15871318781987376E-01  2.47101231268071109E+00
Au  4.17381933790209203E+00  1.27706882790208853E+00  3.70799693460154456E-01
Au  5.71094469709809882E+00  2.82158130709159893E+00  2.16369304030955423E+00
Au  5.86889921641795453E+00  5.59803444820707696E+00  1.80020211824473719E+00
Au  4.69007063015315495E+00  6.67504290223566876E+00 -4.90623006422629737E-01
Au  2.86277035390330381E+00  7.31603508653740242E+00  3.91866453862428932E+00
Au  2.45727524767964889E+00  3.72059406325793285E+00  5.91920500549482753E+00
Au  4.43849008189483207E+00  3.95335596268916856E+00 -1.24094693880901427E-01
26
Distances in Angstroem, Energy =  6.5624722E-02 eV (PBE)
Au  2.66300212209412579E-01  3.49676414641646494E+00 -5.26074888062340817E-01
Au  1.95016966383986667E+00 -2.36293980201290665E+00  6.04139055692967220E-01
Au -5.21204276080870166E-01  3.25625733523804239E+00 -3.26524647143550029E+00
Au  2.38309119662458047E+00 -2.04975376424255940E+00  3.42765167138110538E+00
Au  1.94174914988092673E+00  2.06565724655522232E+00 -2.18227582482543436E+00
Au  3.26988901225287842E-01 -3.62075940038696187E+00  2.54939009883093171E+00
Au -2.27343154270093573E+00  1.55679267413446420E+00 -1.82684210750328524E+00
Au  1.43934779719211581E+00  6.03894072342694099E-01  3.03871139867165851E+00
Au -2.19903334011001528E+00  4.37936321222577440E+00 -1.40841131847676349E+00
Au  3.62739703278095416E+00 -3.52106056994654915E-01  1.61462048751376552E+00
Au -1.85311784901716470E+00 -1.20296248909452097E+00 -1.77107341439279442E+00
Au  2.31141080787720582E-01  2.94638841578587174E+00  2.22145506244491342E+00
Au  2.38532291179296552E+00 -2.74151103314812239E-01 -3.77162617372460929E+00
Au -2.08895570503318684E+00 -4.63966437011247645E+00  1.76716508023365582E+00
```

```
Au  1.2930162896794532E+00  2.0639125404366724E+00 -4.9049101594387245E+00
Au -1.9225611158809749E-01 -1.3173020097563226E+00  4.1969379964944808E+00
Au -3.2435096999390794E-01  3.6614031972954758E-01 -3.3835766285316331E+00
Au -2.1019364481320952E+00 -1.9255968696932557E+00  2.3363732538676354E+00
Au -5.8097897835022350E-01 -3.1809038002299692E+00 -1.2289970016560109E-01
Au -2.1402872411968201E+00  2.6186976042842183E+00  7.7088607168456902E-01
Au  2.1502433745256111E+00  1.7875663890652331E+00  5.5470266837109405E-01
Au -3.3317683177072452E+00 -2.7750944041392218E+00 -3.7077320824824810E-02
Au  9.0102764908286326E-01 -1.8406158364368777E+00 -1.9752193986720821E+00
Au -1.3749503714569653E+00  7.1294015497800423E-01  2.6869189230724984E+00
Au  3.0404802056383384E+00 -3.1095303499538785E-01 -1.0683378213590232E+00
Au -2.9540043139005476E+00 -1.5711697884548267E-03  4.7461945915820630E-01
26
Distances in Angstroem, Energy =  7.5209165E-02 eV (PBE)
Au  2.5530652069752041E-01  3.7059904634653909E+00 -5.8141924771863029E-01
Au  1.9586867767759937E+00 -2.7172856152099487E+00  5.1484949196599628E-01
Au -4.9336833969698635E-01  3.2218486817599734E+00 -3.2721735681904578E+00
Au  2.7623818071322410E+00 -2.9153796835206549E+00  3.2555981305919630E+00
Au  1.8748346029090055E+00  2.0115571777898751E+00 -2.0524850528317269E+00
Au  2.9079383049654634E-01 -3.8598723028006265E+00  2.5213123479488813E+00
Au -2.3472965483831145E+00  1.5403178004043128E+00 -1.9008790737750443E+00
Au  1.4976460081994607E+00  1.3969974571694255E+00  3.4462173266104057E+00
Au -2.2712184836009808E+00  4.3813307597765956E+00 -1.5381958617411244E+00
Au  2.8783745037952344E+00 -4.6955052603566216E-01  1.9301543755781327E+00
Au -1.8901821839322248E+00 -1.2328875803618675E+00 -1.7614318792409394E+00
Au  6.7052313088545770E-02  3.3737723935638986E+00  2.1726449521400539E+00
Au  2.4364508113559613E+00 -3.3662772470710112E-01 -3.5924436630180239E+00
Au -2.2095149654079691E+00 -4.5853266469000689E+00  1.7771943313395580E+00
Au  1.3608714317315591E+00  1.9979214546803452E+00 -4.8187723637989040E+00
Au  5.6816590970781089E-01 -1.2187611619850687E+00  3.4977724982163654E+00
Au -2.8114622114380139E-01  3.0962029433614519E-01 -3.3046892788442404E+00
Au -1.8219343937204380E+00 -1.8392587004931968E+00  2.2538887113802075E+00
Au -6.8797529088238230E-01 -3.2606907933620266E+00 -1.5848433938495438E-01
Au -2.1434626904651735E+00  2.6736356905430605E+00  6.5616951493032238E-01
Au  1.9570802407307655E+00  1.8967747030009151E+00  6.9421736239677079E-01
Au -3.4337997587194069E+00 -2.6787929493830633E+00  5.1171207777315222E-02
Au  8.5101817283424808E-01 -1.9244741734019204E+00 -1.9444159390115721E+00
Au -1.0760870715508464E+00  7.9542509536535433E-01  2.4708778195994343E+00
Au  2.8156681329092930E+00 -3.1713549506207612E-01 -8.0378021184301829E-01
Au -2.9183451148688366E+00  5.0851381361789516E-02  4.8710240892801721E-01
26
Distances in Angstroem, Energy =  1.0048964E-01 eV (PBE)
Au  1.0201047120219062E+00 -2.6021499172289850E+00  3.4398141925994341E-01
Au -4.3790984782208335E+00 -2.5992735892981380E+00 -8.1789997303007589E-01
Au -3.3312257956530189E+00 -6.5334811380567112E-02 -7.1863386472508395E-01
Au  1.4346295601674378E-01  2.4938142911017396E+00  1.2697525966437977E+00
Au -7.1431340794477272E-01  2.7979756739587405E+00 -3.2549786154423095E+00
```

```
Au  3.2585541698465 2813E+00  9.67135307948087553E-01  1.74801230908370409E-01
Au  3.7660661326570 1595E+00 -1.81160460192659700E+00  3.71447433117360670E-01
Au  1.4404822963331 3407E+00  1.87266874218833679E+00 -1.65960685204662983E+00
Au -2.4433718918140 8743E+00 -4.30414610089033312E-01  4.00456855386849853E+00
Au -2.1337462864306 9419E+00  9.57940285061563590E-01  1.63535839422865714E+00
Au  1.9228181290157 3507E-01  4.22986705246353001E+00 -1.00092795564231651E+00
Au  1.9199669668690 06780E+00 -7.90868832156471857E-01  2.29642012958611241E+00
Au  2.4687399780362 2602E+00  1.94424643246484052E+00  2.77279956975393915E+00
Au  4.5706951194758 8403E+00  1.02684992025840367E-02  2.38682905104611143E+00
Au  1.3450968210624 5532E-02  8.49589638146341697E-01  3.56847912310312854E+00
Au  2.7365731831095 5855E+00 -3.46047605077928777E+00 -1.64715618894460825E+00
Au -1.6692660656327 9867E+00 -2.60113684508751541E+00 -2.43503847647814742E-01
Au -2.8765636311169 6312E+00  9.74743357292860213E-01 -3.24361912599567814E+00
Au -2.4763520960668 1439E+00 -1.82301554744496519E+00 -2.84054592350056234E+00
Au  1.2758656974983 9007E-01 -2.76237973344331289E+00 -2.41153594046996522E+00
Au  2.1192073383137 3020E+00 -7.71868616019832365E-01 -1.75146117445937510E+00
Au -7.0642044663031 4780E-01 -1.60970540863913003E+00  2.20587300319860669E+00
Au  2.5318396377647 7405E+00  3.62654942112999068E+00  4.67162678109399376E-01
Au -1.7421780843918 5062E+00  2.16476199172329276E+00 -7.84846259189189488E-01
Au -3.5100120278915 3011E+00 -1.56642274953564775E+00  1.67344862307091513E+00
Au -3.2646633133351 4437E-01  5.09062034858507977E-03 -2.79620608479964927E+00
26
Distances in Angstroem, Energy =  1.1623052E-01 eV (PBE)
Au -4.4631195314222 7547E-01  2.79445451197255812E+00 -1.76091720788761541E+00
Au  1.4341743034852 5876E+00 -2.36818141591355724E+00  1.42925147889884530E-01
Au -1.4588415339728 2978E+00  8.60346274256625132E-01 -3.51648331261373004E+00
Au  8.2239960549864 8545E-01 -2.42612791202700118E+00  2.95351595134002576E+00
Au  1.8220255503608 4357E+00  4.22543152150794299E+00 -2.59884405368868920E+00
Au -9.1277385611243 4667E-01 -3.55304881649481885E+00  1.13455832386341360E+00
Au -2.8555754352510 4718E+00 -1.22859808324235886E+00 -2.43329635502163599E+00
Au  2.0130984093823 6118E+00  1.30505775390631351E+00  2.01507084278694526E+00
Au -2.3157633766677 8009E+00  1.02267127362836074E+00 -8.19549593723873704E-01
Au  3.2168168601834 8767E+00 -1.30636137871087388E+00  1.96949097773479709E+00
Au -7.9327697523041 1407E-01 -1.94898452069108052E+00 -4.16801135205265361E+00
Au -3.2881785125043 6055E-01  1.11608968165618960E+00  3.49983372042308849E+00
Au  2.0992302404738 3810E+00  1.40507057143305270E+00 -2.70283217624115268E+00
Au -1.9436534555197 7775E+00 -1.79446417891915422E+00  2.97165014328848676E+00
Au  3.1928039919289 8412E-01  2.77826584071330229E+00 -4.47531596512888541E+00
Au  2.0174951824563 5695E+00 -2.95150591069201995E-01  4.28424051499200242E+00
Au  8.9433868691429 1081E-01  7.59640951011467114E-02 -4.86693844739695081E+00
Au -4.0461539698364 5181E-01 -1.14303980940943672E+00  5.21753551849233066E+00
Au -7.4825740115354 3888E-01 -2.82691988177655551E+00 -1.50441152997480687E+00
Au -4.6325403885667 8541E-01  2.20401229835384216E+00  9.58409736130942380E-01
Au  1.9884955253139 1760E+00  2.83272691405218779E+00 -2.31625928341882614E-01
Au -2.6532631496144 4432E+00 -1.53680608259449381E+00  2.84661591752299992E-01
Au  1.3984848374139 8089E+00 -1.29296114173742249E+00 -2.45876069569671696E+00
Au -2.7575279720661 1966E+00  2.73694867334427827E-01  4.62089893093525284E+00
```

```
Au  2.7326930311258839 1E+00  5.2872710235485933 7E-02 -3.9921398546251107 8E-01
Au -2.6766002359801102 1E+00  7.7398549843484498 7E-01  1.8834092036017828 8E+00
```

E.2 Hexagonal Motif

```
26
Distances in Angstroem, Energy =  1.2700530E-02 eV (PBE)
Au -2.1031411807086750 6E+00  2.2624686241633660 3E+00 -1.7339010422485354 2E-01
Au  2.2356017853215282 0E+00 -2.2046608047683733 4E+00  2.5255838707915290 2E-01
Au -1.7317381059108745 1E+00  1.9208490645416149 7E+00 -2.9306259085014954 0E+00
Au  2.6014279245570310 9E+00 -1.9705676984931621 6E+00  3.0677614747609771 6E+00
Au  1.7392795794774865 1E-01  3.3018865500277163 6E+00 -1.3478151080765796 7E+00
Au  1.8840269456841449 9E-02 -2.4696356975080764 3E+00  1.9152654782207614 3E+00
Au -2.6702048635778683 5E+00 -4.7924475219690776 4E-01 -1.4862460289295060 8E+00
Au  2.7756331541979495 4E+00  4.3085603537814753 9E-01  1.4984115986350097 2E+00
Au -4.2939971165854817 8E+00  1.7932085288165191 5E+00 -1.9203583844041078 9E+00
Au  4.8038925593872026 1E+00 -1.6162586891031731 6E+00  1.5088011876921489 9E+00
Au -6.3943029432178843 9E-01 -7.8199576180977603 5E-01 -3.3046402787716644 4E+00
Au  5.3545274047972890 6E-01  1.1666460064448099 0E-01  3.1407449118564052 5E+00
Au  3.7525031715239762 8E+00  8.2625103550806922 0E-01 -3.2482814009621665 5E+00
Au -4.3126409141751409 5E+00 -5.2042421395000770 8E-01  3.0158955821053803 3E+00
Au -7.4687995661073558 5E-01  1.1190595393341267 2E+00 -5.3434790322102152 5E+00
Au  5.5055785282411551 2E-01 -2.0847098300551691 9E+00  4.8407896634695868 3E+00
Au  1.1060402369530237 8E+00  1.5818985155578129 1E+00 -3.4005327921551749 0E+00
Au -1.8719393169303761 8E+00 -1.4261396007993496 5E+00  3.8085180370384330 6E+00
Au -3.4857774740338642 9E-01 -1.8413153443439109 8E+00 -7.6216478479047056 2E-01
Au  3.7899118074853926 9E-01  1.7907403804689352 4E+00  9.9149218488568990 9E-01
Au  2.3991662920667913 3E+00  1.7154894579864581 3E+00 -9.1822047293280284 5E-01
Au -2.4849365846191320 3E+00 -1.6285856584882969 5E+00  1.0327128710825321 8E+00
Au  1.8727482076436110 5E+00 -1.0313738712351108 2E+00 -2.2380887841750718 4E+00
Au -1.9757376502291785 8E+00  1.0083657570797119 9E+00  2.2867753601772919 9E+00
Au  4.2637135894927107 3E+00 -4.2580939982983090 4E-01 -8.5226658272906230 5E-01
Au -4.2892731915281778 8E+00  6.1298323307417845 7E-01  5.6638292585982108 5E-01
26
Distances in Angstroem, Energy =  1.0541089E-01 eV (PBE)
Au  8.0929994435271073 2E+00  1.0891565247095154 1E+00  4.8410812551621518 2E+00
Au  9.1060495348530405 8E+00  3.5414088673990136 2E+00  5.5057819854216889 5E+00
Au  5.4500843715517151 0E+00  5.9968506811941404 4E+00  5.2645683840756598 6E+00
Au  4.9422229154001851 0E+00  3.2109667961204002 6E+00  1.9231118845595038 8E+00
Au  4.5827736648676724 6E+00  2.6138653164625296 2E+00  7.9249357395291726 3E+00
Au  1.4036459765566684 5E+00 -8.2217678307616992 6E-01  6.8031694697481137 6E+00
Au  8.1398849026905466 5E+00  5.8538596583444153 5E+00  4.3080676565452273 9E+00
Au  7.4766964625408176 6E+00  2.4768031301191757 4E+00  7.5127573391550983 6E+00
Au  1.6432744217660268 9E+00  3.3208024190361684 6E+00  4.2341894503851760 2E+00
Au  2.2071090010908673 8E+00  1.9087298957167935 2E+00  6.6264670878160183 4E+00
Au  3.0307475630967512 0E-01  1.9425118390240274 6E+00  2.1633238207615042 0E+00
Au  3.5623612600102814 2E+00 -1.4947816697330706 9E-01  5.0790621003858644 8E+00
```

