Ab Initio Modelling of Chemical Vapor and Area-Selective Atomic Layer Deposition

Developing an Automated Exploration of Surface Reaction Networks

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Chemie der Philipps-Universität Marburg

Vorgelegt von

Fabian Pieck, M.Sc.

Aus Dillenburg

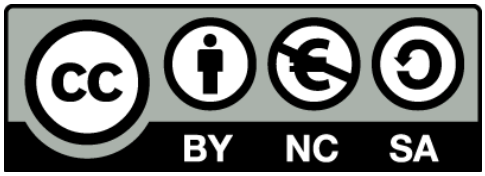
Marburg, 2022

Die vorliegende Dissertation wurde von Mai 2017 bis August 2022 an dem Fachbereich Chemie der Philipps-Universität Marburg und der Fakultät für Chemie und Mineralogie der Universität Leipzig unter Leitung von Prof. Dr. Ralf Tonner-Zech angefertigt.

Vom Fachbereich Chemie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation angenommen am 11.08.2022

Erstgutachter: Prof. Dr. Ralf Tonner-Zech
Zweitgutachterin: Prof. Dr. Kerstin Volz

Tag der Disputation: 20.10.2022



Danksagung

Ich möchte an dieser Stelle all jenen Danken, die mir in den letzten Jahren im Privaten als auch beruflich zur Seite gestanden haben und diese Arbeit erst möglich gemacht haben. Ich möchte Danke sagen für all die Ratschläge, Diskussionen und Hilfestellungen für meine wissenschaftliche Arbeit, ebenso wie für die schönen Ablenkungen und nicht-wissenschaftlichen Aktivitäten.

Mein besonderer Dank gilt Prof. Dr. Ralf Tonner-Zech für die Betreuer meiner Dissertation, die Vielzahl an Diskussionen und hilfreichen Ideen, die zu dieser Promotionsarbeit maßgeblich beigetragen haben. Besonders hervorheben möchte ich die große Freiheit eigene Ideen zu entwickeln und ihnen nachzugehen auch wenn damit der eine oder andere Umweg und entsprechende Zeitaufwand verbunden war. Mir hat die Arbeit an meinem Promotionsprojekt viel Freude bereitet, genauso wie die Einbindung in Kooperationsprojekte in verschiedensten Themenfelder der Chemie. Danken möchte ich auch Prof. Dr. Kerstin Volz für die Erstellung des Zweitgutachtens und Prof. Dr. Michael Gottfried für die Mitwirkung in meiner Prüfungskommission.

Meiner Arbeitsgruppe möchte ich für die schöne Atmosphäre, der auch mehrere Jahre Pandemie nichts anhaben konnte, danken. Ich habe eure kritischen Rückmeldungen und praktischen Hilfestellungen bei meiner alltäglichen wissenschaftlichen Arbeit, meinen Vorträgen und vor allem auch dem Korrekturlesen dieser Arbeit sehr zu schätzen gelernt. Dankbar bin ich für die Möglichkeiten der letzten Jahre in Seminaren, Übungen und Praktika andere Studenten und Studentinnen bei ihren ersten Schritten in der theoretischen Chemie anzuleiten und dem Vertrauen, das mir dabei entgegengebracht wurde. Hervorheben möchte ich hier Daniel Grigjanis, Patrick Maue, Alexander Kafka und Moritz Richter für ihre Mitarbeit an meinen Projekten.

Danken möchte ich auch Prof. Dr. Stacey F. Bent und Prof. Dr. Han-Bo-Ram Lee und ihren Arbeitsgruppen für die interessante und spannende Arbeit innerhalb der gemeinsamen Forschungsprojekte und die vielen teils nächtlichen Treffen mit Diskussionen, Fragen und Anregungen zu meinen Arbeiten bezüglich der Atomlagendeposition. Hierbei möchte ich besonders Josiah Yarbrough und Summal Zoha für die Zusammenarbeit danken. Ebenso möchte ich mich bei Dr. Greta Linden, Dr. Eike Dornsiepen und Dr. Tania E. Sandoval für die gemeinsamen Arbeiten und die spannenden Einblicke bedanken.

An dieser Stelle möchte ich auch die DFG und die Merck KGaA, Darmstadt erwähnen, die mich durch die Finanzierung meiner Promotion innerhalb des GRK 1782 „Functionalization of Se-

miconductors“ beziehungsweise des „350th anniversary Research Grant“ unterstützt haben. Zudem wäre diese Arbeit ohne die Ressourcen der Rechenzentren der Universitäten in Marburg, Frankfurt, Stuttgart und Regensburg nicht möglich gewesen.

Contents

List of Publications

The following list contains all submitted publications and manuscripts in preparation originated during the doctoral program at the Philipps-Universität Marburg in chronological order.

Accepted Publications

Fabian Pieck, Jan-Niclas Luy, Florian Kreuter, Badal Mondal, Ralf Tonner-Zech, *Reactivity of Organic Molecules on Semiconductor Surfaces Revealed by Density Functional Theory* in *High Performance Computing in Science and Engineering ' 21*, (Eds.: W. E. Nagel, D. H. Kröner, M. M. Resch), Springer, Cham, Switzerland, in press.

Josiah Yarbrough, **Fabian Pieck**, Daniel Grigjanis, Il-Kwon Oh, Patrick Maue, Ralf Tonner-Zech, Stacey F. Bent, *Tuning Molecular Inhibitors and Aluminum Precursors for the Area-Selective Atomic Layer Deposition of Al_2O_3*, *Chem. Mater.* **2022**, *34*, 4646–4659, DOI: 10.1021/acs.chemmater.2c00513.

Fabian Pieck, Ralf Tonner-Zech, *Alkyne-Functionalized Cyclooctyne on Si(001): Reactivity Studies and Surface Bonding from an Energy Decomposition Analysis Perspective*, *Molecules* **2021**, *26*, 6653, DOI: 10.3390/molecules26216653.

Tania E. Sandoval, **Fabian Pieck**, Ralf Tonner, and Stacey F. Bent, *Effect of Heteroaromaticity on Adsorption of Pyrazine on the Ge(100)-2x1 Surface*, *J. Phys. Chem. C* **2020**, *124*, 22055–22068, DOI: 10.1021/acs.jpcc.0c04673.

Eike Dornsiepen, **Fabian Pieck**, Ralf Tonner, Stefanie Dehnen, *[{$(PhSn)_3SnS_6$}{$(MCp)_3S_4$}] (M = W, Mo): Minimal Molecular Models of the Covalent Attachment of Metal Chalcogenide Clusters on Doped Transition Metal Dichalcogenide Layers*, *J. Am. Chem. Soc.* **2019**, *141*, 16494–16500, DOI: 10.1021/jacs.9b09209.

Greta Linden, Lei Zhang, **Fabian Pieck**, Uwe Linne, Dmitri Kosenkov, Ralf Tonner, Olalla Vázquez, *Conditional Singlet Oxygen Generation through a Bioorthogonal DNA-targeted Tetrazine Reaction*, *Angew. Chem.* **2019**, *131*, 13000–13005, DOI: 10.1002/ange.201907093. *Angew. Chem. Int. Ed.* **2019**, *58*, 12868–12873, DOI: 10.1002/anie.201907093.

Fabian Pieck, Lisa Pecher, Jan-Niclas Luy, Ralf Tonner, *Inorganic and Organic Functionalisation of Silicon Studied by Density Functional Theory* in *High Performance Computing in Science and Engineering ' 18*, (Eds.: W. E. Nagel, D. H. Kröner, M. M. Resch), Springer, Cham, Switzerland, **2019**, p. 153–166, DOI: 10.1007/978-3-030-13325-2_10.