```
Au  2.19370435833961430E+00  3.77028439437585794E+00  1.41793405020285967E+00
Au  5.03509263741601565E+00  5.31903598553586221E+00  8.01713892844983000E+00
Au  3.02581308844482511E+00  1.32632234878559352E+00  2.73517882677616386E+00
Au  6.12379330894792417E+00  5.75613640129463988E+00  2.49350561273197435E+00
Au  4.07763753745966095E+00  5.64124704342515937E+00  6.89010060333060226E-01
Au  3.45919669611950154E+00  5.31106623689785096E+00  3.45663514374169445E+00
Au  3.73123534484300379E+00 -1.08384614927334269E-01  8.02797312851852318E+00
Au  8.49803112052368359E-01  5.53777596852822795E-01  4.48268037315854873E+00
Au  7.53109880719996294E+00  5.20277998425564991E+00  6.97920914917446034E+00
Au  5.57810703757822335E+00  1.19661379486484964E+00  3.72193501731900467E+00
Au  3.38989629278731730E+00  4.33261217786055752E+00  6.06289095567320313E+00
Au  6.17198943909909481E+00  6.77045573413470825E-01  9.13725257894995480E+00
Au  5.95712112733713628E+00  3.40198734177378614E-01  6.35892121614032657E+00
Au  7.35851430120983707E+00  3.30632208511038383E+00  3.26944518528531125E+00
26
Distances in Angstroem, Energy =  1.0678305E-01 eV (PBE)
Au -2.23419643671283019E+00  2.20826779647641080E+00 -2.96029662125061710E-01
Au  2.23414011787022915E+00 -2.20840266619023007E+00  2.95944413718548627E-01
Au -2.53517771692891980E+00  1.81442971139867337E+00 -3.12213010836283500E+00
Au  2.53524303094004955E+00 -1.81456975270871590E+00  3.12218982672172629E+00
Au -1.66796744460772257E-02  2.48676168342740400E+00 -1.94023055939347810E+00
Au  1.65751723628196172E-02 -2.48679938662712852E+00  1.94014065867375463E+00
Au -2.75027768602491030E+00 -5.12574796709129865E-01 -1.46433788453818781E+00
Au  2.75031222570140477E+00  5.12396545012904214E-01  1.46425871622260062E+00
Au -4.78636039943838565E+00  1.66377141165546227E+00 -1.67990888577790964E+00
Au  4.78631204812979583E+00 -1.66367147354872835E+00  1.67984903970581478E+00
Au -5.25904308829192813E-01 -2.35441687085249124E-01 -3.10084763044676492E+00
Au  5.25933440584760770E-01  2.35248945199778958E-01  3.10079257292040333E+00
Au  4.27974297176518981E+00  5.83224741598997398E-01 -3.13947448944533170E+00
Au -4.27959259970075934E+00 -5.82460752410723037E-01  3.13977772223401219E+00
Au -5.31884452513812223E-01  2.11504789896737000E+00 -4.87412695577317923E+00
Au  5.31830431192404252E-01 -2.11532008736430388E+00  4.87403345579882519E+00
Au  1.79139928653097513E+00  1.30638850399113404E+00 -3.81500345636292648E+00
Au -1.79145987237906601E+00 -1.30648713874859146E+00  3.81499975938436275E+00
Au -3.18261986733422042E-01 -1.69887128194880943E+00 -7.29080176723343798E-01
Au  3.18212865122961486E-01  1.69881943041657513E+00  7.29066233039262990E-01
Au  2.47636026313790625E+00  1.65271552890583817E+00 -1.05098754944040618E+00
Au -2.47639310205930485E+00 -1.65259558944150320E+00  1.05108777953702481E+00
Au  1.97061448620512114E+00 -1.06688758361574298E+00 -2.22025145368308419E+00
Au -1.97060273051235746E+00  1.06703426936787937E+00  2.22038117945720526E+00
Au  4.32896180404013720E+00 -5.12313419237095058E-01 -6.95239699612273609E-01
Au -4.32884717730475899E+00  5.12289149217528883E-01  6.95127154271256376E-01
26
Distances in Angstroem, Energy =  1.0891814E-01 eV (PBE)
Au  8.24336046637321296E+00  1.51051513991974917E+00  4.27707639360196623E+00
Au  7.96228876556518728E+00  2.59163278557710308E+00  6.78713792862220178E+00
Au  5.56191699980282461E+00  5.88245548417216568E+00  5.29225972088569119E+00
```

```
Au  4.7644796355331928 7E+00  3.2792829818387470 0E+00  1.6555477106508407 2E+00
Au  4.3531174441685980 3E+00  2.7895911192270621 5E+00  8.1555814863073550 2E+00
Au  1.8401061009319787 9E+00 -1.0171785028820881 1E+00  6.6477233827574648 8E+00
Au  7.8475869596395382 3E+00  4.3447102261382362 1E+00  4.4966519557640634 0E+00
Au  6.8856842979452395 1E+00  3.6992521763240016 1E+00  9.0788549904214797 0E+00
Au  1.6644067769920940 5E+00  3.3013513946304033 9E+00  4.2771408464787805 7E+00
Au  2.3082071425234618 2E+00  1.8288644289968230 0E+00  6.6028764703462563 5E+00
Au  3.0131265151129149 6E-01  1.9215429829719745 7E+00  2.2265976456826361 8E+00
Au  3.7797842030202719 6E+00 -4.0754155603343819 6E-02  4.9016745165116066 8E+00
Au  2.0449017803590998 6E+00  3.9507323481876466 3E+00  1.5053730788983417 1E+00
Au  5.0080794601552662 8E+00  5.5218219220047970 5E+00  8.0255372732862895 2E+00
Au  3.0024310429397265 1E+00  1.3723176834250039 3E+00  2.6199467259021957 9E+00
Au  6.1986740061137473 0E+00  5.5791461733148315 4E+00  2.5968616189946689 4E+00
Au  4.0457537369897806 9E+00  5.8339443401648543 0E+00  8.3870604151446603 9E-01
Au  3.4864968878041944 2E+00  5.3116969758718104 4E+00  3.5491563899015559 1E+00
Au  4.1536573718273723 8E+00 -2.6823476259007401 2E-02  7.7665510015773753 2E+00
Au  9.8091163752245202 9E-01  4.7428983331321183 9E-01  4.4698822942092881 8E+00
Au  7.7003241219908682 4E+00  5.3711014195650816 9E+00  6.9929567321443437 7E+00
Au  5.6564613680014552 7E+00  1.2376354138537681 8E+00  3.3376361107887051 5E+00
Au  3.3237635510120981 9E+00  4.3983314851939519 4E+00  6.1444108601788736 1E+00
Au  6.5143386357580821 4E+00  1.0101939000463593 4E+00  8.7238624465098482 1E+00
Au  6.2983351759077681 5E+00  4.7619204341831938 8E-01  5.9714632236811251 4E+00
Au  7.4667995796097832 4E+00  3.0065091332624134 2E+00  2.0947595543826729 2E+00
26
Distances in Angstroem, Energy =  1.0999619E-01 eV (PBE)
Au -2.3899316211278405 2E+00  2.1157019813899808 9E+00 -3.0003328970618514 7E-01
Au  1.7293798743752644 1E+00 -3.4856766956150235 8E+00 -3.2106432261695505 6E-01
Au -2.4462870973799808 5E+00  1.8598499503860519 4E+00 -3.1630607124507870 2E+00
Au  2.9212117671377888 5E+00 -1.9181049052383469 8E+00  1.8052840127626268 4E+00
Au -1.0949466734594527 3E-02  2.4213971410951993 2E+00 -1.7284542442938128 9E+00
Au  1.5965629375595330 5E-01 -2.5728914522955257 4E+00  1.7575462854137671 2E+00
Au -2.7189924195329822 1E+00 -4.8579107170388558 0E-01 -1.4584278807412183 2E+00
Au  2.3034063351016786 9E+00  8.7137412168829275 8E-01  2.1990559657428154 8E+00
Au -4.7655328704872914 6E+00  1.4172565629447229 1E+00 -1.7818832371865123 1E+00
Au  4.4675449008544712 0E+00  1.7173029665964278 9E-01  6.9229866370132853 2E-01
Au -4.2825256888290980 1E-01 -1.3548526162814514 3E-01 -3.0431139076477080 6E+00
Au  7.9285788417985156 3E-02  1.3213105954946071 7E+00  3.7149419452590177 7E+00
Au  4.2860076070272894 5E+00  2.4718840465954647 3E-01 -2.1169447029976136 1E+00
Au -4.0287511162722022 8E+00 -8.5114426888563976 5E-01  3.2229251161106748 9E+00
Au -1.3130474851889306 6E-01  2.3193695812702865 8E+00 -4.5773433077799499 1E+00
Au  1.3678438817862199 6E+00 -1.2301267423330815 0E+00  3.9197497229752253 9E+00
Au  2.0648182718241017 1E+00  1.2684105208116311 9E+00 -3.3327042886385132 0E+00
Au -1.3704783485424638 6E+00 -1.2653034603102668 6E+00  3.7504972249855184 1E+00
Au -4.9532851561845836 5E-01 -1.9773693377957961 6E+00 -9.2073650842221721 5E-01
Au  6.9684061730349952 7E-02  2.0659958696838893 7E+00  1.0134572419693572 7E+00
Au  2.3623080984099678 0E+00  1.6081938375217186 3E+00 -5.0212511932347170 3E-01
Au -2.3687886574752963 2E+00 -1.6662937858411246 4E+00  1.0807705129864786 4E+00
```

```
Au  1.88424809031554186E+00 -1.32710142036030865E+00 -2.17271108419813386E+00
Au -2.22446220317960641E+00  1.08632707908358817E+00  2.26686410227420687E+00
Au  4.11843914419511670E+00 -2.16759150124659339E+00 -7.24682088852542372E-01
Au -4.43477448117918627E+00  3.08773960564606931E-01  7.19893900674620557E-01
26
Distances in Angstroem, Energy =  1.2240064E-01 eV (PBE)
Au -2.31579152710879743E+00  2.17356605686846205E+00 -2.91437775549621469E-01
Au  2.13507120713556153E+00 -2.69329160598603279E+00  1.53973717022866302E-01
Au -2.47887131722596754E+00  1.80744762054931951E+00 -3.16395304018374990E+00
Au  2.68235765606789700E+00 -1.41751084869645094E+00  2.63836444111614110E+00
Au -8.17207444539967860E-03  2.48608327980930399E+00 -1.82640053425450621E+00
Au  7.12153136010589966E-02 -2.46958754422450388E+00  1.90165893842318789E+00
Au -2.71458776781659639E+00 -4.39529509701558185E-01 -1.43000024371800238E+00
Au  2.63974025373378884E+00  1.44189618523785534E+00  2.04368898593259019E+00
Au -4.75887215863513813E+00  1.77688146256970003E+00 -1.69743968791767985E+00
Au  3.95499976347855808E+00 -3.25290411668582802E-01  4.48800706410212136E-01
Au -4.26081660579460730E-01 -1.71128398466472820E-01 -3.01427666663330340E+00
Au  4.34976404312000853E-01  2.89405898652175642E-01  3.17875019614703724E+00
Au  4.29572821918349312E+00  3.22838548443172779E-01 -2.29003569190481038E+00
Au -4.19663429815464895E+00 -7.18932491012644448E-01  3.15115235528474580E+00
Au -2.38798725015040431E-01  2.18015850443594861E+00 -4.67585977433323130E+00
Au  8.89655443086519560E-01 -2.11913193418960466E+00  4.59724078624249088E+00
Au  2.02746559867423448E+00  1.27687698230755808E+00 -3.47094228877387190E+00
Au -1.69469039737665805E+00 -1.47020914339944397E+00  3.88060351944341786E+00
Au -3.81637749133999615E-01 -1.83211600701115263E+00 -7.76194921869931442E-01
Au  1.51577616436927631E-01  1.91220180561624353E+00  8.96923472244160358E-01
Au  2.44437493417137031E+00  1.79484336170449943E+00 -7.13164891828547809E-01
Au -2.44385060323573144E+00 -1.63877111146178267E+00  1.07342096666612230E+00
Au  1.93881721674934471E+00 -1.31626387711445525E+00 -2.24347245416651431E+00
Au -2.06843700126858687E+00  9.71523203082855558E-01  2.26978819617815875E+00
Au  4.45111078611367272E+00 -2.30127523505765907E+00 -1.33429127335168984E+00
Au -4.39066513274838677E+00  4.79315208713232577E-01  6.93102963374319425E-01
```

E.3 Filled Cage Motif

```
26
Distances in Angstroem, Energy =  1.3818823E-01 eV (PBE)
Au  5.43919397585806319E+00 -9.09072390477962261E-02  3.07383675954036617E+00
Au  1.20538895604917395E+00  1.98540494023659209E+00  4.23092866931243172E+00
Au  1.37945053643191851E+00  4.73767778828612673E+00  3.39601308609042007E+00
Au  4.23270072297214561E+00  4.49585637938876026E+00  3.74985391305335236E+00
Au  4.60255885908325268E+00  9.27834825429571630E+00  3.73996573488387263E+00
Au  6.11952253518413691E+00  6.90546639777303639E+00  3.77067604636169840E+00
Au  7.13749765103959533E+00  4.29201571188662712E+00  3.99919535136850657E+00
Au  5.48359878320219263E+00  3.53430608819617387E+00  6.24841462764796240E+00
Au -1.18688021314297065E-01  3.97387347813617930E+00  5.63692730248423324E+00
Au  1.37423925359669519E+00  6.30275684735076513E+00  5.81681545961678470E+00
```

```
Au  4.17126904784159258E+00  6.05016764596717227E+00  6.37084785793531783E+00
Au  5.76842608321427708E+00  8.34399932691616897E+00  6.16846696503272174E+00
Au  7.00643606018410559E+00  5.85566083936124393E+00  6.30469447548372397E+00
Au  2.93825066109078081E+00  8.54194453928390196E+00  5.92823719363573076E+00
Au  6.84197181018285860E+00  1.62545707157525299E+00  4.75800423596938948E+00
Au  3.99967890750286115E+00  1.63120705222569784E+00  4.74880654007239400E+00
Au  3.07281840954632646E+00  5.47439998381009119E+00  1.21269742526785218E+00
Au  2.61340733480417242E+00  2.74903573658335532E+00  1.85269084732242906E+00
Au  2.61003948746321379E+00  1.07714944669662804E-01  2.82719499609169400E+00
Au  4.27669937724136240E+00  8.46553672163660065E-01  6.42247991495396797E-01
Au  5.94801616450523518E+00  2.52424440420352214E+00  2.16426121540817196E+00
Au  5.94156767225599403E+00  5.25633517950567786E+00  1.49971477231261652E+00
Au  4.66835699177035401E+00  7.73784203696781514E+00  1.42269706641563998E+00
Au  2.78636254701597696E+00  7.16103275280472662E+00  3.47629902909223665E+00
Au  2.64249722763541550E+00  3.72335839607612584E+00  6.03158954132910008E+00
Au  4.50247723637978314E+00  3.51135959684998467E+00 -1.32161016073419374E-01
26
Distances in Angstroem, Energy =  1.5479731E-01 eV (PBE)
Au  1.06166250330384027E+00 -2.62079375929920610E+00  3.84563917589047621E-01
Au -4.16960552998869627E+00 -1.42628350318094288E+00 -1.21510870777389290E+00
Au -2.69601851074639054E+00  8.55175912714511322E-01 -5.74564886004856223E-01
Au  1.23907331700258250E-01  1.86268880171613527E-01  2.31394845199284105E-01
Au -1.44786993735539715E+00  2.46090280370205194E+00 -2.70536994992434510E+00
Au  3.03215844306558280E+00  1.12442903275808304E+00  2.18622924721711520E-01
Au  3.78536605337528220E+00 -1.62056077515893726E+00  4.26913997464257455E-01
Au  1.22786644937059441E+00  1.88881712540487801E+00 -1.80566758012409889E+00
Au -2.52134293268816290E+00 -5.89044643884301666E-01  4.01810165002210606E+00
Au -1.85739837799904328E+00  1.22334003851526307E+00  2.06211189307264098E+00
Au  2.79440771427906520E-01  4.45153849522025968E+00 -1.74026301467272693E+00
Au  1.94716105028035980E+00 -8.41094564861450400E-01  2.45596766027286062E+00
Au  2.22503858256358411E+00  1.96896516731071602E+00  2.80860310900914500E+00
Au  4.48381034796091260E+00  2.29694691200973855E-01  2.38252468279062102E+00
Au  1.27635281490802333E-01  5.59391319021807498E-01  3.95942337327056926E+00
Au  2.88607640549208089E+00 -3.37981405821277914E+00 -1.55506599323676364E+00
Au -1.62264401421404370E+00 -2.29339109841832034E+00 -3.91359186100996848E-01
Au -3.16322994989830120E+00  3.85282089798667016E-01 -3.28773824992083208E+00
Au -2.28454383416204010E+00 -2.30752125018667842E+00 -3.13816859369456491E+00
Au  2.97644423023066873E-01 -2.92864789986837026E+00 -2.41176991380153050E+00
Au  2.10552746669161328E+00 -7.31119642936737923E-01 -1.76219255357905524E+00
Au -7.41364307053838267E-01 -1.67499063516824553E+00  2.24678502046667239E+00
Au  1.77365641876907465E+00  3.54315150264510370E+00  5.64006222252514733E-01
Au -9.37824826752506779E-01  2.88620992760723594E+00  1.64187957389331229E-01
Au -3.45616522794617920E+00 -1.15910625228123565E+00  1.43380839733417642E+00
Au -4.58944079709587005E-01 -1.90798902612957261E-01 -2.76974702201971512E+00
26
Distances in Angstroem, Energy =  1.8819427E-01 eV (PBE)
Au  4.58839423795187606E+00  1.16587354946236441E-01  3.43964748857297398E+00
```