Manuscripts in Preparation

Summal Zoha, Bonwook Gu, **Fabian Pieck**, Ralf Tonner-Zech, Han-Bo-Ram Lee, *Organothiol Inhibitor Instigated Passivation of Multiple Substrates for Area-Selective Atomic Layer Deposition of HfO_2*.

Josiah Yarbrough, **Fabian Pieck**, Alex Shearer, Patrick Maue, Ralf Tonner-Zech, Stacey F. Bent, *Area-Selective Al_2O_3 Atomic Layer Deposition with Molecular Deactivation of Copper by Methanesulfonic Acid*.

Abbreviations & Acronyms

Theory and Math

DFT	density functional theory
SD	Slater determinant
SCF	self-consistent field
LDA	local density approximation
GGA	generalized gradient approximation
LCAO	linear combination of atomic orbitals
AO	atomic orbital
STO	Slater type orbital
GTO	Gaussian type orbital
PBC	periodic boundary conditions
PAW	projector augmented-wave
PES	potential energy surface
NEB	nudged elastic band
MEP	minimum energy path
CI	climbing image
TS	transition state
CG	conjugate gradient
IRC	internal reaction coordinate
RMSD	root mean square deviation
H-RMSD	root mean square deviation minimized by the Hungarian algorithm
SH-RMSD	root mean square deviation minimized by the Hungarian algorithm under consideration of the surface symmetry
GF	Gaussian function
COM	center of mass
SVD	singular value decomposition

PESE

PESE	PESE - a Potential Energy Surface Explorer for thin film growth
RMG	reaction mechanism generator
MD	molecular dynamic
SSW	stochastic surface walking
BP-CBD	bias-potential driven constrained-Broyden-dimer
GSM	growing string method
GRRM	global reaction route mapping
ADDF	anharmonic downward distortion following
AFIR	artificial force induced reaction
kMC	kinetic Monte Carlo
CPU	central processing unit
RAM	random access memory
GPU	graphic processing unit
RAID	redundant array of inexpensive disks
HDD	hard disk drive
SSD	solid-state drive
MPI	message passing interface
OpenMP	open multi-processing

ML	machine learning
xTB	extended tight-binding

Atomic Layer Deposition

ALD	atomic layer deposition
ASD	area selective deposition
AS-ALD	area selective atomic layer deposition
GS	growth surface
NGS	non-growth surface
SAM	self assembled monolayer
SMI	small molecule inhibitor
MSA	methanesulfonic acid
DES	diethyl ether
MTMS	methoxytrimethylsilane
DMDMS	dimethoxydimethylsilane
TMMS	trimethoxymethylsilane
TMES	trimethoxyethylsilane
TMPS	trimethoxypropylsilane
TMA	trimethylaluminium
TEA	triethylaluminium
DMAI	dimethylaluminum isopropoxide

Chemical Vapor Deposition

fcc	face-centered cubic
CVD	chemical vapor deposition
MOCVD	metalorganic chemical vapor deposition
MOVPE	metalorganic vapor phase epitaxy
LED	light-emitting diode
TBP	tert-butylphosphine
TBAs	tert-butylarsine
DTBAA	di-tertiary-butyl-arsano-amine
TMGa	trimethylgallium
TEGa	triethylgallium
DTBAP	di-tert-butylaminophosphane

1

Introduction

In our modern life semiconductor materials, among which silicon is the most prominent, are of outstanding importance[1]. Their relevance is based on their usage in every electronic device[2]. Most prominent examples are transistors[3–7], which are the most basic building block of every central processing unit (CPU) and random access memory (RAM), and solar cells[8–10], which hold great promise for the supply with renewable energy. Furthermore, without semiconductor materials the production of lasers[11–16] and light-emitting diodes (LEDs)[17–22] for the emission of light at a certain wavelength is unthinkable. Also, electronic devices as photodetectors[23–26] for the detection of light and gas sensors[27–30] are fabricated based on semiconductors.

By now, the most common semiconductor materials used by the industry are silicon (Si), germanium (Ge) and gallium arsendide (GaAs)[31]. Here, Si stands out due to its low costs while Ge and GaAs show more favorable electronic and optical properties like a high hole or electron mobility or a direct band gap[32–34]. Especially, the direct band gap is mandatory for a all electronic devices emitting or absorbing light while aiming at a high efficiency. Consequently, GaAs is used as the active material for these devices instead of silicon due to its direct band gap as shown in Figure 1.1. GaAs is one of the so called III-V compound semiconductors since it is a combination of a group 13 (main group III) with a group 15 (main group V) element. As shown in Figure 1.1, a beneficial characteristic of III-V compound semiconductors is that the band gap and lattice constant of the material can be fine-tuned by changing its composition. Since even ternary and quaternary materials can be produced, a huge amount of different semiconductor materials are available.

Besides Si, Ge and III-V semiconductors also 2D materials[35] as for example MoS_2 and WS_2 are heavily investigated regarding their usage in electronic devices. 2D materials are beneficial since they are intrinsically restricted to the nanoscale size in one dimension[35]. This property stimulates the study of 2D materials with the advancing miniaturisation of electronic components in mind. Noteworthy, the miniaturisation is not only driving the development of

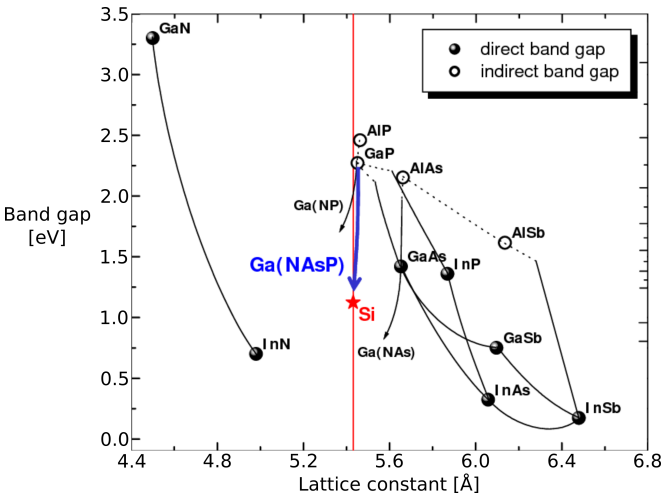


Figure 1.1.: Relation between composition, lattice constant and band gap for III-V compound semiconductors. Adapted reprint with permission from reference [32].

semiconductor materials but also the study of high-κ dielectrics as Al_2O_3 and HfO_2[34,36]. Here, thin and yet functional dielectric layers are needed to separate (semi-) conducting materials as for example in transistors.

Techniques for the deposition of thin material films are atomic layer deposition (ALD)[34,37–41] and chemical vapor deposition (CVD)[2,35,42–46]. Both processes use volatile molecules called precursors or coreactants to transport the desired elements for the material growth to the substrate surface at which the material is deposited. However, the ALD and CVD processes differ in their overall approach as explained with the Figures 1.2 and 1.3.

The ALD process is schematically shown in Figure 1.2. In a first step a single precursor is pulsed into the reaction chamber. Here, the precursor is adsorbing and reacting with the reactive surface sites to form the first deposited material layer. Reaction products are desorbing and removed in a subsequent purging step with an inert carrier gas as N_2 or Ar. The combination of the precursor deposition and purging step is also called the first ALD half-cylcle. As a third step, the second precursor or coreactant is pulsed in the reaction chamber. This coreactant is then reacting with the reactive groups of the surface, which are now the residual groups of the first precursor. Again, a purging step is performed to complete the second half-cylce. By repeating the presented steps material growth is obtained. By combining several half-cycles of different materials in a so called supercycle even ternary and quaternary materials can be grown.[41]

As shown in Figure 1.2, the ALD process is dominated by surface reactions. These are ideally self-limiting in the sense that in every deposition step no more than a single monolayer is

formed. This requires that neither the precursor nor the coreactant are reacting with themselves or thermally decomposing. Also, lower temperatures of less than $350\,^{\circ}\mathrm{C}$ are common. Furthermore, the temperature has to be tuned for every ALD process to prevent slow reaction kinetics and the condensation of the gaseous species at too low temperatures as well as thermal decomposition and desorption of precursors at too high temperatures.