```
Au  6.06941422176019429E+00  5.08111604301120856E+00  7.36545288879871585E+00
Au  4.46771492106012413E+00  2.49152025183535475E+00  5.25404623875071142E+00
Au  1.80445250129742618E+00  3.77343506992184352E+00  1.91602928776510395E+00
Au  7.60837613077039521E+00  3.48316197654822446E+00  9.02855407698907975E+00
Au  4.17728287796594167E+00 -1.73388991820453775E+00  5.57214403658835611E+00
Au  1.27275152079288145E+00  5.02327306322657918E+00  6.57974256584609041E+00
Au  7.35904487221297998E+00  2.75490247561186186E+00  6.31878775962718109E+00
Au  1.56708882130169047E+00  3.02354215978769458E+00  4.61620612881713388E+00
Au  4.94256784569084040E+00  2.78387189970126192E+00  8.71334163062946132E+00
Au  2.34041828602114954E+00  4.81773362231163182E-01  5.39621415872778964E+00
Au  6.06765390182845454E+00 -1.69242449846084564E+00  7.58712206921175358E+00
Au  1.67657030564239662E+00  5.73827111066363127E+00  3.90111852499823986E+00
Au  6.34122701799514399E+00  4.75317987368553219E+00  4.60030149747488259E+00
Au  1.97847558919570576E+00  1.07081221623481770E+00  2.57586168376158753E+00
Au  6.73002044155653056E+00  2.08135899092227650E+00  3.64165517404169536E+00
Au  4.14927312677172822E+00  5.22034934602065981E+00  2.88926070390163225E+00
Au  6.57813589455982406E+00  4.20080957835287006E+00  1.87502261742089993E+00
Au  6.94230329075823693E+00  8.13878172654778131E-01  8.43342258350076435E+00
Au  2.29325630050166973E+00 -1.52908729940469001E+00  3.50356974428903412E+00
Au  3.90232355210446702E+00  5.59333222977753053E+00  5.66907784254396763E+00
Au  4.35310290148040924E+00  2.50938603005535832E+00  2.09851442235968744E+00
Au  3.38120498906185984E+00  5.02855354971932922E+00  8.40260684653488532E+00
Au  2.52010146371607879E+00  2.60727698004965314E+00  7.27039208307751750E+00
Au  4.21271146984396783E+00  3.90330405368356614E-01  7.53426921199546307E+00
Au  6.42271927469699477E+00  1.94905408207465297E-01  5.61407885828795994E+00
26
Distances in Angstroem, Energy =  1.9882961E-01 eV (PBE)
Au  5.33458331836968291E+00 -2.44704885041589423E-02  2.81923968262955604E+00
Au  1.17542500234568648E+00  2.08058505958248663E+00  4.19475610710120517E+00
Au  1.42765095122868702E+00  4.88636253461964820E+00  3.60725705030393540E+00
Au  4.30721350338613451E+00  4.49965468711803851E+00  3.88289641398296981E+00
Au  4.88397734223305235E+00  7.92858495642216266E+00  1.98373641747560159E+00
Au  4.93641825112136878E+00  9.13500979956547710E+00  4.39546329568375871E+00
Au  7.18288317898935613E+00  4.20930768553946066E+00  4.11492045198787793E+00
Au  5.50360468520353319E+00  3.32979331787880994E+00  6.27042020572064018E+00
Au -9.19817829479320642E-02  3.99925217331493199E+00  5.73982960542534926E+00
Au  1.47123876133029841E+00  6.25918923444814279E+00  6.15992056114839848E+00
Au  4.27987169281986635E+00  5.87176951389584101E+00  6.57419916918743574E+00
Au  5.93175830678587079E+00  8.08887742693065448E+00  6.77599656387986293E+00
Au  7.06453625354797943E+00  5.61464237832813318E+00  6.47291068099721301E+00
Au  3.11534130375505924E+00  8.42999154202798451E+00  6.46387897486036422E+00
Au  6.77553122632107385E+00  1.46138330431023711E+00  4.67504428631983160E+00
Au  3.95309497735679871E+00  1.56359481978274184E+00  4.69456177921374351E+00
Au  3.09777880329346722E+00  5.71129042202417381E+00  1.55093596529286781E+00
Au  2.49543818002057405E+00  2.97301364339748675E+00  1.81405575407057373E+00
Au  2.50264106857106317E+00  2.89153763429172317E-01  2.62181479611503665E+00
Au  4.18179331947829169E+00  1.14200753314530412E+00  5.13798872565100817E-01
```

```
Au  5.95335931345012970E+00  2.62334807904201162E+00  2.14324334167566866E+00
Au  6.12882001424035128E+00  5.37929392581197696E+00  1.85614418909405265E+00
Au  4.88751678213512530E+00  6.49882782603121800E+00 -4.05432417156858016E-01
Au  3.01860087975246572E+00  7.20021380277449907E+00  3.95463938561243644E+00
Au  2.66665302751066990E+00  3.61105608934781142E+00  6.12596406566955221E+00
Au  4.45999000970124904E+00  3.79337796973563135E+00 -6.12791988559719114E-02
26
Distances in Angstroem, Energy =  2.1746228E-01 eV (PBE)
Au  1.47028154248345899E+00 -9.79565150276169971E-01 -2.28530120006551218E+00
Au -1.77377596874137700E+00 -4.10991425351817075E+00 -4.35174910298080697E-01
Au -2.19227861473195151E+00 -1.74153178367178585E+00  9.80651808608419384E-01
Au -2.85605079848826815E+00  4.07096885051261570E-01 -8.49783093079631957E-01
Au -2.85709599876897091E-01  5.11129204488240596E+00  2.24706167778917387E-01
Au  2.08716784319218140E+00  1.86163987144919685E+00 -2.38671969317928667E+00
Au  1.82085018393644726E+00 -1.63670861057361305E+00  1.79972451846575576E+00
Au -1.93980524497726958E-01 -3.60126541130372813E+00  1.87565217505144233E+00
Au -5.64585213868113556E-01 -1.20068647789864791E+00  3.24528498198477555E+00
Au -2.02234071485006117E+00  8.84644112603621635E-01  1.90435281233314546E+00
Au -5.31614212469872505E-01  3.34564690014392685E+00  2.37226336804413140E+00
Au  1.19140635420836905E+00  2.76407159340715092E+00  1.30030762422628315E-01
Au  2.78110399135061304E+00  4.65716891791576826E-01  1.40685220496392482E-01
Au  1.18592081035704222E+00  1.02495474348090676E+00  2.38732013281066946E+00
Au -7.33003764518936562E-01  1.37160688737416026E+00  4.32651721408974588E+00
Au  8.59479559158993212E-01 -3.38074113795438258E+00 -8.24156363747376997E-01
Au -2.72050013151867009E+00  2.60920664229577515E+00 -2.67221871713625925E+00
Au -6.08688879770629576E-01  8.52615782494299590E-01 -2.73576401024827387E+00
Au -3.75343305224586743E+00 -2.16226762176472942E+00 -1.32780493569801661E+00
Au -1.14071779638112858E+00 -1.80303868344005136E+00 -2.02327640236003337E+00
Au  2.50130258766427271E+00 -4.23629185552000553E+00  1.24512162983901575E+00
Au  4.07236288223132714E+00 -9.36707908119543270E-02 -2.23759667841482868E+00
Au -6.63258237280175827E-02  3.62235503176580709E+00 -2.23506028784861943E+00
Au -1.88241283040144669E+00  2.88138752183738500E+00 -7.50564231850338465E-02
Au -4.33412387084088913E-02 -3.42506836798617883E-04  1.42075409505251094E-02
Au  3.39888341025522012E+00 -2.25621062502400749E+00 -5.58605617609090221E-01
26
Distances in Angstroem, Energy =  2.4209993E-01 eV (PBE)
Au -1.84978698923934148E+00  2.15487611673442681E+00  9.97521815119864597E-01
Au  1.06997398034283542E+00 -3.88564502673302137E+00 -1.20354131098797046E+00
Au -9.43381389582412466E-01  2.53899398609135241E-01 -2.74377550939663539E+00
Au  2.19588557958833119E+00 -2.07854299259963948E+00  6.98119510451084557E-01
Au -5.60733389331829146E-01  2.94678513127318409E+00 -3.55064001073730573E+00
Au  4.93232391092111977E-01 -1.02134602115394713E+00  2.67774213140328454E+00
Au -2.97253965899616190E+00 -4.45554735699113913E-01 -8.91599341478565899E-01
Au  8.83709781642890824E-01  1.88373166482676102E+00  1.68239140542448795E+00
Au -2.57557854306842282E+00  2.40648761144636447E+00 -1.73794480846381494E+00
Au  3.92843473603034310E+00 -8.14157568227600464E-02 -2.53336127910531506E+00
Au -1.22063878895226097E+00 -2.40662028958019869E+00 -1.72431968641975852E+00
```

```
Au  1.49848041388510778E+00  1.11087699685479957E+00  4.26146458493167479E+00
Au  1.95536675719763825E-01  2.50095889701227536E+00 -9.31105238686888437E-01
Au -1.17533802040336033E+00  1.15838540468030193E+00  3.57181514647790221E+00
Au  1.36315905481260957E+00 -1.34041987738492518E+00 -2.45649242140240176E+00
Au  3.09540445533902364E+00  1.15330570219817852E-01  2.22232424395727746E+00
Au  2.61876935267126409E+00  1.16728765197891859E+00 -4.12714620496254625E-01
Au -2.23133992016556304E+00 -1.09769045671009380E+00  1.90674391840882085E+00
Au -5.31479347969354365E-01 -3.01954843430837849E+00  8.93571884973819119E-01
Au -4.46618997611546931E+00  1.76964711648139361E+00  9.62243957219532653E-02
Au  4.62636987193921723E+00 -7.01674752120170941E-01  9.61295263334333017E-02
Au -3.76284811555256793E+00  1.14225175383867383E+00  2.77076558308616727E+00
Au  3.62085362850710846E+00 -2.77563079760638054E+00 -1.61231448531970045E+00
Au -1.16724879210588176E-01 -2.60708053632643821E-01 -3.42589360551767072E-02
Au  1.68280704438761841E+00  1.37457074752348851E+00 -3.12815717901367663E+00
Au -4.86603794737102469E+00 -8.70291867128505259E-01  1.08541068127397455E+00
26
Distances in Angstroem, Energy =  2.5292878E-01 eV (PBE)
Au  6.05691769212369380E+00  3.03946154861430284E-01  2.35541587839740796E+00
Au  1.14547746868672584E-01  3.12554386032443432E+00  5.62346050247536411E+00
Au  1.46102888390888541E+00  4.43753837559010478E+00  3.62978869407203986E+00
Au  4.10998148544132480E+00  9.21814341453299591E+00  4.97905482418781364E+00
Au  4.31940942539037209E+00  8.24623793892214429E+00  2.40654454123494510E+00
Au  5.89341855794961322E+00  7.11979230335358260E+00  4.51248631901197950E+00
Au  7.14561182042624665E+00  4.60935578379289712E+00  4.34775137323982719E+00
Au  5.60810648521449373E+00  3.10698412383861067E+00  6.21599605018254575E+00
Au  3.21551552761373971E+00  2.28914714737318087E-01  2.07921456408138638E+00
Au  1.60112342956128262E+00  5.39267604984952964E+00  6.39219128249286950E+00
Au  3.60638769329285314E+00  7.10406606815595953E+00  6.92027368516767094E+00
Au  6.01044955556358929E+00  8.41363862745401114E+00  7.00461064473164274E+00
Au  6.27536754778870609E+00  5.66043322332164589E+00  6.84861782559748900E+00
Au  2.66522243679423587E+00  6.92721488609729352E+00  4.31030008284588906E+00
Au  7.31704963944227416E+00  1.84551959816090672E+00  4.34431453251031563E+00
Au  4.43915523406048340E+00  1.37459319409737124E+00  4.34363450376833171E+00
Au  2.94843131065976438E+00  5.78822387079856426E+00  1.72736026082037974E+00
Au  2.86177555408209239E+00  2.96980054959276041E+00  1.66281493671498115E+00
Au  1.55302114552300674E+00  1.57906965578757164E+00  3.76595541508200959E+00
Au  4.80671143633445475E+00  1.58435721535147600E+00  1.70354074489156543E-01
Au  6.29142088288993673E+00  3.10854220293164296E+00  2.06417616025974171E+00
Au  5.97452229941879143E+00  5.88023476598287420E+00  2.02975339189133352E+00
Au  4.56813591023428334E+00  7.07390121653304238E+00 -1.02504089788763880E-01
Au  4.30167906066082306E+00  4.45082721354358579E+00  3.97721743201638267E+00
Au  2.88214780699496975E+00  2.92760431588528425E+00  5.99009648047050280E+00
Au  4.74809563176117333E+00  4.31064906650323820E+00 -2.23994359530160676E-02
26
Distances in Angstroem, Energy =  2.5772540E-01 eV (PBE)
Au  6.81079367174662487E+00  7.78108300175294199E-01  6.16987374215972917E+00
Au  8.41305118466434365E+00  1.52389689453556643E+00  8.32687763828192296E+00
```

```
Au  4.14995250514385372E+00  2.30768443935633538E+00  5.04634564923246032E+00
Au  3.37340314123080054E+00  5.32772176614206305E+00  5.33069468831117632E+00
Au  4.33670618512835215E+00  3.23156262551571016E+00  9.60032080242886465E+00
Au  2.80362606178279217E+00 -1.29602299354168005E+00  4.68905063260144317E+00
Au  3.29621653170006468E+00  3.83584572928639433E+00  7.88970013327175757E-01
Au  6.69046377533781911E+00  3.62621350848396418E+00  8.17028384322819434E+00
Au  1.56348200494709211E+00  8.17682564748401619E-01  3.35458278540027077E+00
Au  2.61970018709783226E+00  3.38660264486512474E+00  7.30040422711085846E+00
Au  8.14847027134080304E-01  3.11287386168860936E+00  2.03395950839285211E+00
Au  4.37507307083675379E+00 -3.74462609415336600E-01  6.77321103448120798E+00
Au  1.45894783672128803E+00  3.31209725329031546E+00  4.72067482766622692E+00
Au  6.07904643649841603E+00  4.58818959262143355E+00  5.53765952449445820E+00
Au  2.31941360134946128E+00  1.22897382467733829E+00  6.13776968737546724E-01
Au  6.73170863114976736E+00  2.41465535327527281E+00  3.84181435696426288E+00
Au  2.13678412878650370E+00  5.40467978944584537E+00  2.87028422891586477E+00
Au  4.70862051620857791E+00  1.75199895805387351E+00  2.02017187494371653E+00
Au  5.74696695152574666E+00  9.35789152661834156E-01  8.83021816276704286E+00
Au  3.47058392746989997E+00 -8.33391434640158524E-01  2.01833314343869707E+00
Au  4.80405270185090938E+00  4.38600593739165934E+00  3.04897600678356318E+00
Au  5.29671758575335261E+00 -1.37832001112836611E-01  4.04845360677917565E+00
Au  4.58130780387414038E+00  5.34509035657532738E+00  7.81749503900538478E+00
Au  1.87113058616804429E+00  9.15622090631036056E-01  6.17040929274116579E+00
Au  2.93974224102220649E+00  1.00341895351272647E+00  8.76223637633824382E+00
Au  8.41836872782027612E+00  3.07770605253478857E+00  5.94917072952009907E+00
26
Distances in Angstroem, Energy =  2.5874108E-01 eV (PBE)
Au -1.34319829032882287E+00  2.37107371182792281E+00 -6.65891899564554968E-01
Au  1.39681089473148945E+00 -2.85259146014788678E+00 -9.62633857710792129E-01
Au -1.87439034478928113E+00  1.04627143170902515E+00 -3.15403518362848789E+00
Au  2.44201943776236963E+00 -1.47888049923148368E+00  1.25927099168799961E+00
Au  7.52874811024168977E-01  1.76147841098596869E+00 -2.36855429154695019E+00
Au  7.34519194514315132E-01 -3.84365650658532942E+00  1.52373763134965934E+00
Au -2.67192026739857402E+00 -3.21442815649656155E-01 -6.88021500433718014E-01
Au  1.65016802732789358E+00  8.68513335544934262E-01  2.76803861008554275E+00
Au -3.96262670936536798E+00  1.97392201658170729E+00 -1.61582428051811511E+00
Au  3.63992375234375531E+00  1.05160186240857212E+00  9.00587468149390169E-01
Au -6.59136556425868325E-01 -1.43938776952903313E+00 -2.35045039380315091E+00
Au -5.49316115529617233E-01  6.19526496630017687E-01  4.49081782833988807E+00
Au  3.42079401949249462E+00  2.24422352327645180E+00 -1.58779171445672684E+00
Au -3.04379915920822164E+00  1.56704337963669005E+00  3.78071072046264378E+00
Au  2.33323808937105942E-01  3.96523872443222472E-02 -4.58233032730024714E+00
Au  2.87960026523984836E-01 -1.55503807827716867E+00  3.04266591638923112E+00
Au  2.86928413502940582E+00  7.58651146054259429E-01 -3.84376883038992245E+00
Au -1.84212998302253061E+00 -3.31618297745187141E+00  2.79559416788685189E+00
Au -1.26619670676259166E+00 -2.50266368521280391E+00  1.86278311474467578E-01
Au -8.45724835228390481E-01  1.99263941004678036E+00  2.06191279986164000E+00
Au  1.33398758216531599E+00  2.51467320052645604E+00  3.48377586893082403E-01
```

```
Au -2.33496077795398449E+00 -6.49979217412091970E-01  2.23502518224248670E+00
Au  1.91755909049512074E+00 -1.89677044739116396E+00 -3.47597908178274251E+00
Au  1.62385927357118709E-01 -5.05141611479633726E-02  1.15885371184719185E-01
Au  3.09883833707908796E+00 -6.11373648862069574E-01 -1.31293548816419481E+00
Au -3.54704929877041186E+00  1.70921095442549276E+00  1.09931426329201165E+00
26
Distances in Angstroem, Energy =  2.7306527E-01 eV (PBE)
Au -4.23909683869863108E-01  2.56995149765127096E+00  1.36977413880812637E+00
Au  1.17657529737181865E+00 -3.96482559704932935E+00 -1.09092774669520232E+00
Au -9.87346249211224825E-01  1.05058823687447581E-01 -2.81877517119238119E+00
Au  1.51088875544028789E+00 -2.16636907222653985E+00  9.48254033417379349E-01
Au -6.97645544035581877E-01  2.74827241105334075E+00 -3.51555818870353187E+00
Au -8.42318310074884447E-01 -1.54956888217056665E+00  2.38017655111663240E+00
Au -3.24400396844681183E+00 -2.88082923694458148E-01 -1.19342092131556843E+00
Au  2.42636963492469437E+00  2.61206048929402623E+00  1.18609034042620243E+00
Au -1.89586261458725480E+00  2.21649179794247164E+00 -1.07558653245197577E+00
Au  3.88893269825302079E+00 -1.53753674887071007E-01 -2.55917519909754221E+00
Au -1.15322558925050767E+00 -2.21946733523313577E+00 -1.34499871538518700E+00
Au  1.10771770547911652E+00  2.59548408405595010E+00  3.68557002690492075E+00
Au  8.48680306340188984E-01  2.81008638844228464E+00 -1.13774724593579335E+00
Au -1.15417898288327403E+00  9.35295215875005859E-01  3.67999132541959773E+00
Au  1.33273219838741741E+00 -1.41886898089597602E+00 -2.41801351644493234E+00
Au  1.38270600586731618E+00  1.84654969753420578E-01  2.42906383335187970E+00
Au  3.58723645239797140E+00  2.35955564616685853E+00 -1.35232099795582794E+00
Au -3.09218440017559937E+00 -2.37139641577364646E+00  9.01116606376502816E-01
Au -8.17478333411778446E-01 -3.88606556770573164E+00  8.27667835849541178E-01
Au -2.78222660333356275E+00  1.10588304586119568E+00  1.26690216675483813E+00
Au  3.01343226472256731E+00  5.41170229720098622E-02  1.04126256985726043E-01
Au -3.33554771736961886E+00 -7.31059955670741446E-01  3.27337185431070665E+00
Au  3.56718821914267892E+00 -2.45506664518832318E+00 -1.06462830659838681E+00
Au -2.86244686019452160E-02  1.14101997873262265E-01 -7.69457129261068434E-02
Au  1.65986098431215878E+00  1.22888465033994954E+00 -3.34638683459150510E+00
Au -5.04776805738748369E+00 -4.35372990473152144E-01  9.42380119572222119E-01
26
Distances in Angstroem, Energy =  2.7441049E-01 eV (PBE)
Au -1.86506931574139068E+00  1.94510010809694611E+00 -1.33087227929535246E+00
Au  2.33461497775765581E-01  7.43133298708543388E-01 -2.74808526965412891E+00
Au  4.15111491849256264E+00 -1.26295659725255383E+00 -2.10124348287445928E+00
Au  3.25799166441670707E+00 -2.66725650978861184E+00  1.20791226312274610E-01
Au -2.38813667513574829E-01 -1.79385865925885524E+00 -3.90672244334981666E+00
Au -1.96195785811220613E+00 -7.77702506654165182E-01 -1.85693709562980724E+00
Au -5.59557082331735067E-01 -1.68389851462181883E+00  2.76202133047057297E+00
Au  2.02725000775003394E+00 -9.55605360977369589E-01  1.95508326131861021E+00
Au  1.97558650523506341E+00  1.69316671942949593E+00  8.78445287089787863E-01
Au  3.02083627089387519E-01  1.02179489938413637E+00  3.07954756185730316E+00
Au -2.56902361828637327E+00  1.61325768952211934E+00  2.95221560151030893E+00
Au -1.53653194350035771E+00 -6.49705161477441978E-02  4.84071638849408181E+00
```

```
Au -5.3431221821276075E-01  2.8730649724723256E+00  1.1998058716590662E+00
Au  4.1182543150029991E+00  4.9247568405493401E-02  3.7701988689444227E-01
Au -1.0574057400584300E+00 -3.3854831856478599E+00 -1.7422368946228661E+00
Au -2.2792905870047697E+00 -2.3418784009341311E+00  5.3787536006723336E-01
Au -3.2223550817559241E+00  3.2351846530066453E-01  5.3248102185774282E-01
Au -3.3149223176907796E+00  3.1775862876929355E+00  8.5230048126670044E-01
Au -1.4736665443189245E+00  4.6135442352216529E+00 -7.6402299852579148E-01
Au  8.2482454253226123E-01  3.0734589651699098E+00 -1.3150581477680605E+00
Au  2.8380203209544752E+00  1.2285087331435218E+00 -1.8478982846698715E+00
Au  5.2903389770945597E-01 -2.8618658207576546E+00  4.6770213236732428E-01
Au  1.6960047738768900E+00 -2.7539354128882669E+00 -2.1302260407866704E+00
Au -3.3387421676203868E+00 -1.1837769449191993E+00  2.8488403557254984E+00
Au -2.8647634136224504E-01  3.0161331908914714E-02  3.1363283745444964E-01
Au  2.2844984126737273E+00 -6.5235484460865179E-01 -3.9751756671692196E+00
26
Distances in Angstroem, Energy =  2.7688857E-01 eV (PBE)
Au -3.3527856346392637E-01  2.5959512310388431E+00  7.9798047503213787E-01
Au  1.2940192101643739E+00 -3.9270978329980077E+00 -8.5646356275665458E-01
Au -1.0785199594180273E+00 -1.6989040511699188E-01 -2.9837744982666545E+00
Au  2.6154322936187144E+00 -1.6654645484685335E+00  9.0987810870621369E-02
Au  1.8108048450750328E-01  2.2213466940309892E+00 -2.0025235882443391E+00
Au -1.1133520350753163E+00 -1.0060279438436823E+00  3.1494938658921146E+00
Au -3.4285363825885580E+00 -9.6440422759589084E-01 -1.7571157345012758E+00
Au  2.0751603485620946E+00  2.5850674979492543E+00  2.2175994729016701E+00
Au -2.2517662639058726E+00  1.3494815286208857E+00 -7.7822113443836860E-01
Au  3.5914050973182836E+00 -3.2291542603429241E-01 -2.3704401093462559E+00
Au -1.1371763191862701E+00 -2.5736546026609601E+00 -1.3883482925761082E+00
Au -1.5341668391868898E-01  1.6961205557278138E+00  3.5358349597646837E+00
Au  1.9883882700594764E+00  3.6669353901636050E+00 -3.7280177381137047E-01
Au -2.4297832345478461E+00  8.2412493369812245E-01  4.7752533009698563E+00
Au  1.2200353639264079E+00 -1.8535941504729288E+00 -2.7844834793642437E+00
Au  1.4609748634753010E+00 -1.8024485903746504E-01  2.1694336369075864E+00
Au  2.9774742999999919E+00  2.3549418759621839E+00 -2.6001833744629010E+00
Au -2.5867365114456282E+00 -1.7465421746385423E+00  8.6959061063164633E-01
Au  7.2210142468588867E-02 -2.5571883291168565E+00  1.2301859960544876E+00
Au -2.5788063404934425E+00  1.5622920284003403E+00  2.0252363600378915E+00
Au  2.9409279477642509E+00  1.0617579071948804E+00 -5.4264267653319790E-03
Au -3.9856958749565834E+00 -6.6212274246459824E-01  3.0204776727559124E+00
Au  3.6962907318427658E+00 -3.0519891686280967E+00 -2.0636841120227891E+00
Au  1.7275634591750555E-01 -1.4231316395608029E-01 -3.6176450404874516E-01
Au  1.3584511012035536E+00  6.2463513597800180E-01 -4.0106441126557684E+00
Au -4.5655383318286500E+00  2.8079479626792958E-01  4.5380054144239723E-01
26
Distances in Angstroem, Energy =  2.7763545E-01 eV (PBE)
Au  8.1122143198785182E+00  3.4359411490572187E+00  6.3191649937640042E+00
Au  1.2801393956145692E+00  1.9388153221279265E+00  4.1572434311761750E+00
Au  1.3642336158951676E+00  4.7196228332593231E+00  3.3827972564682325E+00
```

```
Au  4.24882971757014172E+00  4.54469996352215233E+00  3.74670141066344486E+00
Au  4.58635903837970993E+00  9.38719790403565568E+00  3.69817685510103100E+00
Au  6.03573200271250077E+00  6.93929676804379358E+00  3.81564580427173805E+00
Au  7.15044250534037928E+00  4.32381086545637139E+00  3.87220999366284957E+00
Au  5.33777346594667446E+00  3.51661832753245118E+00  6.24028394682458298E+00
Au -1.11173011235809777E-01  4.00544665358057905E+00  5.62096262870973185E+00
Au  1.33401636826343961E+00  6.35552925422612525E+00  5.83884266990358736E+00
Au  4.13213186308035763E+00  6.09284794859735346E+00  6.30362588642767996E+00
Au  5.80701700984426594E+00  8.39445861551361183E+00  6.25743279176438083E+00
Au  6.97886853648589334E+00  5.96111782905778131E+00  6.31889569445707888E+00
Au  2.82187566578165949E+00  8.60487046635887864E+00  5.99576478234949128E+00
Au  6.72039373905088677E+00  1.56713637684439289E+00  4.61860203875806530E+00
Au  3.98658280180961100E+00  1.56038168300738334E+00  4.40123357358694278E+00
Au  2.96801263355964728E+00  5.43136111447979708E+00  1.28893202347388613E+00
Au  2.81398666342903203E+00  2.51162560189448003E+00  1.95371798116229622E+00
Au  4.48166338426438404E+00  1.07755141011084348E+01  6.05151783005048394E+00
Au  4.73935829514952722E+00  6.16644440798249160E+00 -7.16197513797235175E-01
Au  5.67632615892508419E+00  2.31468472147018112E+00  2.19497084130928499E+00
Au  6.03125150549285927E+00  5.22114073458663608E+00  1.54743586530440380E+00
Au  4.66457763586735652E+00  7.83745270506058400E+00  1.48644443959603478E+00
Au  2.81483924585764367E+00  7.16020015143819055E+00  3.54205743220528291E+00
Au  2.62110657581578810E+00  3.70454451365248660E+00  6.00479561130738571E+00
Au  4.49029496722075994E+00  3.36778228810571756E+00 -4.35109185008728250E-02
26
Distances in Angstroem, Energy =  2.7799505E-01 eV (PBE)
Au  9.44337405568231425E+00  1.20205049581101275E+00  5.17594983762346761E+00
Au  7.65028206771318331E+00  3.22324034426945838E+00  5.70142705329320698E+00
Au  3.53827377848394686E+00  6.86817362077664395E+00  6.48695921313967006E+00
Au  2.99567974353176192E+00  2.36184535277734442E+00  2.23278812780167568E+00
Au  5.76002935126751314E+00  2.85155634940508840E+00  7.81919206543418355E+00
Au  4.16917366468293871E+00  3.30533014165587780E-01  3.79050687289440136E+00
Au  4.10364021417040625E+00  5.84131864031264314E+00  3.94401735156327993E+00
Au  5.63055937063349532E+00  4.43714926802048770E-01  9.19502170465518276E+00
Au  1.69921556394546025E+00  5.11437892466864330E+00  5.37108521615060219E+00
Au  3.19831426372304861E+00  1.60763849140684933E+00  8.39041491956239760E+00
Au  9.68151403459989734E-01  2.79976874949269661E+00  4.23462264614741191E+00
Au  4.92272318060681702E+00  2.04256845674298804E-01  6.52119176372248965E+00
Au  1.89833803372754550E+00  4.89197880351019432E+00  2.55031610771828010E+00
Au  3.40310126644662292E+00  4.28568606870352031E+00  7.55226858544771407E+00
Au  1.44457136152663179E+00  2.55924118055064487E-01  3.23755358712160835E+00
Au  7.22732240448061525E+00  4.01362599311838952E+00  8.84466907991326656E-01
Au  4.53486110860756142E+00  4.58926499611543548E+00  1.45617097496443026E+00
Au  8.46779831402846916E+00  2.59653964262985992E+00  2.95634654653269324E+00
Au  7.64017184227436275E+00  7.53450361869736795E-01  7.26864397127939998E+00
Au  6.84054772852756265E+00  6.74082651161739155E-01  4.44973122764125062E+00
Au  5.69759797166480908E+00  5.14279781941715619E+00  6.15650011591202340E+00
Au  5.75083697354464274E+00  2.05914428309277264E+00  2.19299918945080208E+00
```

```
Au  4.85051075516881980E+00  2.76805674959330394E+00  4.96386550369059343E+00
Au  2.30341797516695701E+00  8.01949073393788514E-01  5.84487890519119890E+00
Au  1.02353340871576060E+00  2.94045261648783862E+00  7.06762382889686247E+00
Au  6.59360970477677633E+00  4.70756793369201265E+00  3.49794696641898062E+00
26
Distances in Angstroem, Energy =  2.8256243E-01 eV (PBE)
Au  1.01408616797445692E+01  1.86003937816816123E+00  5.38133215276272470E+00
Au  8.18052916228369575E+00  3.58850354095763002E+00  6.31575002616668968E+00
Au  3.34873774374319222E+00  6.22186219953392516E+00  6.27485235987649936E+00
Au  2.68637835605974784E+00  1.11419255136802020E+00  1.74578672553180581E+00
Au  5.95474697003408071E+00  2.41221775245105219E+00  7.64270355120371825E+00
Au  3.07234789460425306E+00  4.00462283709397737E-01  4.44180780989218604E+00
Au  4.85071674934952934E+00  5.60944892189743083E+00  3.97033392301849419E+00
Au  5.99399574102204991E+00 -3.62339646567508611E-01  8.11936109673783157E+00
Au  2.09060690240300229E+00  5.16505920145668451E+00  3.97975930906910547E+00
Au  3.66314219243709660E+00  7.94375394041056393E-01  7.16169677040317332E+00
Au -2.17806500655635282E-01  3.61538833560173334E+00  4.49177819772000753E+00
Au  5.73384298501224521E+00  8.62916647654226460E-02  5.32435209286109945E+00
Au  1.08043271959061848E+00  3.41473752311652667E+00  2.06766941892038947E+00
Au  3.38949193720133746E+00  3.63509117649136781E+00  7.39352402387119056E+00
Au  5.02860689292863539E-01  1.08361266322039906E+00  3.58430267556803717E+00
Au  6.56383186826248810E+00  3.67307994539357585E+00  2.79392234349869018E+00
Au  3.83278850665366377E+00  3.64044643496994036E+00  2.29080707799180239E+00
Au  9.25026452271742272E+00  3.75637419316647581E+00  3.56981456738433200E+00
Au  8.11017168427417445E+00  7.81792940596563346E-01  6.82125645751193765E+00
Au  7.78083656901508558E+00  1.36786103594384589E+00  3.90325203450189662E+00
Au  5.79860135543078759E+00  4.99268725623267073E+00  6.63600187116419793E+00
Au  5.25955404773166890E+00  1.13105365955294968E+00  2.72088529822816483E+00
Au  4.71107547859636266E+00  2.80711129153992234E+00  4.98795377658234518E+00
Au  1.43178469650394891E+00  2.02719232727452869E+00  6.09410868732908906E+00
Au  9.63578036467484300E-01  4.81118703186381946E+00  6.63509967959861058E+00
Au  7.58226351878210636E+00  5.70126780965731417E+00  4.59437726285130932E+00
26
Distances in Angstroem, Energy =  2.8751312E-01 eV (PBE)
Au -1.38177376561936938E+00  2.38665805331166103E+00 -5.43977593789062408E-01
Au  1.77743629928350488E+00 -3.42674696350476671E+00 -2.55097377566838701E-01
Au -2.00864042617693972E+00  1.64311113432672928E+00 -3.25736904284103490E+00
Au  2.19994579744999452E+00 -1.30653381782002809E+00  1.57569521227459308E+00
Au  6.92910955700985265E-01  1.97768300771601213E+00 -2.38295058445184527E+00
Au  4.18971934486553777E-03 -3.19946387360711215E+00  1.79446617907650641E+00
Au -2.51016712547380072E+00 -3.28243564781490860E-01 -1.21785702944383711E+00
Au  1.85928920588133240E+00  1.25692530419713866E+00  2.89724426740603969E+00
Au -4.01818107626367382E+00  2.01358575232833736E+00 -1.37390660770045270E+00
Au  3.62557153067842730E+00  1.03133560738628383E+00  8.42517417683357195E-01
Au -4.39822914848483293E-01 -7.38908717083039623E-01 -3.06339537301182263E+00
Au -2.44319641223297773E-01  1.45193687575839170E+00  4.77534815549740621E+00
Au  3.38889706947827340E+00  1.75322563810550758E+00 -1.82041729009222553E+00
```