The huge advantage of the cyclic, self-limiting surface reactions within the ALD process is that a high control over the reactions and thereby the conformality , film thickness and composition is achieved. However, the requirement of self-limiting reactions results in a tedious development of new precursors and coreactants. Also, slow deposition rates are common for the ALD process due to the time consuming combination of growth and purge steps. Still, approaches like spatial ALD[47] are developed with the aim to overcome this shortcoming. Overall, the ALD process offers an excellent level of control over the growth process but is limited by the available precursors and suited surface reaction pathways.

In comparison to the ALD process, the CVD process shows two key differences. On the one hand, the CVD process is not solely relying on adsorbate-surface reactions. Here, the supply of precursors is not strictly separated by a purge step although such an approach can be used[48]. As a consequence, reactions between the precursors in the gas phase or with a reactive carrier gas as H_2 occur within the CVD process. On the other hand, CVD processes are commonly conducted at higher temperatures than ALD processes. Thereby, the thermal decomposition of the precursors is the pursued reaction pathway in the CVD process and the precursors are

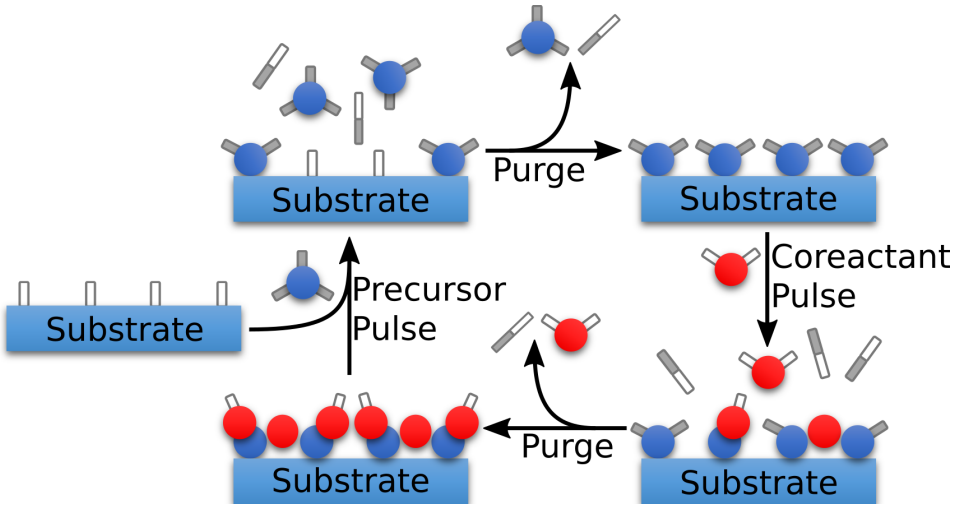


Figure 1.2.: Schematic illustration of the ALD process. The first and second half-cycle are shown containing the precursor and coreactant pulse and a subsequent purge step.

no longer self-limiting in their reactivity.

In Figure 1.3 the elemental reaction steps, which take place at the substrate surface, are shown for the CVD process. In contrast to the ALD process, adsorption, diffusion and decomposition are possible reaction steps for the precursors as well as for every obtained gas phase decomposition product. Consequently, the CVD process is less controlled and more complex than the ALD process. Also, the diffusion or mobility of all species is highly important due to the possible accumulation or clustering of a single species[49].

An advantage of the CVD process is that the higher reactivity results in higher growth rates. Furthermore, the CVD process is less dependent on finding suitable precursors. Still, it is harder to achieve high conformality, precise film thickness and composition in comparison to the ALD process. Furthermore, more side products are formed in the thermal decomposition of the precursors and thereby the spurious incoorporation of elements like carbon or oxygen has to be suppressed. Overall, the CVD process can handle a wide range of materials but exercise less control over growth process in comparison to the ALD process.

Also, different variants of the CVD process exists. In case metal-organic precursor like triethylgallium (TEGa) or tert-butylphosphine (TBP) are used, the CVD process is called metalorganic chemical vapor deposition (MOCVD). In addition, in case a crystalline film is grown on top of a single crystal substrate as for example for the growth of GaAs on a Si wafer the process is termed metalorganic vapor phase epitaxy (MOVPE).

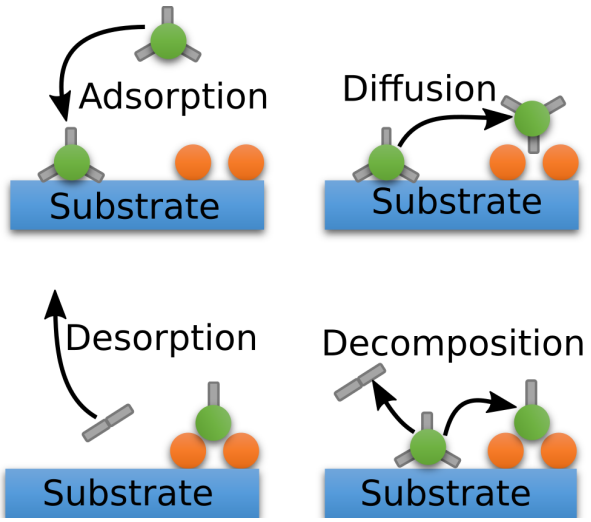


Figure 1.3.: Schematic illustration of the CVD process. The four most important elemental steps of the precursor adsorption, diffusion and decomposition and the desorption of fragments are shown.

Motivated by their relevance for the fabrication of electronic devices, the ALD and CVD processes are also modelled by various theories ranging from the reactivity of precursors at an atomic scale up to the fluid dynamics in the growth reactor[50–55].

Modelling both processes at the atomic scale is worthwhile to gain fundamental insights, which are not only used to understand but also to guide the exploration of new materials. Here, to fine-tune optoelectronic devices the nature and size of the band gap is for example of interest[56,57]. Also, structural information like possible surface reconstructions[58–61] under experimental conditions, the surface termination by hydrogen[62] or the structure of the formed interface[63] are obtained by theory. In addition, understanding and predicting the reactivity of precursors and coreactants is a valuable contribution by theory. Here, suitable precursors are predicted by studying their surface reactivity[64] or gas phase decomposition[65–70]. Initial adsorption reactions[71,72] and elemental reactions on the substrate surface[73–77] are investigated to understand the dependence of the precursor reactivity on the surface termination. Especially, modelling the elemental reactions is highly valuable since the knowledge gained thereby is used to model the material growth itself at a larger scale[78,79].

However, modelling these elemental reactions is a huge challenge since the ALD and CVD growth processes are controlled by a vast number of individual reactions. Here, a high manual effort is necessary to find all relevant reaction structures and paths. To guide the manual exploration of possible reactions, scientists commonly rely on their chemical intuition. On the one hand, chemical intuition based on knowledge and experience can point the way to important reaction channels. On the other hand, restricting the search of possible reactions to what is intuitive results in a bias, which increases the chance to miss unexpected but yet important reaction channels. Consequently, approaches to drastically reduce the manual effort and even preventing a bias in the exploration of possible reaction channels are desirable.

As discussed in detail in the next sections, this thesis is focusing on the reactivity of small molecules within the ALD and CVD process. In a first part a software for the automated exploration of reaction networks within the CVD process is presented. The software development is aiming at a high rate of automatization meaning that all elemental steps of the adsorption, diffusion and decomposition of the precursor of interest are derived with minimal human intervention. The capabilities of my software are presented by studying Bi, GaH_3 and PH_3 on the GaP(001) surface.

In a second part of this thesis, three novel small molecule inhibitor (SMI) for the area selective atomic layer deposition (AS-ALD) of Al_2O_3 and HfO_2 are introduced. The SMIs are studied

regarding their surface reactivity on a silicon dioxide and copper substrate. Consequently, the observed surface reactivity enables to propose suitable combinations of a SMI with a substrate to maximize the blocking behavior and thereby area selectivity.

2

Theory

This chapter provides a basic overview of the used theories and methods. While established text-book knowledge[80–85] is not explicitly cited in this chapter, selected publications will be referenced where appropriate. In a first section, the quantum mechanical description of chemical systems within the density functional theory (DFT) is introduced. Also the consideration of dispersion interactions and the extension to periodic systems are addressed in this section. As a second part of this chapter, the derivation of intramolecular forces and their utilization to derive reaction paths are presented.

2.1. Density Functional Theory

Within DFT the electron density ρ is used as the decisive quantity to determine every property of the studied system. While this theory is nowadays, especially in the material sciences, the most heavily used approach to model the electronic structure of the studied system based on the theory of quantum mechanics, historically approaches based on the wave function Ψ of the system were introduced first. Here, the wave function comprises every property of the system and is determined by the famous Schrödinger equation. In equation 2.1 the time independent and nonrelativistic formulation is shown. The Hamilton operator $\hat{H}$, which is based on the kinetic and potential energy of nuclei and electrons, is acting on the wave function leading to the total energy of the system as the eigenvalue E times the wave function as the eigenfunction. Here, commonly the Born-Oppenheimer approximation is used to decouple the motion of nuclei and electrons by assuming that the electrons move in a field of static nuclei. This assumption is based on the significantly smaller mass and thereby larger speed of motion of the electrons in comparison to the nuclei. As a consequence, an electronic Schrödinger equation is obtained confining the quantum mechanical description to the electrons while the kinetic energy of the nuclei is set to zero and the potential energy of the nuclei is incorporated as a constant.

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

Finding solutions to the wave function with a selected Hamilton operator and thereby obtaining the energy and all other properties of the system is the objective of several approaches. In the Hartree-Fock approach a single Slater determinant (SD) (equation 2.4) is used to approximate the wave function and to solve the electronic Schrödinger equation. As a consequence of using a single SD, the electron-electron interaction is described by an average interaction, i.e. every electron interacts with the mean-field of all other electrons. However, within this mean-field approximation the electron correlation is not completely captured. Consequently, post-Hartree-Fock methods are formulated to improve on the Hartree-Fock approach. In the configuration interaction and coupled cluster methods several SDs are used to describe the wave function and thereby capture the electron correlation. Another method is the Møller-Plesset perturbation theory. Here, a sum over the one-electron Fock operators of the Hartree-Fock approach is used as a reference Hamilton operator. The electron correlation is then captured by using the difference between the exact and reference Hamilton operator as a perturbation. Independent from the applied approach, a link between the wave function and the electron density is obtained as shown in equation 2.2 by integrating the wave function over all coordinates **x** of the N electrons except one set of spatial coordinates.

$$\rho(\mathbf{r}) = N \int \cdots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)|^2 \mathrm{d}s_1 \mathrm{d}\mathbf{x}_2 ... \mathrm{d}\mathbf{x}_N \tag{2.2}$$

2.1.1. Kohn-Sham Approach

The theoretical foundation to use the electron density instead of a wave function was provided by the Hohnberg-Kohn theorems. The first theorem states that the ground state of a system can be described entirely by its electron density ρ_0 by showing that every ground state electron density is uniquely assigned to an external potential and the wave function of the system. The second theorem demonstrates that the ground state electron density can be obtained by minimizing the total energy equivalent to the variational principle for wave functions. As a consequence of these theorems, the total energy can be formulated as a functional of the electron density ρ as shown in equation 2.3. However, only for the classical Coulomb J and the electron-nucleus interaction E_{Ne} a formulation is known while it is unknown for the kinetic energy T and the non-classical electron-electron interaction E_{ncl}.

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{ncl}[\rho(\mathbf{r})] + E_{Ne}[\rho(\mathbf{r})] \tag{2.3}$$

This is particularly problematic since the kinetic energy is a crucial part to the total energy. Walter Kohn and Lu Jeu Sham circumvented this problem by introducing a reference system of

non-interacting electrons moving in an external potential. The advantage of this approach is that this reference system is correctly described by a single SD Θ (equation 2.4). Here, the SD is built of one-electron spin orbitals χ the so called Kohn-Sham orbitals. Furthermore, within this reference system the kinetic energy of the electrons is known.

$$\Theta = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.4}$$

The connection between the reference system and the real system is introduced by requesting that the electron density has to be identical for both systems. In practice this is done by constructing an appropriate effective potential. The correct form of the potential is obtained by minimizing the expression for the total energy shown in equation 2.5. Here, the kinetic energy expression of the reference system, the classical Coulomb energy and the electron-nucleus attraction is formulated in terms of the Kohn-Sham orbitals. All unknown parts are collected in the exchange-correlation energy E_{XC}. This energy term contains the residual part of the true kinetic energy, the self-interaction correction, exchange interaction and correlation of the electrons.

$$\begin{aligned} E[\rho(\mathbf{r})] = &-\frac{1}{2}\sum_{i}^{N} \left\langle \chi_i \left| \nabla^2 \right| \chi_i \right\rangle + \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N} \int\int |\chi_i(\mathbf{r}_1)|^2 \frac{1}{r_{12}} |\chi_j(\mathbf{r}_2)|^2 \mathsf{d}\mathbf{r}_1 \mathsf{d}\mathbf{r}_2 \\ &+ E_{XC}[\rho(\mathbf{r})] - \sum_{i}^{N} \int \sum_{A}^{M} \frac{Z_A}{r_{1A}} |\chi_i(\mathbf{r}_1)|^2 \mathsf{d}\mathbf{r}_1 \end{aligned} \tag{2.5}$$

The minimization of the total energy under the constraint of orthogonal orbitals leads to the expression of the effective potential and enables the formulation of the Hamilton operator of the reference system $\hat{f}^{KS}$ shown in equation 2.6. Consequently, the Kohn-Sham orbitals can be obtained as solutions to the so called Kohn-Sham equations (equation 2.7).

$$\hat{f}^{KS} = -\frac{1}{2}\nabla^2 + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} \mathsf{d}\mathbf{r}_2 + V_{\mathsf{XC}}(\mathbf{r}_1) - \sum_{A}^{M} \frac{Z_A}{r_{1A}} \tag{2.6}$$

$$\hat{f}^{KS}\chi_i = \varepsilon_i \chi_i \tag{2.7}$$

However, an iterative procedure the so called self-consistent field (SCF) method is necessary to solve the Kohn-Sham equations since the electron density, which is constructed based on the

orbitals, is already incorporated in the formulation of the Hamilton operator. Furthermore, the exact form of the exchange-correlation potential is unknown. Hence, approximations to this functional are needed. Most prominent are functionals belonging to the group of local density approximations (LDAs), generalized gradient approximations (GGAs) or hybrid functionals. Here, the exchange and correlation contributions are commonly approximated separately.