```
Au -2.4528128991962936E+00  1.3504064054670014E-01  3.5844320522291744E+00
Au  1.1007985745987185E-01  1.1990767495248398E+00 -5.0005956446688851E+00
Au  2.2756461554503127E-01 -9.6452093150126073E-01  3.4982697284267257E+00
Au  2.2718802127606996E+00 -3.7669583041762968E-02 -3.6844069024199166E+00
Au -1.9597515690414660E+00 -2.6513025378850297E+00  3.7561560982082769E+00
Au -7.4088634470351689E-01 -2.5709109831291985E+00 -9.5209328022159001E-01
Au -7.8199123429160566E-01  1.8806794380073160E+00  2.1054874699070130E+00
Au  1.3571932386787935E+00  2.4842742564071285E+00  3.0065539369786543E-01
Au -2.1809854564294771E+00 -1.4814916969073624E+00  1.2703331217128895E+00
Au  1.5561695331562107E+00 -2.6445132178335871E+00 -2.9027791344874010E+00
Au  1.3546017954426967E-01 -4.8258743908043932E-02 -1.8790617345900880E-02
Au  2.8998941119850414E+00 -9.3655434726874720E-01 -1.0921850190073150E+00
Au -3.3871498736793191E+00  1.2158652065544651E+00  1.1652164009282933E+00
26
Distances in Angstroem, Energy =  1.9958403E-01 eV (LDA)
Au  4.8934737145986240E+00  2.0202576696148158E-01  3.3468463439195326E+00
Au  5.4734659805302288E-01  1.9504214703909095E+00  3.2831650133748851E+00
Au  1.3894438902061421E+00  4.5987435064625020E+00  3.2200493243086177E+00
Au  4.1016418309755505E+00  4.4791531978771841E+00  3.8057734527988920E+00
Au  4.3910450546237536E+00  9.1870383917132816E+00  3.6539208375050936E+00
Au  6.0026890059059133E+00  6.9817006049391308E+00  3.7772904843401514E+00
Au  6.8482080688292397E+00  4.4254064442118297E+00  4.2228867737290337E+00
Au  6.2996729694650152E+00  3.9751114287829630E+00  6.8723649778504221E+00
Au  9.6330966127858775E-01  3.2711551373330621E+00  5.5817949167162597E+00
Au  1.9972110324297496E+00  5.7251301761003113E+00  5.7164424055446803E+00
Au  4.5832159426685069E+00  6.1019795384592319E+00  6.5410922141777066E+00
Au  5.7264408646131422E+00  8.5460260583685059E+00  5.9704773758792245E+00
Au  7.2956890051598142E+00  6.3638933221991136E+00  6.1658744903292915E+00
Au  3.0080835049349366E+00  8.2081687979926361E+00  5.8242729764064505E+00
Au  5.5830633386188202E+00  2.0921482456380054E+00  5.1248442571794044E+00
Au  2.9351621362401823E+00  1.6179939504156915E+00  4.6084331460136268E+00
Au  3.1862724761185217E+00  5.3806074328758839E+00  1.1945104367338904E+00
Au  2.6169890407978498E+00  2.7687253185706031E+00  1.6071860475590569E+00
Au  2.3409741078095441E+00  1.7557618880341297E-01  2.3363226930094170E+00
Au  4.4390190676639811E+00  9.2636062992007573E-01  7.2962709343004317E-01
Au  5.8272181047752394E+00  2.6087807167250499E+00  2.4210285339061612E+00
Au  5.9370227252477097E+00  5.2619011583255437E+00  1.6896308698927041E+00
Au  4.6141989357301432E+00  7.6420105176945299E+00  1.4528846810928320E+00
Au  2.7316068549178763E+00  6.9987468135747646E+00  3.3770690019941299E+00
Au  3.5995452749088273E+00  3.5633454637309616E+00  6.3981351620621885E+00
Au  4.7852007366851055E+00  3.5029615473994356E+00  1.6992577397170573E-02
```

E.4 Tubular Motif

```
26
Distances in Angstroem, Energy =  3.7159014E-01 eV (PBE)
Au  2.5197210529647354E+00  2.9168634638069007E+00  8.1593311215020474E+00
```

```
Au  2.6382988304897894E+00  3.0796884822342467E+00  3.47660351179517046E-01
Au  4.9651256292647964E+00  1.5626954085070348E+00  7.8283232082775005E+00
Au  2.5674601160112402 5E+00  1.3249508809533266 9E-01  7.7594071785606031 4E+00
Au  1.2877231947307768 4E-01  1.4926361829762382 3E+00  7.7501732997524417 4E+00
Au  8.6037185909217717 1E-02  4.2840117531009500 6E+00  7.8070335227447058 6E+00
Au  2.4824923263365712 4E+00  5.7171339717121751 9E+00  7.8715059025141354 0E+00
Au  4.9211605333211654 3E+00  4.3547338972932276 8E+00  7.8808388701043972 1E+00
Au  4.0626285776601669 4E+00  6.7339697447597257 6E-01  6.6154773212940454 2E-01
Au  3.9959769559853941 5E+00  5.5083953056735230 9E+00  7.6278024552037926 3E-01
Au  1.2023156967599648 6E+00  5.4665987577250936 3E+00  7.1920949489291963 2E-01
Au -1.6242266916960646 0E-01  3.0291429897388337 8E+00  6.4922173284427398 0E-01
Au  1.2724271135307343 4E+00  6.3536206205213252 6E-01  6.2151965115469121 1E-01
Au  5.4262805883206466 4E+00  3.1084365162885121 4E+00  7.3584895439328768 0E-01
Au  3.9445252913427042 4E+00  5.4460357506699725 9E+00  5.5000518858027236 4E+00
Au  5.4110420608142062 4E+00  3.0102355311230808 4E+00  5.4734836322756432 7E+00
Au  4.0180550246839121 8E+00  5.3608383859315778 0E-01  5.4010827302422050 9E+00
Au  1.1769419677446584 2E+00  4.8858660445554297 2E-01  5.3555568533295545 6E+00
Au -2.8085972222055816 1E-01  2.9294079830113077 8E+00  5.3852002472925137 9E+00
Au  1.0119896088537457 5E+00  5.4138697497492316 3E+00  5.4549754491271151 8E+00
Au  2.6392082011372313 1E+00  1.8076275879767220 2E-01  3.0208080781398489 7E+00
Au  5.0831051913548961 4E+00  1.6305852784806633 7E+00  3.0886101993654788 7E+00
Au  5.0372266330144199 7E+00  4.4720330350430117 5E+00  3.1470607227005169 5E+00
Au  2.5571704786025555 3E+00  5.8585740939735808 7E+00  3.1369799653715242 4E+00
Au  1.1910094686118925 8E-01  4.4000668172659516 1E+00  3.0706545527539557 1E+00
Au  1.5114905839394718 3E-01  1.5592797051566809 1E+00  3.0128143942184308 9E+00
```

F Coordinates of Au_{26}^-

In this section the coordinates of the isomers for which the computed photoeletron spectra yield a good match with the experimental photoelectron spectra are given in the xyz file format.

F.1 Isomer 3

```
26
isomer 3, distances in Angstroem (LDA-Geometry)
Au  8.56390007473861914E+00  1.62813118597455753E+00  3.84506264722230595E+00
Au  7.52979683230909291E+00  3.61538281507366221E+00  5.51783586113459901E+00
Au  2.85997906184454775E+00  6.17406817842427014E+00  6.68529792208702922E+00
Au  3.93558816482195928E+00  1.98320570700093279E+00  2.43570266613816999E+00
Au  5.95362997777755609E+00  2.72728732899271575E+00  7.54977648580234639E+00
Au  3.46719761444288643E+00  1.16619807625252384E-01  4.56858100675232937E+00
Au  4.28613558147161822E+00  6.87726001742204396E+00  4.51575465339584792E+00
Au  6.23438324546289557E+00  1.83983467930523392E-01  8.43265257065652918E+00
Au  1.90360882346215132E+00  5.54272118205560727E+00  4.17360250679909939E+00
Au  3.89103844881442473E+00  8.78985411941761718E-01  7.21049032959466452E+00
Au  1.56205299977200313E-01  3.41430447715891727E+00  4.21877053316152040E+00
Au  5.68008003334728606E+00 -7.53842709450097126E-01  5.90677845311759953E+00
Au  1.85116786515310539E+00  3.72499073955051285E+00  2.13876431500338970E+00
Au  3.31661237365390926E+00  3.57269994756447007E+00  7.42763163436437424E+00
Au  1.43657372828245533E+00  1.24087234757609455E+00  3.21620736946902941E+00
Au  6.43236137740602487E+00  3.08608574091134313E+00  2.75838552597992637E+00
Au  4.22198370976999016E+00  4.70577676278272072E+00  2.90996353646191253E+00
Au  8.83379997882074619E+00  4.24557988608681569E+00  3.22816634050113516E+00
Au  7.85018331346632792E+00  1.06820247054917616E+00  6.44834014190638971E+00
Au  8.04412258312049566E+00 -9.10652067820296285E-01  4.58905686093019938E+00
Au  5.34506046260942558E+00  5.09675812216348056E+00  6.31289258162645073E+00
Au  5.98947645023099895E+00  5.02952070417775343E-01  3.50865712877924674E+00
Au  4.71046439372265713E+00  2.65508207862393464E+00  5.02980509494596006E+00
Au  1.76718106634718386E+00  1.88759898781572910E+00  5.86751162536211446E+00
```

```
Au  8.74571922037037930E-01  4.39649028434815659E+00  6.56096303463436836E+00
Au  6.62053312346721423E+00  5.66845262568273256E+00  3.88583836441861052E+00
```

F.2 Isomer 4

```
26
isomer 4, distances in Angstroem (LDA-Geometry)
Au  5.38462062383746165E+00  6.87071251529801086E-02  3.08488386989383434E+00
Au  1.23924742408907695E+00  2.03813094396562056E+00  4.23489726203083894E+00
Au  1.44020641051322928E+00  4.71984179543080717E+00  3.44100790040505222E+00
Au  4.23653948604446473E+00  4.49516471653930516E+00  3.74533957785136185E+00
Au  4.59065079903130258E+00  9.11103041552364878E+00  3.75879907237170796E+00
Au  6.28634818765099901E+00  6.93659620232828278E+00  3.70436381540503401E+00
Au  7.08301348821806975E+00  4.31727992871775790E+00  3.94251918234840559E+00
Au  5.42043094000065562E+00  3.56917250238682504E+00  6.15537773382419662E+00
Au -4.18522025302283177E-02  3.98508543664259740E+00  5.60031242118994310E+00
Au  1.40280962362958883E+00  6.25969759858646846E+00  5.78483071847732511E+00
Au  4.13671837102026529E+00  5.99830183267298533E+00  6.34780145255364392E+00
Au  5.70176559362258306E+00  8.21711537178995499E+00  6.07617318655689242E+00
Au  6.88768609790032205E+00  5.81435552055627358E+00  6.19366228278825481E+00
Au  2.95503053321028242E+00  8.42939235829805966E+00  5.87749577453457128E+00
Au  6.72931308034893405E+00  1.73789749474130728E+00  4.69696196041840075E+00
Au  3.96898937946083574E+00  1.68166578380588949E+00  4.77022534179680680E+00
Au  3.10064969217627473E+00  5.43531181488001280E+00  1.31772376235887467E+00
Au  2.58998123572877281E+00  2.79061429315253351E+00  1.89528667013679897E+00
Au  2.63666133593633578E+00  2.23921165131357763E-01  2.86515329882660374E+00
Au  4.27307170339855968E+00  9.62538660179732330E-01  7.67572810132032934E-01
Au  6.11465072707510693E+00  2.53445491238267140E+00  2.08851167472807919E+00
Au  5.94867713080938199E+00  5.22971904289193468E+00  1.57705589925054745E+00
Au  4.65442954466400405E+00  7.61090129948944405E+00  1.53642201828652336E+00
Au  2.76312273565879440E+00  7.10330085529342270E+00  3.48327484032691670E+00
Au  2.64504789756137004E+00  3.75068807436583995E+00  5.95322968675589514E+00
Au  4.49592853094340938E+00  3.53422585509427600E+00  4.00337867514279980E-02
```

F.3 Isomer 18

```
26
isomer 18, distances in Angstroem (LDA-Geometry)
Au  5.24024699637489988E-01  1.78022707359568577E+00 -3.38028719003049893E+00
Au  1.37705324340677282E+00  3.66060789941122477E+00 -1.64885228994592348E+00
Au  2.16808203069707561E+00  1.06385430165422523E+00 -1.29308608924425528E+00
Au  1.16902995289582590E+00 -3.45556296108731242E+00 -2.40848854379831190E+00
Au -1.15087190118835814E+00 -2.25542394157155712E+00 -1.57624381170290251E+00
Au -1.53502746620485131E+00  3.58908899800555570E-01 -2.33366274053112210E+00
Au  9.13949953606714854E-02 -1.75878096140044404E-01  3.90543141715241904E-02
Au  8.91833967336320366E-01 -4.05569499199379191E+00  2.11022447411338598E+00
```