2.1.2. Linear Combination of Atomic Orbitals Approach

With equation 2.7 a blueprint to apply DFT is obtained. However, this equation is hard to solve since the form of the one-electron wave functions χ_i is unknown. As a solution to this obstacle the wave functions are described based on a linear combination of atomic orbitals (LCAO) (equation 2.8). Since the atomic orbitals (AOs) are known functions, only the search for optimal coefficients $c_{\alpha i}$ remains.

$$\chi_i(\mathbf{r}) = \sum_{\alpha} c_{\alpha i}\phi_{\alpha}(\mathbf{r}) \tag{2.8}$$

Inserting this approach into equation 2.7 results in the Roothaan-Hall equations shown in equation 2.9. The search for finding the optimal coefficients has now turned into a matrix problem. Here, $\mathbb{F}$ (equation 2.10) is the Kohn-Sham matrix, $\mathbb{C}$ the coefficient matrix and $\mathbb{S}$ (equation 2.11) the overlap matrix of the AOs. By choosing an orthonormal set of AOs or by performing a respective basis set transformation, it is furthermore possible to transform this generalized eigenvalue problem to a standard eigenvalue problem. In this case the desired coefficients are obtained by diagonalizing the transformed Kohn-Sham matrix.

$$\mathbb{F}\mathbb{C} = \mathbb{S}\mathbb{C}\varepsilon \tag{2.9}$$

$$\mathbb{F}_{\alpha\beta} = \langle\phi_{\alpha}|\hat{f}^{KS}|\phi_{\beta}\rangle \tag{2.10}$$

$$\mathbb{S}_{\alpha\beta} = \langle\phi_{\alpha}|\phi_{\beta}\rangle \tag{2.11}$$

2.1.3. Periodic Boundary Conditions

To simplify the theoretical description of solids and surfaces, periodic boundary conditions (PBC) are utilized. In this approach the structure of the solid or surface is completely constructed based on a unit cell with the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The periodic structure is

then obtained by all possible translations **R** composed of multiples of the basis vectors (equation 2.12). Here, the translation vector **R** describes the translational symmetry of the solid or surface while the position of the atoms has to be defined only within the unit cell.

$$\mathbf{R} = t_1\mathbf{a}_1 + t_2\mathbf{a}_2 + t_3\mathbf{a}_3 \qquad t_i \in \mathbb{Z} \tag{2.12}$$

To describe the electronic structure of a periodic system the reciprocal space is used. Similar to the unit cell in real space, the reciprocal space is constructed by three basis vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ with a corresponding translation vector **K** (equation 2.13).

$$\mathbf{K} = s_1\mathbf{b}_1 + s_2\mathbf{b}_2 + s_3\mathbf{b}_3 \qquad s_i \in \mathbb{Z} \tag{2.13}$$

The basis vectors of the real and reciprocal space are orthogonal to each other, i.e. $\mathbf{a}_i\mathbf{b}_j = 2\pi\delta_{ij}$, and connected by the relation shown in equation 2.14. Here, V is the volume of the unit cell.

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{V} \qquad \mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{V} \qquad \mathbf{b}_3 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{V} \tag{2.14}$$

Following the Bloch theorem, a (one-electron) wave function $\chi_{\mathbf{k}}(\mathbf{r})$ of a periodic system can be obtained as a product of a plane wave with a cell periodic function $u_{\mathbf{k}}(\mathbf{r})$ (equation 2.15) also called Bloch factor.

$$\chi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \tag{2.15}$$

Here, the cell periodic function has to be invariant to translation by every translation vector **R**. As a consequence of the Bloch theorem the previously introduced Roothaan-Hall equations receive a **k**-dependence (equation 2.16). However, only for values of **k** within the first Brillouin zone unique wave functions are obtained. In addition, sampling the first Brillouin zone by some k-points is sufficient to converge the wave function and thereby also the electron density.

$$\mathbb{F}_{\mathbf{k}}\mathbb{C}_{\mathbf{k}} = \mathbb{S}_{\mathbf{k}}\mathbb{C}_{\mathbf{k}}\varepsilon_{\mathbf{k}} \tag{2.16}$$

2.1.4. Projector Augmented-Wave Method

In principle every type of basis function can be used to expand the cell periodic function. Nevertheless, periodic functions like plane waves, which fulfill the requirement of being invariant to translations by any translation vector **R**, are a well suited basis set. However, due to the rapid oscillations of the wave function in core regions, plane waves with a large kinetic energy are necessary. As shown in equation 2.17, the kinetic energy of a plane wave at the k-point **k** is

defined by the reduced Planck constant $\hbar$, the electron mass m_e and the wave vector **G**. Also, the size of the plane wave basis set is defined by a cutoff energy E_{cut}. Consequently, all plane waves with a kinetic energy less than the cutoff energy are included in the basis set. Thereby, large basis sets are needed to include plane waves with a large kinetic energy. However, such basis sets are computationally too demanding for most systems. A possible solution to this problem is the projector augmented-wave (PAW) method[86,87], which is a generalization of the linear augmented-plane-wave method and the pseudopotential method.

$$E_{\text{cut}} > E_{\text{kin}} = \frac{\hbar^2}{2m_e}|\mathbf{k}+\mathbf{G}|^2 \tag{2.17}$$

The strategy of the PAW method is to replace the true one-electron Kohn-Sham wave functions $|\chi\rangle$ by computationally convenient pseudo wave functions $|\tilde{\chi}\rangle$, which can be described by a small plane wave basis set. The connection between the true and the pseudo wave functions is a linear transformation $\mathcal{T}$ as shown in equation 2.18.

$$|\chi\rangle = \mathcal{T}\,|\tilde{\chi}\rangle \tag{2.18}$$

This approach has the advantage that once the transformation is known every operator $\hat{A}$ can be transformed to the respective pseudo operator $\tilde{A}$ (equation 2.19) and its expectation value can be obtained based on the pseudo wave functions.

$$\tilde{A} = \mathcal{T}^{\dagger}\hat{A}\mathcal{T} \tag{2.19}$$

To define the transformation, Blöchl introduced an augmented region around every core. Outside this augmented region the true and the pseudo wave function are identical. This implies that one part of the transformation $\mathcal{T}$ is a unit operator. Inside the augmented region the true and the pseudo wave function differ. Therefore, as a second part, atom-centered contributions to the transformation are necessary for every augmented region. Furthermore, within the augmented region the transformation is defined to map so called partial pseudo waves $|\tilde{\phi}_i\rangle$ on partial waves $|\phi_i\rangle$.

The transformation is now derived by expanding the (pseudo) wave functions as a linear combination of (pseudo) partial waves within the augmented region. The true wave functions can then be obtained by subtracting the pseudo partial wave from the pseudo wave function and

adding the partial waves as shown in equation 2.20.

$$|\chi\rangle = |\tilde{\chi}\rangle - \sum_i |\tilde{\phi}_i\rangle\, c_i + \sum_i |\phi_i\rangle\, c_i \tag{2.20}$$

Here, only the knowledge of the coefficients c_i is missing to derive the complete transformation. To maintain a linear transformation the coefficient are obtained by a scalar product shown in equation 2.21. The functions $\tilde{p}_i$ are introduced as projector functions fulfilling the relation $\langle \tilde{p}_i | \tilde{\phi}_j \rangle = \delta_{ij}$.

$$c_i = \langle \tilde{p}_i | \tilde{\chi} \rangle \tag{2.21}$$

Finally the transformation can be defined as equation 2.22.

$$\mathcal{T} = 1 + \sum_i \left[|\phi_i\rangle - |\tilde{\phi}_i\rangle \right] \langle \tilde{p}_i | \tag{2.22}$$

The advantage of the PAW method is that the necessary functions, i.e. the (pseudo) partial waves and the projector functions, can be defined for isolated atoms and kept constant during a calculation. In this case, only the plane wave basis set defining the pseudo wave functions has to be optimized. As a possible choice for the partial waves, Blöchl used solutions of the radial Schrödinger equation. In case a relativistic Schrödinger equation is used, also the mass-velocity and Darwin corrections can be included. However, the description of the spin-orbit coupling would require a perturbation operator[88] to be added. Like the partial waves also the pseudo partial waves are obtained as solutions to a radial Schrödinger equation. Here, a pseudo potential is used to smooth the obtained pseudo partial waves. The projector functions can then be derived based on the pseudo partial waves.

2.1.5. Atom Centered Basis Functions

In the previous section plane waves were introduced as a natural choice for basis functions. However, they are not the only possible choice. In molecular systems Slater type orbitals (STOs) and Gaussian type orbitals (GTOs) are commonly used. They are shown in polar coordinates in equations 2.23 and 2.24, respectively. Here, N is a normalization constant, Y are the spherical harmonic functions and n, l and m are the principal, angular momentum and magnetic quantum number. In addition, the size of the orbital is controlled by ζ. In contrast

to plane waves, STOs and GTOs are commonly centered at the position of the nuclei.

$$\phi^{STO}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}e^{-\zeta r} \tag{2.23}$$

$$\phi^{GTO}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}e^{-\zeta r^2} \tag{2.24}$$

The advantage of STOs is that they resemble the eigenfunctions of the hydrogen atom and consequently offer a high accuracy for the description of molecular orbitals. However, the calculation of four-center integrals is computationally demanding for STOs since these integrals can not be solved analytically. In contrast, GTOs are less accurate than STOs and as a consequence larger basis sets are necessary. However, since the four-center integrals can be solved analytically, GTOs are computationally significantly less demanding than STOs. Also, by combining several GTOs to one contracted GTO, which resembles the form of a STO, the accuracy can be increased.

GTOs and STOs can be used in equation 2.8 to build molecular orbitals. In addition, similar to a plane wave basis set, crystal orbitals of equation 2.15 can be derived based on atom centered basis functions as shown in equation 2.25. Here, a summation over all translation vectors **R** is necessary to fulfill the Bloch theorem.

$$\chi_{i,\mathbf{k}}(\mathbf{r}) = \sum_{\alpha}\sum_{\mathbf{R}} c_{\alpha i}e^{i\mathbf{k}\mathbf{R}}\phi^{GTO}_{\alpha}(\mathbf{r}+\mathbf{R}) \tag{2.25}$$

As already introduced in the LCAO approach, only the coefficients $c_{\alpha i}$ are optimized while the size of the GTO or STO controlled by ζ is fixed. Also, several basis functions with different values of ζ are combined with diffuse and polarization functions for the description of a single electron.

2.1.6. Dispersion Correction

A common problem of density functionals is that they are not able to account for London dispersion forces[89]. Therefore, the semiempirical dispersion correction approach DFT-D3[90,91] with improved damping function is used. The final energy of the DFT-D3 method is derived as a sum of the Kohn-Sham energy and the contribution from the dispersion correction (equation 2.26).

$$E^{D3}_{DFT} = E_{DFT} + E^{D3}_{Disp} \tag{2.26}$$

As shown in equation 2.27, the dispersion correction is calculated based on atom-pair wise dispersion coefficients C_{AB} weighted by their distance R_{AB} and a damping function f_{damp}.

Furthermore, two fitting parameters s_6 and s_8 are necessary. While s_6 turns out to be 1 for GGA and hybrid functionals, s_8 is used to adapt the dispersion correction to the character of the used exchange correlation functional.

$$E_{Disp}^{D3} = -\frac{1}{2}\sum_{A\neq B}\sum_{n=6,8} s_n \frac{C_{n,AB}}{R_{AB}^n} f_{damp}(R_{AB}) \tag{2.27}$$

$$f_{damp}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 R_{0,AB} + a_2)^n} \tag{2.28}$$

A common choice for the damping function $f_{damp}(R_{AB})$ is the function proposed by Becke and Johnson[92] (equation 2.28). Here, a cutoff radius $R_{0,AB}$ and two adjustable fitting parameters a_1 and a_2 are included. The cutoff radius defines in which area the dispersion energy is decreasing since the interaction is well described by the density functional. The parameter a_1 serves as a scaling factor of the cutoff radius, while a_2 is used to fine tune the function. By combining equations 2.27 and 2.28 the final expression for the dispersion correction is obtained as shown in equation 2.29.

$$E_{Disp}^{D3} = -\frac{1}{2}\sum_{A\neq B} s_6 \frac{C_{6,AB}}{R_{AB}^6 + (a_1 R_{0,AB} + a_2)^6} + s_8 \frac{C_{8,AB}}{R_{AB}^8 + (a_1 R_{0,AB} + a_2)^8} \tag{2.29}$$

The cutoff radius $R_{0,AB}$ in these equations is derived from the C_6 and C_8 dispersion coefficients as shown in equation 2.30.

$$R_{0,AB} = \sqrt{\frac{C_{8,AB}}{C_{6,AB}}} \tag{2.30}$$

The most important component of the DFT-D3 method are the $C_{6,AB}$ dispersion coefficients. Their derivation is based on the Casimir-Polder formula shown in equation 2.31. Here, the coefficients are calculated using the averaged dipole polarizability $\alpha(i\omega)$ of isolated atoms at the imaginary frequency ω. However, in the DFT-D3 method the dipole polarizabilities of simple hydrides for every element are used to derive the dispersion coefficients. This allows the introduction of a coordination number dependence to the dispersion coefficients. Higher order $C_{8,AB}$ coefficients are then recursively derived from the $C_{6,AB}$ coefficients and multipole-type expectation values.

$$C_{6,AB} = \frac{3}{\pi}\int_0^\infty \alpha_A(i\omega)\alpha_B(i\omega)\mathrm{d}\omega \tag{2.31}$$

2.1.7. Gibbs Free Energy

To add temperature and pressure effects to the DFT energies E_{el}, thermodynamic corrections can be applied[93]. These corrections are based on statistical thermodynamics to derive a correction to the enthalpy H_{corr} and the entropy S. With these terms the Gibbs free energy G at the temperature T is derived as shown in equation 2.32.

$$G = E_{el} + H_{corr} - T \cdot S \tag{2.32}$$

The correction to the enthalpy and the entropy term are calculated as a sum over their individual vibrational, rotational and translational contributions. Here, electronic contributions can be neglected unless the electronic states are close in energy. To obtain all contributions the well-known models of the harmonic oscillator, rigid rotator and ideal gas are applied. In this way the final equations 2.33 and 2.34 are obtained for the enthalpy and entropy respectively.

$$\begin{aligned} H_{vib} &= R\sum_{i}^{3M}\left[\frac{h\nu_i}{k_B}\left(\frac{1}{2}+\frac{1}{e^{h\nu_i/k_BT-1}}\right)\right] \\ H_{rot} &= \frac{3}{2}RT \\ H_{trans} &= \frac{5}{2}RT \end{aligned} \tag{2.33}$$

The equations for the enthalpy correction depend only on the temperature T and the vibrational frequencies ν_i. For the entropy contributions, the symmetry number σ, the moments of inertia I, the volume the molecule occupies V and its mass m are needed in addition. With these equations the Gibbs free energy can be approximated based on the DFT energy at the additional cost of calculating the frequencies (see section 2.2.2). However, this additional calculation can be computationally expensive in case the frequencies have to be derived numerically.