```
Au  2.56569532334106443E+00  1.11700826017554128E-01  1.32949018401557528E+00
Au  8.78303571059449895E-01  1.20417762690235497E+00  3.16123953394850332E+00
Au -1.08401660095107766E+00  2.70554296144324979E+00  2.06914463188249487E+00
Au -1.44098858973818422E+00 -1.73626406320385313E-02  2.39927113882867138E+00
Au -2.93395661428852872E-01  4.57991232352976674E+00  2.14449299022703987E-01
Au  2.51822094696784182E+00 -1.59741594349177096E+00 -8.21584223913149336E-01
Au  3.11890450850427170E+00 -2.47279002022756567E+00  1.78735585699738464E+00
Au -2.71507173197165352E+00 -6.08231852836758624E-01 -1.49384695606041423E-02
Au -3.50894031271358964E+00  1.50203735854508258E+00  1.52171858098811952E+00
Au -2.83284839608949568E+00  3.81498953566238264E+00  2.73042641699039190E-01
Au -3.56452300617550577E+00  1.68989663215910313E+00 -1.19545625578935866E+00
Au -1.34271219381925166E+00  3.10817871364175780E+00 -1.92739018384635785E+00
Au  1.40262078528042977E+00  2.56696150998901551E+00  8.62383641780024801E-01
Au  2.44796584676877282E+00 -4.20806931936364048E+00 -1.49443683727887272E-01
Au -2.79134320383089107E-01 -4.50218122392749898E+00 -3.11816071111768067E-01
Au -1.13130924464269778E+00 -2.46300285124460627E+00  1.25219494236252138E+00
Au  8.64621462243715611E-01 -1.48171911041906967E+00  2.96534494370778079E+00
Au  8.61087465912502070E-01 -8.53669668404558046E-01 -2.92365781752549880E+00
```

Bibliography

[1] D. J. Wales, M. A. Miller, and T. R. Walsh, "Archetypal Energy Landscapes", Nature **394**, 758 (1998).

[2] J. P. K. Doye, M. A. Miller, and D. J. Wales, "Evolution of the Potential Energy Surface with Size for Lennard-Jones Clusters", The Journal of Chemical Physics **111**, 8417 (1999).

[3] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, Cambridge, 2003).

[4] W. J. Wales, "Decoding the Energy Landscape: Extracting Structure, Dynamics and Thermodynamics", Philosophical Transactions of the Royal Society of London A **370**, 2877 (2012).

[5] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (MIT Press, Cambridge, Massachusetts, 1992).

[6] S. M. Woodley, P. D. Battle, J. D. Gale, and C. Richard A. Catlow, "The Prediction of Inorganic Crystal Structures Using a Genetic Algorithm and Energy Minimisation", Physical Chemistry Chemical Physics **1**, 2535 (1999).

[7] V. E. Bazterra, M. B. Ferraro, and J. C. Facelli, "Modified Genetic Algorithm to Model Crystal Structures. I. Benzene, Naphthalene and Anthracene", The Journal of Chemical Physics **116**, 5984 (2002).

[8] A. R. Oganov and C. W. Glass, "Crystal Structure Prediction Using Ab Initio Evolutionary Techniques: Principles and Applications", The Journal of Chemical Physics **124**, 244704 (2006).

[9] S. E. Schönborn, S. Goedecker, S. Roy, and A. R. Oganov, "The Performance of Minima Hopping and Evolutionary Algorithms for Cluster Structure Prediction", The Journal of Chemical Physics **130**, 144108 (2009).

[10] Z. Li and H. A. Scheraga, "Monte Carlo-Minimization Approach to the Multiple-Minima Problem in Protein Folding", Proceedings of the National Academy of Sciences **84**, 6611 (1987).

[11] C. Baysal and H. Meirovitch, "New Conformational Search Method Based on Local Torsional Deformations for Cyclic Molecules, Loops in Proteins, and Dense Polymer Systems", The Journal of Chemical Physics **105**, 7868 (1996).

[12] D. J. Wales and J. P. K. Doye, "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms", Journal of Physical Chemistry A **101**, 5111 (1997).

[13] G. T. Barkema and N. Mousseau, "Event-Based Relaxation of Continuous Disordered Systems", Physical Review Letters **77**, 4358 (1996).

[14] N. Mousseau and G. T. Barkema, "Traveling Through Potential Energy Landscapes of Disordered Materials: The Activation-Relaxation Technique", Physical Review E **57**, 2419 (1998).

[15] R. Malek and N. Mousseau, "Dynamics of Lennard-Jones Clusters: A Characterization of The Activation-Relaxation Technique", Physical Review E **62**, 7723 (2000).

[16] E. Machado-Charry, L. K. Béland, D. Caliste, L. Genovese, T. Deutsch, N. Mousseau, and P. Pochet, "Optimized Energy Landscape Exploration Using the Ab Initio Based Activation-Relaxation Technique", The Journal of Chemical Physics **135**, 34102 (2011).

[17] N. Mousseau, L. K. Béland, P. Brommer, J.-F. Joly, F. El-Mellouhi, E. Machado-Charry, M.-C. Marinica, and P. Pochet, "The Activation-Relaxation Technique: ART Nouveau and Kinetic ART", Journal of Atomic and Molecular Physics **2012**, 925278 (2012).

[18] S. Goedecker, "Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems", The Journal of Chemical Physics **120**, 9911 (2004).

[19] S. Roy, S. Goedecker, and V. Hellmann, "Bell-Evans-Polanyi Principle for Molecular Dynamics Trajectories and its Implications for Global Optimization", Physical Review E **77**, 056707 (2008).

[20] M. Amsler and S. Goedecker, "Crystal Structure Prediction Using the Minima Hopping Method", The Journal of Chemical Physics **133**, 224104 (2010).

[21] A. Groß, "Reactions at Surfaces Studied by Ab Initio Dynamics Calculations", Surface Science Reports **32**, 291 (1998).

[22] T. C. Berkelbach, H. Lee, and M. E. Tuckerman, "Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study", Physical Review Letters **103**, 238302 (2009).

[23] A. Groß, "Ab Initio Molecular Dynamics Simulations of the Adsorption of H2 on Palladium Surfaces", ChemPhysChem **11**, 1374 (2010).

[24] J. K. Clark II and S. J. Paddison, "Ab Initio Molecular Dynamics Simulations of Water and an Excess Proton in Water Confined in Carbon Nanotubes", Physical Chemistry Chemical Physics **16**, 17756 (2014).

[25] A. Groß, "Ab Initio Molecular Dynamics Simulations of the O_2/Pt(111) Interaction", Catalysis Today, (2015).

[26] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, "Transition Path Sampling and the Calculation of Rate Constants", The Journal of Chemical Physics **108**, 1964 (1998).

[27] C. Dellago, P. G. Bolhuis, and D. Chandler, "Efficient Transition Path Sampling: Application to Lennard-Jones Cluster Rearrangements", The Journal of Chemical Physics **108**, 9236 (1998).

[28] M. R. Sørensen and A. F. Voter, "Temperature-Accelerated Dynamics for Simulation of Infrequent Events", The Journal of Chemical Physics **112**, 9599 (2000).

[29] A. Laio and M. Parrinello, "Escaping Free-Energy Minima", Proceedings of the National Academy of Sciences **99**, 12562 (2002).

[30] T. F. Miller and C. Predescu, "Sampling Diffusive Transition Paths", The Journal of Chemical Physics **126**, 144102 (2007).

[31] C. Dellago and P. Bolhuis, "Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events", in *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Vol. 221, edited by C. Holm and K. Kremer, Advances in Polymer Science (Springer Berlin Heidelberg, 2009), pp. 167–233.

[32] D. J. Wales, "Perspective: Insight Into Reaction Coordinates and Dynamics from the Potential Energy Landscape", The Journal of Chemical Physics **142**, 130901 (2015).

[33] J. P. K. Doye and D. J. Wales, "On Potential Energy Surfaces and Relaxation to the Global Minimum", The Journal of Chemical Physics **105**, 8428 (1996).

[34] K. D. Ball, R. S. Berry, R. E. Kunz, F.-Y. Li, A. Proykova, and D. J. Wales, "From Topographies to Dynamics on Multidimensional Potential Energy Surfaces of Atomic Clusters", Science **271**, 963 (1996).

[35] J. P. K. Doye, M. A. Miller, and D. J. Wales, "The Double-Funnel Energy Landscape of the 38-Atom Lennard-Jones Cluster", The Journal of Chemical Physics **110**, 6896 (1999).

[36] T. F. Middleton and D. J. Wales, "Energy Landscapes of Some Model Glass Formers", Physical Review B **64**, 024205 (2001).

[37] P. N. Mortenson, D. A. Evans, and D. J. Wales, "Energy Landscapes of Model Polyalanines", The Journal of Chemical Physics **117**, 1363 (2002).

[38] D. J. Wales, "Discrete Path Sampling", Molecular Physics **100**, 3285 (2002).

[39] D. J. Wales, "Some Further Applications of Discrete Path Sampling to Cluster Isomerization", Molecular Physics **102**, 891 (2004).

[40] D. J. Wales, "Energy Landscapes: Calculating Pathways and Rates", International Reviews in Physical Chemistry **25**, 237 (2006).

[41] J. Doye and D. Wales, "Surveying a Potential Energy Surface by Eigenvector-Following", Zeitschrift für Physik D Atoms, Molecules and Clusters **40**, 194 (1997).

[42] M.-C. Marinica, F. Willaime, and J.-P. Crocombette, "Irradiation-Induced Formation of Nanocrystallites with *C*15 Laves Phase Structure in bcc Iron", Physical Review Letters **108**, 025501 (2012).

[43] F. Calvo, A. Fortunelli, F. Negreiros, and D. J. Wales, "Communication: Kinetics of Chemical Ordering in Ag-Au and Ag-Ni Nanoalloys", The Journal of Chemical Physics **139**, 111102 (2013).

[44] M. T. Oakley and R. L. Johnston, "Exploring the Energy Landscapes of Cyclic Tetrapeptides with Discrete Path Sampling", Journal of Chemical Theory and Computation **9**, 650 (2013).

[45] L. V. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales, and R. L. Johnston, "Structures and Energy Landscapes of Hydrated Sulfate Clusters", Journal of Chemical Theory and Computation **11**, 2377 (2015).

[46] A. Alavi, P. Hu, T. Deutsch, P. L. Silvestrelli, and J. Hutter, "CO Oxidation on Pt(111): An Ab Initio Density Functional Theory Study", Physical Review Letters **80**, 3650 (1998).

[47] Y. Xia, B. Zhang, J. Ye, Q. Ge, and Z. Zhang, "Acetone-Assisted Oxygen Vacancy Diffusion on TiO2(110)", The Journal of Physical Chemistry Letters **3**, 2970 (2012).

[48] G. Zhou, D.-W. Wang, L.-C. Yin, N. Li, F. Li, and H. Cheng, "Oxygen Bridges between NiO Nanosheets and Graphene for Improvement of Lithium Storage", ACS Nano **6**, 3214 (2012).

[49] D. J. Wales, "Locating Stationary Points for Clusters in Cartesian Coordinates", Journal of the Chemical Society, Faraday Transactions **89**, 1305 (1993).

[50] D. J. Wales, "Rearrangements of 55-Atom Lennard-Jones and $(C_{60})_{55}$ Clusters", The Journal of Chemical Physics **101**, 3750 (1994).

[51] G. Mills and H. Jónsson, "Quantum and Thermal Effects in H_2 Dissociative Adsorption: Evaluation of Free Energy Barriers in Multidimensional Quantum Systems", Physical Review Letters **72**, 1124 (1994).

[52] G. Mills, H. Jónsson, and G. K. Schenter, "Reversible Work Transition State Theory : Application to Dissociative Adsorption of Hydrogen", Surface Science **324**, 305 (1995).

[53] G. Henkelman and H. Jónsson, "Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points", The Journal of Chemical Physics **113**, 9978 (2000).

[54] G. Henkelman, B. P. Uberuaga, and H. Jónsson, "A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths", The Journal of Chemical Physics **113**, 9901 (2000).

[55] G. Henkelman and H. Jónsson, "A Dimer Method for Finding Saddle Points on High Dimensional Potential Surfaces Using Only First Derivatives", The Journal of Chemical Physics **111**, 7010 (1999).

[56] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson, "Comparison of Methods for Finding Saddle Points without Knowledge of the Final States", The Journal of Chemical Physics **121**, 9776 (2004).

[57] A. Heyden, A. T. Bell, and F. J. Keil, "Efficient Methods for Finding Transition States in Chemical Reactions: Comparison of Improved Dimer Method and Partitioned Rational Function Optimization Method", The Journal of Chemical Physics **123**, 224101 (2005).

[58] J. Kästner and P. Sherwood, "Superlinearly Converging Dimer Method for Transition State Search", The Journal of Chemical Physics **128**, 014106 (2008).

[59] S. Sakong, C. Mosch, A. Lozano, H. F. Busnengo, and A. Groß, "Lowering Energy Barriers in Surface Reactions through Concerted Reaction Mechanisms", ChemPhysChem **13**, 3467 (2012).

[60] Y.-F. Li, U. Aschauer, J. Chen, and A. Selloni, "Adsorption and Reactions of O2 on Anatase TiO2", Accounts of Chemical Research **47**, 3361 (2014).

[61] B. Schaefer, R. Pal, N. S. Khetrapal, M. Amsler, A. Sadeghi, V. Blum, X. C. Zeng, S. Goedecker, and L. Wang, "Isomerism and Structural Fluxionality in the Au_{26} and Au_{26}^- Nanoclusters", ACS Nano **8**, 7413 (2014).

[62] B. Schaefer, S. Mohr, M. Amsler, and S. Goedecker, "Minima Hopping Guided Path Search: An Efficient Method for Finding Complex Chemical Reaction Pathways", The Journal of Chemical Physics **140**, 214102 (2014).

[63] J. P. K. Doye, "Physical Perspectives on the Global Optimization of Atomic Clusters", in *Global Optimization – Scientific and Engineering Case Studies*, edited by J. D. Pintér (Springer, 2006), pp. 103–139.

[64] B. Schaefer, S. A. Ghasemi, S. Roy, and S. Goedecker, "Stabilized Quasi-Newton Optimization of Noisy Potential Energy Surfaces", The Journal of Chemical Physics **142**, 034112 (2015).

[65] See http://bigdft.org for source codes of SQNM, SQNS and MHGPS.

[66] S. De, B. Schaefer, A. Sadeghi, M. Sicher, D. G. Kanhere, and S. Goedecker, "Relation between the Dynamics of Glassy Clusters and Characteristic Features of their Energy Landscape", Physical Review Letters **112**, 083401 (2014).

[67] H. Eyring, "The Energy of Activation for Bimolecular Reactions Involving Hydrogen and the Halogens, According to the Quantum Mechanics", Journal of the American Chemical Society **53**, 2537 (1931).

[68] C. B. Anfinsen, "Principles that Govern the Folding of Protein Chains", Science **181**, 223 (1973).

[69] F. H. Stillinger and T. A. Weber, "Structural Aspects of the Melting Transition", in Proceedings of the Tenth Mexican Winter Meeting on Statistical Mechanics, Vol. 159 (1981).

[70] F. H. Stillinger and T. A. Weber, "Hidden Structure in Liquids", Physical Review A **25**, 978 (1982).

[71] L. Dixon and G. Szegö, "The Global Optimization Problem: An Introduction", in *Towards global optimisation 2* (North-Holland Pub. Co., 1978).

[72] P. G. Mezey, "Catchment Region Partitioning of Energy Hypersurfaces, I", Theoretica Chimica Acta **58**, 309 (1981).

[73] O. M. Becker and M. Karplus, "The Topology of Multidimensional Potential Energy Surfaces: Theory and Application to Peptide Structure and Kinetics", The Journal of Chemical Physics **106**, 1495 (1997).

[74] J. P. K. Doye and D. J. Wales, "Saddle Points and Dynamics of Lennard-Jones Clusters, Solids, and Supercooled Liquids", The Journal of Chemical Physics **116**, 3777 (2002).

[75] J. N. Murrell and K. J. Laidler, "Symmetries of Activated Complexes", Trans. Faraday Soc. **64**, 371 (1968).

[76] D. J. Wales and R. S. Berry, "Limitations of the Murrell-Laidler Theorem", Journal of the Chemical Society, Faraday Transactions **88**, 543 (1992).

[77] K. Fukui, "Formulation of the Reaction Coordinate", The Journal of Physical Chemistry **74**, 4161 (1970).

[78] W. Quapp and D. Heidrich, "Analysis of the Concept of Minimum Energy Path on the Potential Energy Surface of Chemically Reacting Systems", Theoretica Chimica Acta **66**, 245 (1984).

[79] P. Pechukas, "On Simple Saddle Points of a Potential Surface, the Conservation of Nuclear Symmetry Along Paths of Steepest Descent, and the Symmetry of Transition States", The Journal of Chemical Physics **64**, 1516 (1976).

[80] Y. Kumeda, D. J. Wales, and L. J. Munro, "Transition States and Rearrangement Mechanisms from Hybrid Eigenvector-Following and Density Functional Theory.: Application to $C_{10}H_{10}$ and Defect Migration in Crystalline Silicon", Chemical Physics Letters **341**, 185 (2001).

[81] D. A. Evans and D. J. Wales, "The Free Energy Landscape and Dynamics of Met-Enkephalin", The Journal of Chemical Physics **119**, 9947 (2003).

[82] S. C. Harvey, R. K.-Z. Tan, and T. E. Cheatham, "The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition", Journal of Computational Chemistry **19**, 726 (1998).

[83] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln", Annalen der Physik **389**, 457 (1927).

[84] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2004).

[85] G. A. Worth and L. S. Cederbaum, "Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection", Annual Review of Physical Chemistry **55**, 127 (2004).

[86] R. G. P. Weitao and Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1994).

[87] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Mineola N.Y., 2012).

[88] M. P. Nightingale and C. J. Umrigar, *Quantum Monte Carlo Methods in Physics and Chemistry* (Springer, Berlin, 1998).

[89] L. H. Thomas, "The Calculation of Atomic Fields", Mathematical Proceedings of the Cambridge Philosophical Society **23**, 542 (1927).