$$\begin{aligned} S_{vib} &= R\sum_{i}^{3M}\left[\frac{h\nu_i}{k_BT}\frac{1}{e^{h\nu_i/k_BT-1}} - \ln\left(1-e^{-h\nu_i/k_BT}\right)\right] \\ S_{rot} &= R\left[\frac{3}{2}+\ln\left(\frac{\sqrt{\pi}}{\sigma}\left[\frac{8\pi^2k_BT}{h^2}\right]^{\frac{3}{2}}\sqrt{I_1I_2I_3}\right)\right] \\ S_{trans} &= R\left[\frac{5}{2}+\ln\left(V\left[\frac{2\pi mk_BT}{h^2}\right]^{\frac{3}{2}}\right)\right] \end{aligned} \tag{2.34}$$

2.2. Potential Energy Surface

In the previous sections the calculation of the total energy for a fixed arrangement of atoms was introduced. The dependence of the total energy with respect to the atomic coordinates $E(\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_M)$ defines the potential energy surface (PES). In the exploration of a PES minima and minimum energy paths (MEPs) are especially of interest. A minimum corresponds to a stable structure while a MEP corresponds to a reaction path linking two structures. Along every MEP a transition state (TS) is observed. The TS is the point, which is a maximum in energy along the MEP but a minimum in all other directions. In addition to important structures, also energetic information like reaction energies and activation barriers are derived as energy differences between two points of the PES. To explore the PES, the forces $\mathbf{F}_A$ acting on each atom A are calculated. Forces can serve as guidance since they are always pointing in the direction of the closest minimum. As shown by equation 2.35 they are obtained as the first derivative of the energy E with respect to the coordinates $\mathbf{R}_A$ of atom A.

$$\mathbf{F}_A = -\frac{\partial E}{\partial \mathbf{R}_A} \tag{2.35}$$

2.2.1. Hellmann-Feynman Forces

In equation 2.36 the Hellmann-Feynmann theorem is shown. It states that for the calculation of the first derivative of the energy only the derivative of the Hamilton operator is needed.

$$\frac{\partial E}{\partial \mathbf{R}} = \langle \Psi | \frac{\partial \hat{H}}{\partial \mathbf{R}} | \Psi \rangle \tag{2.36}$$

In an ideal case this means that only the nucleus-nucleus repulsion and Coulomb interaction of the electron density with the nucleus contributes to the force. However, if for example an incomplete basis set or pseudo potentials are used additional contributions to the force arise[87].

2.2.2. Vibrational Frequencies

In addition to the calculation of the forces acting on every atom, also the second derivative of the energy with respect to the atomic coordinates is of interest. Here, the Hessian matrix $\mathbb{H}$ shown in equation 2.37 is obtained. Every element of the matrix corresponds to a partial derivative of the energy E with respect to two coordinates x. However, these elements are

for periodic systems commonly calculated numerically by using the method of finite differences (see section A.2). In case the central difference scheme is applied, two calculations are necessary for every degree of freedom of the system. Consequently, the calculation of the Hessian matrix is computationally demanding for larger systems.

$$\mathbb{H} = \begin{pmatrix} \frac{\partial^2 E}{\partial x_1 \partial x_1} & \frac{\partial^2 E}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_1 \partial x_{3M}} \\ \frac{\partial^2 E}{\partial x_2 \partial x_1} & \frac{\partial^2 E}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_2 \partial x_{3M}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x_{3M} \partial x_1} & \frac{\partial^2 E}{\partial x_{3M} \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_{3M} \partial x_{3M}} \end{pmatrix} \tag{2.37}$$

Still, once the Hessian matrix is calculated the vibrational frequencies can be derived. For this the elements of the Hessian matrix are weighted by the mass m_A and m_B of the displaced atoms. Here, the Hessian matrix is multiplied element-wise with the matrix $\mathbb{G}$, in which every element is defined according to equation 2.38.

$$G_{AB} = \frac{1}{\sqrt{m_A m_B}} \tag{2.38}$$

The eigenvectors of the mass-weighted Hessian matrix are the vibrational normal coordinates, which are used to describe the normal modes of the system. The vibrational frequencies ν_i are derived based on the eigenvalues ε_i of the mass-weighted Hessian matrix as shown in equation 2.39.

$$\nu_i = \frac{1}{2\pi}\sqrt{\varepsilon_i} \tag{2.39}$$

The benefits of deriving the vibrational frequencies are that thermodynamic corrections (see section 2.1.7) can be calculated and that the presence of a minimum structure (see section 2.2.3) can be confirmed. Here, a minimum structure is present if only real frequencies are observed. Otherwise, a first-order TS structure is present in case a single imaginary frequency is observed.

2.2.3. Structure Optimization

With the help of the forces, a minimum on the PES can be obtained by minimizing the energy of an initial structure $\mathbb{R}_0$ with respect to the atomic coordinates. Here, $\mathbb{R}$ corresponds to a matrix storing the coordinates $\mathbf{R}_A$ of every atom. A corresponding matrix $\mathbb{F}$ storing the forces $\mathbf{F}_A$ acting on every atom can also be defined. In the minimization shown in equation 2.40, a new structure $\mathbb{R}_{i+1}$ is obtained by performing a step λ_i in a certain direction $\mathbb{D}_i$ with respect to the current structure $\mathbb{R}_i$. This process is repeated until a minimum on the PES is reached.

The step direction within this process is defined by the used minimization algorithm.

$$\mathbb{R}_{i+1} = \mathbb{R}_i + \lambda_i \mathbb{D}_i \tag{2.40}$$

For all minimizations within this work the conjugate gradient (CG) algorithm[94] is used, which is schematically shown in Figure 2.1. Consequently, as shown in equation 2.41 a new search direction is obtained by two steps. In a first step, the force history $\mathbb{H}_i$ is updated. For this a small portion γ of $\mathbb{H}_{i-1}$, which contains the previously observed forces and thereby the former optimization direction, is added to the actual force $\mathbb{F}_i$. In a second step, the dimensionless direction $\mathbb{D}_i$ is obtained as a normalization of $\mathbb{H}_i$.

$$\begin{aligned} \mathbb{H}_i &= \mathbb{F}_i + \gamma \mathbb{H}_{i-1} \\ \mathbb{D}_i &= \frac{\mathbb{H}_i}{\|\mathbb{H}_i\|} \end{aligned} \tag{2.41}$$

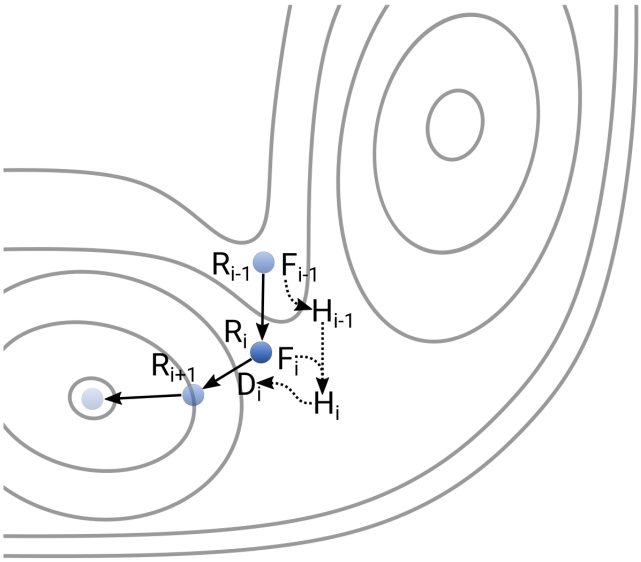


Figure 2.1.: Illustration of the conjugate gradient algorithm. A blue point represents structure i on the PES (grey lines) with the coordinates $\mathbb{R}_i$ and force $\mathbb{F}_i$. The direction $\mathbb{D}_i$ of the next optimization step is derived based on the history $\mathbb{H}_i$ of all previously obtained forces. The trial step in the direction $\mathbb{D}_i$ to derive the actual step length is not shown.