[90] E. Fermi, "Un Metodo Statistico per la Determinazione di alcune Prioprietà dell'Atomo", Rendicondi Accademia Nazionale de Lincei **6**, 602 (1927).

[91] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas", Physical Review **136**, B864 (1964).

[92] M. Levy, "Universal Variational Functionals of Electron Densities, First-Order Density Matrices, and Natural Spin-Orbitals and Solution of the V-Representability Problem", Proceedings of the National Academy of Sciences **76**, 6062 (1979).

[93] E. E. Lieb, "Density Functionals for Coulomb Systems", International Journal of Quantum Chemistry **24**, 243 (1983).

[94] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, International Series of Monographs on Chemistry (Oxford University Press, USA, Oxford, 1989).

[95] E. Engel and R. Dreizler, *Density Functional Theory: An Advanced Course*, Theoretical and Mathematical Physics (Springer, Berlin, 2011).

[96] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects", Physical Review **140**, A1133 (1965).

[97] R. Peverati and D. G. Truhlar, "Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad Spectrum of Databases in Chemistry and Physics", Philosophical Transactions of the Royal Society of London A **372**, 20120476 (2014).

[98] D. M. Ceperley and B. J. Alder, "Ground State of the Electron Gas by a Stochastic Method", Physical Review Letters **45**, 566 (1980).

[99] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis", Canadian Journal of Physics **58**, 1200 (1980).

[100] J. P. Perdew and A. Zunger, "Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems", Physical Review B **23**, 5048 (1981).

[101] L. A. Cole and J. P. Perdew, "Calculated Electron Affinities of the Elements", Physical Review A **25**, 1265 (1982).

[102] J. P. Perdew and Y. Wang, "Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy", Physical Review B **45**, 13244 (1992).

[103] A. D. Becke, "Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior", Physical Review A **38**, 3098 (1988).

[104] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation", Physical Review B **46**, 6671 (1992).

[105] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple", Physical Review Letters **77**, 3865 (1996).

[106] A. D. Becke, "Density-Functional Thermochemistry. III. The Role of Exact Exchange", The Journal of Chemical Physics **98**, 5648 (1993).

[107] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for Mixing Exact Exchange with Density Functional Approximations", The Journal of Chemical Physics **105**, 9982 (1996).

[108] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid Functionals Based on a Screened Coulomb Potential", The Journal of Chemical Physics **118**, 8207 (2003).

[109] S. A. Ghasemi, M. Amsler, R. G. Hennig, S. Roy, S. Goedecker, T. Lenosky, C. J. Umrigar, L. Genovese, T. Morishita, and K. Nishio, "Energy Landscape of Silicon Systems and Its Description by Force Fields, Tight Binding Schemes, Density Functional Methods, and Quantum Monte Carlo Methods", Physical Review B **81**, 214107 (2010).

[110] S. Goedecker, W. Hellmann, and T. Lenosky, "Global Minimum Determination of the Born-Oppenheimer Surface within Density Functional Theory", Physical Review Letters **95**, 055501 (2005).

[111] J. E. Jones, "On the Determination of Molecular Fields. II. From the Equation of State of a Gas", Proceedings of the Royal Society of London A **106**, 463 (1924).

[112] J. E. Jones and A. E. Ingham, "On the Calculation of Certain Crystal Potential Constants, and on the Cubic Crystal of Least Potential Energy", Proceedings of the Royal Society of London A **107**, 636 (1925).

[113] E. Whalley and W. G. Schneider, "Intermolecular Potentials of Argon, Krypton, and Xenon", The Journal of Chemical Physics **23**, 1644 (1955).

[114] A. Rahman, "Correlations in the Motion of Atoms in Liquid Argon", Physical Review **136**, A405 (1964).

[115] L. Rowley, D. Nicholson, and N. Parsonage, "Monte Carlo Grand Canonical Ensemble Calculation in a Gas-Liquid Transition Region For 12-6 Argon", Journal of Computational Physics **17**, 401 (1975).

[116] J. A. White, "Lennard-Jones as a Model for Argon and Test of Extended Renormalization Group Calculations", The Journal of Chemical Physics **111**, 9352 (1999).

[117] O. Talu and A. L. Myers, "Reference Potentials for Adsorption of Helium, Argon, Methane, and Krypton in High-Silica Zeolites", Colloids and Surfaces A: Physicochemical and Engineering Aspects **187-188**, 83 (2001).

[118] M. Born and J. Mayer, "Zur Gittertheorie der Ionenkristalle", Zeitschrift für Physik **75**, 1 (1932).

[119] J. E. Mayer, "Dispersion and Polarizability and the van der Waals Potential in the Alkali Halides", The Journal of Chemical Physics **1**, 270 (1933).

[120] M. L. Huggins and J. E. Mayer, "Interatomic Distances in Crystals of the Alkali Halides", The Journal of Chemical Physics **1**, 643 (1933).

[121] F. Fumi and M. Tosi, "Ionic Sizes and Born Repulsive Parameters in the NaCl-Type Alkali Halides-I: The Huggins-Mayer and Pauling Forms", Journal of Physics and Chemistry of Solids **25**, 31 (1964).

[122] M. Tosi and F. Fumi, "Ionic Sizes and Born Repulsive Parameters in the NaCl-Type Alkali Halides-II: The Generalized Huggins-Mayer Form", Journal of Physics and Chemistry of Solids **25**, 45 (1964).

[123] D. J. Adams and I. R. McDonald, "Rigid-Ion Models of the Interionic Potential in the Alkali Halides", Journal of Physics C: Solid State Physics **8**, 2198 (1975).

[124] L. Pauling, "The Influence of Relative Ionic Sizes on the Properties of Ionic Compounds", Journal of the American Chemical Society **50**, 1036 (1928).

[125] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules", Journal of the American Chemical Society **117**, 5179 (1995).

[126] D. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham, III, T. Darden, R. Duke, H. Gohlke, A. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. Merz, F. Paesani, D. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. Simmerling, W. Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, and P. Kollman, *AMBER 14* (University of California, San Francisco, 2014).

[127] T. J. Lenosky, B. Sadigh, E. Alonso, V. V. Bulatov, T. D. d. l. Rubia, J. Kim, A. F. Voter, and J. D. Kress, "Highly Optimized Empirical Potential Model of Silicon", Modelling and Simulation in Materials Science and Engineering **8**, 825 (2000).

[128] S. Goedecker, "Optimization and Parallelization of a Force Field for Silicon Using OpenMP", Computer Physics Communications **148**, 124 (2002).

[129] M. Miller, D. Wales, and V. de Souza, disconnectionDPS (Fortran program to generate disconnectivity graphs from stationary point databases, http://www-wales.ch.cam.ac.uk/software.html).

[130] H. Eyring, "The Activated Complex in Chemical Reactions", The Journal of Chemical Physics **3**, 107 (1935).

[131] M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems", Journal of Research of the National Bureau of Standards **49**, 409 (1952).

[132] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, "Structural Relaxation Made Simple", Phys. Rev. Lett. **97**, 170201 (2006).

[133] C. G. Broyden, "The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations", IMA Journal of Applied Mathematics **6**, 76 (1970).

[134] R. Fletcher, "A New Approach to Variable Metric Algorithms", The Computer Journal **13**, 317 (1970).

[135] D. Goldfarb, "A Family of Variable-Metric Methods Derived by Variational Means", Mathematics of Computation **24**, 23 (1970).

[136] D. F. Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization", Mathematics of Computation **24**, 647 (1970).

[137] J. Nocedal, "Updating Quasi-Newton Matrices with Limited Storage", Mathematics of Computation **35**, 773 (1980).

[138] D. C. Liu and J. Nocedal, "On the Limited Memory BFGS Method for Large Scale Optimization", Mathematical Programming **45**, 503 (1989).

[139] G. Crippen and H. Scheraga, "Minimization of Polypeptide Energy: XI. The Method of Gentlest Ascent", Archives of Biochemistry and Biophysics **144**, 462 (1971).

[140] D. Asenjo, J. D. Stevenson, D. J. Wales, and D. Frenkel, "Visualizing Basins of Attraction for Different Minimization Algorithms", The Journal of Physical Chemistry B **117**, 12717 (2013).

[141] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer, Berlin, 2006).

[142] A. J. Shepherd, *Second-Order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons* (Springer, Berlin, 2012).

[143] R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, New York, 2013).

[144] E. Haber, *Computational Methods in Geophysical Electromagnetics*, Mathematics in Industry (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014).

[145] L. Nazareth, "A Relationship between the BFGS and Conjugate Gradient Algorithms and its Implications for new Algorithms", SIAM Journal on Numerical Analysis **16**, 794 (1979).

[146] S. T. Chill, J. Stevenson, V. Ruehle, C. Shang, P. Xiao, J. D. Farrell, D. J. Wales, and G. Henkelman, "Benchmarks for Characterization of Minima, Transition States, and Pathways in Atomic, Molecular, and Condensed Matter Systems", Journal of Chemical Theory and Computation **10**, 5476 (2014).

[147] See http://users.iems.northwestern.edu/~nocedal/lbfgs.html for an implementation of L-BFGS.

[148] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and et al., "Daubechies Wavelets as a Basis Set for Density Functional Pseudopotential Calculations", The Journal of Chemical Physics **129**, 014109 (2008).

[149] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, and S. Goedecker, "Daubechies Wavelets for Linear Scaling Density Functional Theory", The Journal of Chemical Physics **140**, 204110 (2014).

[150] J. Tao, "Catalytic Activity Trends of CO Oxidation – A DFT Study", PhD thesis (Technical University of Denmark (DTU), 2011).

[151] M. K. Amsler, "Crystal Structure Prediction Based on Density Functional Theory", PhD thesis (University of Basel, 2011).

[152] F. Tassone, F. Mauri, and R. Car, "Acceleration Schemes for Ab Initio Molecular-Dynamics Simulations and Electronic-Structure Calculations", Physical Review B **50**, 10561 (1994).

[153] M. Probert, "Improved Algorithm for Geometry Optimisation Using Damped Molecular Dynamics", Journal of Computational Physics **191**, 130 (2003).

[154] C. J. Cerjan, "On Finding Transition States", The Journal of Chemical Physics **75**, 2800 (1981).

[155] P. Pulay, "Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration", Chemical Physics Letters **73**, 393 (1980).

[156] C. Lanczos, *Applied Analysis*, Dover Books on Mathematics (Dover Publications, Mineola N.Y., 1988).

[157] R. P. Brent, *Algorithms for Minimization without Derivatives* (Prentice-Hall, Englewood Cliffs, N.J, 1972).

[158] A. Behn, P. M. Zimmerman, A. T. Bell, and M. Head-Gordon, "Efficient Exploration of Reaction Paths via a Freezing String Method", The Journal of Chemical Physics **135**, 224108 (2011).

[159] P. Löwdin, "Quantum Theory of Cohesive Properties of Solids", Advances in Physics **5**, 1 (1956).

[160] I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry* (Springer, New York, 2003).

[161] F. Jensen, *Introduction to Computational Chemistry* (John Wiley & Sons, New York, 2007).

[162] D. Weinstein, "Modified Ritz Method", Proceedings of the National Academy of Sciences **20**, 529 (1934).

[163] Y. Suzuki and K. Varga, *Stochastic Variational Approach to Quantum-Mechanical Few-Body Problems* (Springer, Berlin, 1998).

[164] S. Goedecker, F. Lancon, and T. Deutsch, "Linear Scaling Relaxation of the Atomic Positions in Nanostructures", Physical Review B **64**, 161102 (2001).

[165] M. Page and J. W. McIver, "On Evaluating the Reaction Path Hamiltonian", The Journal of Chemical Physics **88**, 922 (1988).

[166] G. Mills and K. W. Jacobsen, "Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions", in *Classical and quantum dynamics in condensed phase simulations*, edited by G. C. B. J. Berne and D. F. Coker (World Scientific, 1998), pp. 385–404.

[167] A. Sadeghi, S. A. Ghasemi, B. Schaefer, S. Mohr, M. A. Lill, and S. Goedecker, "Metrics for Measuring Distances in Configuration Spaces", The Journal of Chemical Physics **139**, 184118 (2013).

[168] S. T. Chill, M. Welborn, R. Terrell, L. Zhang, J.-C. Berthet, A. Pedersen, H. Jónsson, and G. Henkelman, "EON: Software for Long Time Simulations of Atomic Scale Systems", Modelling and Simulation in Materials Science and Engineering **22**, 055002 (2014).

[169] G. Kresse and J. Hafner, "Ab Initio Molecular Dynamics for Liquid Metals", Physical Review B **47**, 558 (1993).

[170] G. Kresse and J. Hafner, "Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium", Physical Review B **49**, 14251 (1994).

[171] G. Kresse and J. Furthmüller, "Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set", Computational Materials Science **6**, 15 (1996).

[172] G. Kresse, "Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set", Physical Review B **54**, 11169 (1996).

[173] G. Kresse, "From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method", Physical Review B **59**, 1758 (1999).

[174] Z. Wang, L. L. Daemen, Y. Zhao, C. S. Zha, R. T. Downs, X. Wang, Z. L. Wang, and R. J. Hemley, "Morphology-Tuned Wurtzite-Rype ZnS Nanobelts", Nature Materials **4**, 922 (2005).

[175] C. R. Hickenboth, J. S. Moore, S. R. White, N. R. Sottos, J. Baudry, and S. R. Wilson, "Biasing Reaction Pathways with Mechanical Force", Nature **446**, 423 (2007).

[176] M. Moura, L. Broadbelt, and K. Tyo, "Computational Tools for Guided Discovery and Engineering of Metabolic Pathways", in *Systems metabolic engineering*, Vol. 985, edited by H. S. Alper, Methods in Molecular Biology (Humana Press, 2013), pp. 123–147.

[177] R. Granot and R. Baer, "A Spline for Your Saddle", The Journal of Chemical Physics **128**, 184111 (2008).

[178] S. A. Ghasemi and S. Goedecker, "An Enhanced Splined Saddle Method", The Journal of Chemical Physics **135**, 014108 (2011).

[179] E. F. Koslover and D. J. Wales, "Comparison of Double-Ended Transition State Search Methods", The Journal of Chemical Physics **127**, 134102 (2007).

[180] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark", Annual Review of Physical Chemistry **53**, 291 (2002).

[181] C. Dellago, P. G. Bolhuis, and P. L. Geissler, "Transition Path Sampling", in *Advances in chemical physics* (John Wiley & Sons, Inc., 2003), pp. 1–78.

[182] M. Grünwald, C. Dellago, and P. L. Geissler, "Precision Shooting: Sampling Long Transition Pathways", The Journal of Chemical Physics **129**, 194101 (2008).

[183] M. Grünwald and C. Dellago, "Nucleation and Growth in Structural Transformations of Nanocrystals", Nano Letters **9**, 2099 (2009).

[184] W. Lechner, C. Dellago, and P. G. Bolhuis, "Role of the Prestructured Surface Cloud in Crystal Nucleation", Physical Review Letters **106**, 085701 (2011).

[185] X.-J. Zhang and Z. Liu, "Reaction Sampling and Reactivity Prediction Using The Stochastic Surface Walking Method", Physical Chemistry Chemical Physics **17**, 2757 (2015).

[186] M. Qingming and P. Steenkiste, "On Path Selection for Traffic with Bandwidth Guarantees", in Proceedings 1997, International Conference on Network Protocols, edited by M. Ammar and U. Shankar (1997), pp. 191–202.

[187] E. Dijkstra, "A Note on Two Problems in Connexion With Graphs", Numer. Math. **1**, 269 (1959).

[188] C. Shang and Z. Liu, "Stochastic Surface Walking Method for Structure Prediction and Pathway Searching", Journal of Chemical Theory and Computation **9**, 1838 (2013).

[189] C. Shang and Z. Liu, "Constrained Broyden Dimer Method with Bias Potential for Exploring Potential Energy Surface of Multistep Reaction Process", Journal of Chemical Theory and Computation **8**, 2215 (2012).

[190] X.-J. Zhang, C. Shang, and Z. Liu, "Double-Ended Surface Walking Method for Pathway Building and Transition State Location of Complex Reactions", Journal of Chemical Theory and Computation **9**, 5745 (2013).

[191] C. Shang and Z. Liu, "Constrained Broyden Minimization Combined with the Dimer Method for Locating Transition State of Complex Reactions", Journal of Chemical Theory and Computation **6**, 1136 (2010).

[192] C. Dellago, "Transition Path Sampling and the Calculation of Free Energies", in *Free energy calculations*, Vol. 86, edited by C. Chipot and A. Pohorille, Springer Series in Chemical Physics (Springer Berlin Heidelberg, 2007), pp. 249–276.

[193] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines", The Journal of Chemical Physics **21**, 1087 (1953).

[194] C. Dellago, P. Bolhuis, and P. Geissler, "Transition Path Sampling Methods", in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, Vol. 703, edited by M. Ferrario, G. Ciccotti, and K. Binder, Lecture Notes in Physics (Springer Berlin Heidelberg, 2006), pp. 349–391.

[195] P. G. Bolhuis, C. Dellago, P. L. Geissler, and D. Chandler, "Transition Path Sampling: Throwing Ropes Over Mountains in the Dark", Journal of Physics: Condensed Matter **12**, A147 (2000).

[196] M. Grünwald, S. Jungblut, and C. Dellago, "Transition Path Sampling of Phase Transitions–Nucleation and Growth in Materials Hard and Soft", in *Hierarchical methods for dynamics in complex molecular systems*, Vol. 10, edited by J. Grotendorst, G. Sutmann, G. Gompper, and D. Marx (Forschungszentrum Jülich, 2012).

[197] W. Hellmann, R. G. Hennig, S. Goedecker, C. J. Umrigar, B. Delley, and T. Lenosky, "Questioning the Existence of a Unique Ground-State Structure for Si Clusters", Physical Review B **75**, 085411 (2007).

[198] K. Bao, S. Goedecker, K. Koga, F. Lançon, and A. Neelov, "Structure of Large Gold Clusters Obtained by Global Optimization Using the Minima Hopping Method", Physical Review B **79**, 041405 (2009).

[199] S. Goedecker, "Global Optimization with the Minima Hopping Method", in *Modern methods of crystal structure prediction*, edited by A. R. Oganov (Wiley-VCH, 2011), pp. 131–145.

[200] M. Sicher, S. Mohr, and S. Goedecker, "Efficient Moves for Global Geometry Optimization Methods and their Application to Binary Systems", The Journal of Chemical Physics **134**, 044106 (2011).

[201] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[202] M. D. Wolf and U. Landman, "Genetic Algorithms for Structural Cluster Optimization", Journal of Physical Chemistry A **102**, 6129 (1998).

[203] B. Hammer and J. K. Norskov, "Why Gold is the Noblest of All the Metals", Nature **376**, 238 (1995).

[204] X. Gu, S. Bulusu, X. Li, X. Zeng, J. Li, X. Gong, and L. S. Wang, "Au_{34}^-: A Fluxional Core-Shell Cluster", Journal of Physical Chemistry C **111**, 8228 (2007).

[205] A. Sanchez, S. Abbet, U. Heiz, W.-D. Schneider, H. Häkkinen, R. N. Barnett, and U. Landman, "When Gold is Not Noble: Nanoscale Gold Catalysts", Journal of Physical Chemistry A **103**, 9573 (1999).

[206] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, "A DNA-Based Method for Rationally Assembling Nanoparticles into Macroscopic Materials", Nature **382**, 607 (1996).

[207] A. P. Alivisatos, K. P. Johnsson, X. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez, and P. G. Schultz, "Organization of 'Nanocrystal Molecules' Using DNA", Nature **382**, 609 (1996).

[208] S. Chen, "Gold Nanoelectrodes of Varied Size: Transition to Molecule-Like Charging", Science **280**, 2098 (1998).

[209] P. Pyykkö, "Theoretical Chemistry of Gold", Angewandte Chemie. International Edition in English **43**, 4412 (2004).

[210] M. Haruta, "Size- and Support-Dependency in the Catalysis of Gold", Catalysis Today **36**, 153 (1997).

[211] C. Cleveland, U. Landman, T. Schaaff, M. Shafigullin, P. Stephens, and R. Whetten, "Structural Evolution of Smaller Gold Nanocrystals: The Truncated Decahedral Motif", Physical Review Letters **79**, 1873 (1997).

[212] I. Garzón, K. Michaelian, M. Beltrán, A. Posada-Amarillas, P. Ordejón, E. Artacho, D. Sánchez-Portal, and J. Soler, "Lowest Energy Structures of Gold Nanoclusters", Physical Review Letters **81**, 1600 (1998).

[213] G. Bravo-Pérez, I. Garzón, and O. Novaro, "Ab Initio Study of Small Gold Clusters", Journal of Molecular Structure: THEOCHEM **493**, 225 (1999).

[214] H. Häkkinen and U. Landman, "Gold Clusters (Au_N, $2 \leq N \leq 10$) and their Anions", Physical Review B **62**, R2287 (2000).

[215] F. Furche, R. Ahlrichs, P. Weis, C. Jacob, S. Gilb, T. Bierweiler, and M. M. Kappes, "The Structures of Small Gold Cluster Anions as Determined by a Combination of Ion Mobility Measurements and Density Functional Calculations", The Journal of Chemical Physics **117**, 6982 (2002).

[216] S. Gilb, P. Weis, F. Furche, R. Ahlrichs, and M. M. Kappes, "Structures of Small Gold Cluster Cations (Au_n^+, $n < 14$): Ion Mobility Measurements Versus Density Functional Calculations", The Journal of Chemical Physics **116**, 4094 (2002).

[217] B. Yoon, H. Häkkinen, and U. Landman, "Interaction of O_2 with Gold Clusters: Molecular and Dissociative Adsorption", Journal of Physical Chemistry A **107**, 4066 (2003).

[218] H. Häkkinen, B. Yoon, U. Landman, X. Li, H. Zhai, and L. S. Wang, "On the Electronic and Atomic Structures of Small Au_N^- ($N = 4-14$) Clusters: A Photoelectron Spectroscopy and Density-Functional Study", Journal of Physical Chemistry A **107**, 6168 (2003).

[219] H. Häkkinen, S. Abbet, A. Sanchez, U. Heiz, and U. Landman, "Structural, Electronic, and Impurity-Doping Effects in Nanoscale Chemistry: Supported Gold Nanoclusters", Angewandte Chemie. International Edition in English **42**, 1297 (2003).

[220] H. M. Lee, M. Ge, B. R. Sahu, P. Tarakeshwar, and K. S. Kim, "Geometrical and Electronic Structures of Gold, Silver, and Gold-Silver Binary Clusters: Origins of Ductility of Gold and Gold-Silver Alloy Formation", The Journal of Physical Chemistry B **107**, 9994 (2003).

[221] J. Li, X. Li, H. Zhai, and L. S. Wang, "Au_{20}: A Tetrahedral Cluster", Science **299**, 864 (2003).

[222] E. Fernández, J. Soler, I. Garzón, and L. Balbás, "Trends in the Structure and Bonding of Noble Metal Clusters", Physical Review B **70**, 165403 (2004).

[223] M. Neumaier, F. Weigend, O. Hampe, and M. M. Kappes, "Binding Energies of CO on Gold Cluster Cations Au_n^+ ($n = 1-65$): A Radiative Association Kinetics Study", The Journal of Chemical Physics **122**, 104702 (2005).

[224] A. Lechtken, D. Schooss, J. R. Stairs, M. N. Blom, F. Furche, N. Morgner, O. Kostko, B. von Issendorff, and M. M. Kappes, "Au_{34}^-: A Chiral Gold Cluster?", Angewandte Chemie. International Edition in English **46**, 2944 (2007).

[225] S. Bulusu, X. Li, L. S. Wang, and X. C. Zeng, "Structural Transitions from Pyramidal to Fused Planar to Tubular to Core/Shell Compact in Gold Clusters: Au_n^- (n = 21-25)", Journal of Physical Chemistry C **111**, 4190 (2007).

[226] A. F. Jalbout, F. F. Contreras-Torres, L. A. Pérez, and I. L. Garzón, "Low-Symmetry Structures of Au_{32}^Z (Z = +1, 0, -1) Clusters", Journal of Physical Chemistry A **112**, 353 (2008).

[227] I. E. Santizo, F. Hidalgo, L. A. Pérez, C. Noguez, and I. L. Garzón, "Intrinsic Chirality in Bare Gold Nanoclusters: The Au_{34}^- Case", Journal of Physical Chemistry C **112**, 17533 (2008).

[228] M. Johansson, A. Lechtken, D. Schooss, M. Kappes, and F. Furche, "2D-3D Transition of Gold Cluster Anions Resolved", Physical Review A **77**, 053202 (2008).

[229] W. Huang and L. S. Wang, "Probing the 2D to 3D Structural Transition in Gold Cluster Anions Using Argon Tagging", Physical Review Letters **102**, 153401 (2009).

[230] W. Huang, S. Bulusu, R. Pal, X. C. Zeng, and L. S. Wang, "Structural Transition of Gold Nanoclusters: From the Golden Cage to the Golden Pyramid", ACS Nano **3**, 1225 (2009).

[231] A. Lechtken, C. Neiss, M. M. Kappes, and D. Schooss, "Structure Determination of Gold Clusters by Trapped Ion Electron Diffraction: $Au_{14}^- - Au_{19}^-$", Physical Chemistry Chemical Physics **11**, 4344 (2009).

[232] D. Schooss, P. Weis, O. Hampe, and M. M. Kappes, "Determining the Size-Dependent Structure of Ligand-Free Gold-Cluster Ions", Philosophical Transactions of the Royal Society of London A **368**, 1211 (2010).

[233] L. Wang, R. Pal, W. Huang, X. C. Zeng, and L. S. Wang, "Observation of Earlier Two-to-Three Dimensional Structural Transition in Gold Cluster Anions by Isoelectronic Substitution: MAu_n^- ($n = 8-11$; M=Ag,Cu)", The Journal of Chemical Physics **132**, 114306 (2010).

[234] W. Huang, R. Pal, L. Wang, X. C. Zeng, and L. S. Wang, "Isomer Identification and Resolution in Small Gold Clusters", The Journal of Chemical Physics **132**, 054305 (2010).

[235] N. Shao, W. Huang, Y. Gao, L. Wang, X. Li, L. S. Wang, and X. C. Zeng, "Probing the Structural Evolution of Medium-Sized Gold Clusters: Au_n^- ($n = 27-35$)", Journal of the American Chemical Society **132**, 6596 (2010).

[236] R. Pal, L. Wang, W. Huang, L. S. Wang, and X. C. Zeng, "Structure Evolution of Gold Cluster Anions between the Planar and Cage Structures by Isoelectronic Substitution: Au_n^- ($n = 13-15$) and MAu_n^- ($n = 12-14$; M = Ag, Cu)", The Journal of Chemical Physics **134**, 054306 (2011).

[237] L. Wang and L. S. Wang, "Probing the Electronic Properties and Structural Evolution of Anionic Gold Clusters in the Gas Phase", Nanoscale **4**, 4038 (2012).

[238] I. León, Z. Yang, and L. S. Wang, "High Resolution Photoelectron Imaging of Au_2^-", The Journal of Chemical Physics **138**, 184304 (2013).

[239] Z. Yang, I. León, and L. S. Wang, "Communication: Vibrational Spectroscopy of Au_4 from High Resolution Photoelectron Imaging", The Journal of Chemical Physics **139**, 021106 (2013).

[240] N. Shao, W. Huang, W. Mei, L. S. Wang, Q. Wu, and X. C. Zeng, "Structural Evolution of Medium-Sized Gold Clusters Au_n^- (n =36, 37, 38): Appearance of Bulk-Like Face Centered Cubic Fragment", Journal of Physical Chemistry C **118**, 6887 (2014).

[241] C. Hartwigsen, S. Goedecker, and J. Hutter, "Relativistic Separable Dual-Space Gaussian Pseudopotentials From H to Rn", Physical Review B **58**, 3641 (1998).

[242] M. Krack, "Pseudopotentials for H to Kr Optimized for Gradient-Corrected Exchange-Correlation Functionals", Theoretical Chemistry Accounts **114**, 145 (2005).

[243] W. Fa and J. Dong, "Possible Ground-State Structure of Au_{26}: A Highly Symmetric Tubelike Cage", The Journal of Chemical Physics **124**, 114310 (2006).

[244] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals", Computer Physics Communications **180**, 2175 (2009).

[245] Y. Zhao, N. González-García, and D. G. Truhlar, "Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and its Use to Test Theoretical Methods", Journal of Physical Chemistry A **109**, 2012 (2005).

[246] S. Andersson and M. Grüning, "Performance of Density Functionals for Calculating Barrier Heights of Chemical Reactions Relevant to Astrophysics", Journal of Physical Chemistry A **108**, 7621 (2004).

[247] M. Grüning, O. V. Gritsenko, and E. J. Baerends, "Improved Description of Chemical Barriers with Generalized Gradient Approximations (GGAs) and Meta-GGAs", Journal of Physical Chemistry A **108**, 4459 (2004).

[248] A. Zupan, K. Burke, M. Ernzerhof, and J. P. Perdew, "Distributions and Averages of Electron Density Parameters: Explaining the Effects of Gradient Corrections", The Journal of Chemical Physics **106**, 10184 (1997).

[249] Y. Zhao and D. G. Truhlar, "The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals", Theoretical Chemistry Accounts **120**, 215 (2007).

[250] G. Barcaro and A. Fortunelli, "Structure and Diffusion of Small Ag and Au Clusters on the Regular MgO (100) Surface", New J. Phys. **9**, 22 (2007).

[251] A. Dal Corso, A. Pasquarello, and A. Baldereschi, "Density-Functional Perturbation Theory for Lattice Dynamics with Ultrasoft Pseudopotentials", Physical Review B **56**, R11369 (1997).

[252] P. Haas, F. Tran, and P. Blaha, "Calculation of the Lattice Constant of Solids with Semilocal Functionals", Physical Review B **79**, 85104 (2009).

[253] A. Dal Corso, "Ab Initio Phonon Dispersions of Transition and Noble Metals: Effects of the Exchange and Correlation Functional", Journal of Physics: Condensed Matter **25**, 145401 (2013).

[254] T. T. Järvi, A. Kuronen, M. Hakala, K. Nordlund, A. C. van Duin, W. A. Goddard, and T. Jacob, "Development of a ReaxFF Description for Gold", European Physical Journal B: Condensed Matter Physics **66**, 75 (2008).

[255] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, "NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations", Computer Physics Communications **181**, 1477 (2010).

[256] L. S. Wang, H. S. Cheng, and J. Fan, "Photoelectron Spectroscopy of Size-Selected Transition Metal Clusters: Fe_n^-, $n = 3-24$", The Journal of Chemical Physics **102**, 9480 (1995).

[257] J. Akola, M. Manninen, H. Häkkinen, U. Landman, X. Li, and L. S. Wang, "Photoelectron Spectra of Aluminum Cluster Anions: Temperature Effects and Ab Initio Simulations", Physical Review B **60**, R11297 (1999).

[258] W. Huang and L. S. Wang, "Au_{10}^-: Isomerism and Structure-Dependent O_2 reactivity", Physical Chemistry Chemical Physics **11**, 2663 (2009).

[259] D. Tian and J. Zhao, "Competition Among fcc-Like, Double-Layered Flat, Tubular Cage, and Close-Packed Structural Motifs for Medium-Sized Au_n ($n = 21-28$) Clusters", Journal of Physical Chemistry A **112**, 3141 (2008).

[260] S. De, A. Willand, M. Amsler, P. Pochet, L. Genovese, and S. Goedecker, "Energy Landscape of Fullerene Materials: A Comparison of Boron to Boron Nitride and Carbon", Physical Review Letters **106**, 225502 (2011).

[261] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, "Prescription for the Design and Selection of Density Functional Approximations: More Constraint Satisfaction with Fewer Fits", The Journal of Chemical Physics **123**, 062201 (2005).

[262] L. Ferrighi, B. Hammer, and G. K. H. Madsen, "2D-3D Transition for Cationic and Anionic Gold Clusters: A Kinetic Energy Density Functional Study", Journal of the American Chemical Society **131**, 10605 (2009).

[263] M. Mantina, R. Valero, and D. G. Truhlar, "Validation Study of the Ability of Density Functionals to Predict the Planar-to-Three-Dimensional Structural Transition in Anionic Gold Clusters", The Journal of Chemical Physics **131**, 064706 (2009).

[264] K. J. Taylor, C. L. Pettiette-Hall, O. Cheshnovsky, and R. E. Smalley, "Ultraviolet Photoelectron Spectra of Coinage Metal Clusters", The Journal of Chemical Physics **96**, 3319 (1992).

[265] C. Levinthal, "How to Fold Graciously", in Mössbauer Spectroscopy in Biological Systems: Proceedings of a Meeting held at Allerton House, Monticello, Illinois, edited by P. Debrunner, J. C. M. Tsibris, and E. Münck (1969), pp. 22–24.

[266] K. A. Dill and H. S. Chan, "From Levinthal to Pathways to Funnels", Nature Structural Biology **4**, 10 (1997).

[267] R. P. Bell, "The Theory of Reactions Involving Proton Transfers", Proceedings of the Royal Society of London A **154**, 414 (1936).

[268] M. G. Evans and M. Polanyi, "Further Considerations on the Thermodynamics of Chemical Equilibria and Reaction Rates", Trans. Faraday Soc. **32**, 1333 (1936).

[269] G. S. Hammond, "A Correlation of Reaction Rates", Journal of the American Chemical Society **77**, 334 (1955).

[270] R. A. Marcus, "Theoretical Relations Among Rate Constants, Barriers, and Brønsted Slopes of Chemical Reactions", The Journal of Physical Chemistry **72**, 891 (1968).

[271] K. Levenberg, "A Method for the Solution of Certain Non-linear Problems in Least Squares", Quarterly of Applied Mathematics **II**, 164 (1944).

[272] D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", Journal of the Society for Industrial and Applied Mathematics **11**, 431 (1963).

[273] T. Williams, C. Kelley, and many others, *Gnuplot 4.6: An Interactive Plotting Program*, http://gnuplot.sourceforge.net, 2014.

[274] G. Jorgen Pedersen, "Über die Entwickelung Reeller Funktionen in Reihen Mittelst der Methode der Kleinsten Quadrate", Journal für die reine und angewandte Mathematik **94**, 41 (1883).

[275] S. Erhard, "Zur Theorie der Linearen und Nichtlinearen Integralgleichungen. I. Teil: Entwicklung Willkürlicher Funktionen Nach Systemen Vorgeschriebener", Mathematische Annalen **63**, 433 (1907).

[276] W. Kabsch, "A Discussion of the Solution for the Best Rotation to Relate Two Sets of Vectors", Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography **34**, 827 (1978).

[277] E. A. Coutsias, C. Seok, and K. A. Dill, "Using Quaternions to Calculate RMSD", Journal of Computational Chemistry **25**, 1849 (2004).

[278] H. W. Kuhn, "The Hungarian Method for the Assignment Problem", Naval Research Logistics Quarterly **2**, 83 (1955).

[279] H. W. Kuhn, "Variants of the Hungarian Method for Assignment Problems", Naval Research Logistics Quarterly **3**, 253 (1956).