Several ways to define γ are available[95]. Here, I followed the common definition of Polak and Ribiere[96] formulated using the Frobenius product as shown in equation 2.42.

$$\gamma = \frac{\langle \mathbb{F}_i - \mathbb{F}_{i-1}, \mathbb{F}_i \rangle_F}{\langle \mathbb{F}_{i-1}, \mathbb{F}_{i-1} \rangle_F} \tag{2.42}$$

Following reference[94] the step size λ is calculated based on Newton's method. For this a trial step with a fixed step size λ_T has to be performed in the current direction.

$$\mathbb{R}_{T,i} = \mathbb{R}_i + \lambda_T \mathbb{D}_i \tag{2.43}$$

By the trial step the structure $\mathbb{R}_{T,i}$ (equation 2.43) and force $\mathbb{F}_{T,i}$ is obtained. Based on the method of finite differences (for details see A.2) the magnitude of the force m_F and curvature κ along the direction of the trial step can be derived for an intermediate point as shown in equation 2.44 and 2.45, respectively.

$$m_F = \frac{1}{2} \langle \mathbb{F}_i + \mathbb{F}_{T,i}, \mathbb{D}_i \rangle_F \tag{2.44}$$

$$\kappa = \frac{1}{\lambda_T} \langle \mathbb{F}_i - \mathbb{F}_{T,i}, \mathbb{D}_i \rangle_F \tag{2.45}$$

Finally, with both values the step size λ can be derived as shown in equation 2.46.

$$\lambda_i = \frac{m_F}{\kappa} - \frac{1}{2}\lambda_T \tag{2.46}$$

Here, the addition of $-\frac{1}{2}\lambda_T$ is necessary to account for the correct step length respective to $\mathbb{R}_{T,i}$.

While the CG algorithm relies only on the structure and force information, more complex optimization techniques as the Newton-Raphson method also consider the second derivative of the energy with respect to the atomic coordinates. For this, the Hessian matrix $\mathbb{H}$, previously introduced in equation 2.37, is calculated. As shown in equation 2.47 and 2.48 the product of the inverse Hessian matrix and the gradient **g** is used to define the optimization direction **d** or the optimization step directly. Here, in contrast to the introduction of the CG algorithm, **g**, **d** and **R** correspond to vectors storing the $3M$ atomic coordinates and gradients with M the number of atoms. Since the calculation of the Hessian matrix in every step is computationally too demanding, even by the method of finite differences (section A.2), approaches to estimate the Hessian matrix based on gradient information are used. A common example for such a quasi-Newton approach is the algorithm developed by Broyden, Fletcher, Goldfarb

and Shanno (BFGS)[97] and extended by Nocedal (L-BFGS)[98]. A pleasant feature of a quasi-Newton method is that the approximated Hessian matrix is, in contrast to the true Hessian matrix, always positive defined. As a consequence, the optimization converges to a minimum on the PES instead of the closest stationary point, which could be a TS.

$$\mathbf{d}_i = -\mathbb{H}^{-1}\mathbf{g}_i \tag{2.47}$$

$$\mathbf{R}_{i+1} = \mathbf{R}_i - \mathbb{H}^{-1}\mathbf{g}_i \tag{2.48}$$

2.2.4. Nudged Elastic Band

Once two minima on the PES are found they can be linked by a MEP optimized by the nudged elastic band (NEB) method[94,99,100]. The MEP corresponds to a reaction path describing the change from the reactant to the product structure. Since both, the start and the end point of the MEP have to be known in advance, the NEB method belongs to the group of double ended methods. The actual MEP is obtained by optimizing N structures - so called images - in

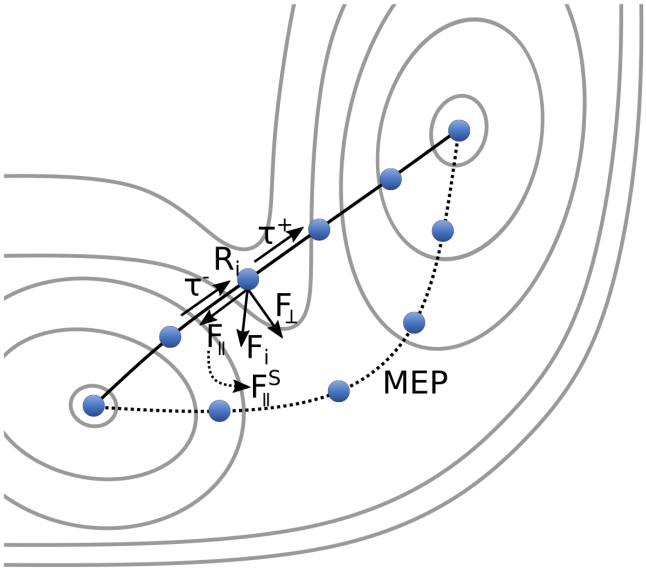


Figure 2.2.: Illustration of the nudged elastic band method. A blue point represents a structure i on the PES (grey lines) with the coordinates R_i and force F_i. As indicated for a single structure, the force F_i is projected based on a tangent τ_i in a component parallel $F_\parallel$ and perpendicular $F_\perp$ to the reaction path. In the optimization of the reaction path, the parallel component is replaced by a spring force $F_\parallel^S$. As a result, the minimum energy path (MEP) is obtained, which is evenly sampled by the structures.

between the minima to the MEP. To quickly converge the images to the MEP the NEB method follows the idea to split the true force in components parallel and perpendicular to the MEP as schematically shown if Figure 2.2. For this the force $\mathbb{F}_i$ at every image i is projected on the tangent τ_i at this image. Since the exact tangent of the MEP is unknown, an estimate is made based on the structures of the neighboring images.

Equation 2.49 shows the definition of the tangent for images, which are not extrema. Here, the tangent is estimated based on one neighboring image, which is higher in energy than the image i.

$$\tau_i = \begin{cases} \tau_i^+ & = \mathbb{R}_{i+1} - \mathbb{R}_i \quad \text{if } E_{i+1} > E_i > E_{i-1} \\ \tau_i^- & = \mathbb{R}_i - \mathbb{R}_{i-1} \quad \text{if } E_{i+1} < E_i < E_{i-1} \end{cases} \tag{2.49}$$

For extrema, both the previous and the next image are used. As shown in equation 2.50, the tangent to the previous and next image are weighted based on an energy difference ΔE_i. This definition ensures that the image, which is closer in energy to the maximum is always weighted by ΔE_i^{max} and therefore more strongly contributing to the tangent.

$$\tau_i = \begin{cases} \tau_i^+ \Delta E_i^{\text{max}} + \tau_i^- \Delta E_i^{\text{min}} & \text{if } E_{i+1} > E_{i-1} \\ \tau_i^+ \Delta E_i^{\text{min}} + \tau_i^- \Delta E_i^{\text{max}} & \text{if } E_{i+1} < E_{i-1} \end{cases} \tag{2.50}$$

The weighting factors are simply derived based on the difference in energy of image i to the previous and next image as shown in equation 2.51.

$$\begin{aligned} \Delta E_i^{\text{max}} &= \text{max}(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \\ \Delta E_i^{\text{min}} &= \text{min}(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \end{aligned} \tag{2.51}$$

Furthermore, following equation 2.52, the tangent is normalized before the force projection.

$$\hat{\tau}_i = \frac{\tau_i}{|\tau_i|} \tag{2.52}$$

With the normalized tangent, the parallel component of the force $\mathbb{F}_{i,\parallel}$ is obtained by a projection of the true force on the tangent (equation 2.53).

$$\mathbb{F}_{i,\parallel} = \frac{\mathbb{F}_i \cdot \hat{\tau}_i}{\underbrace{\hat{\tau}_i \cdot \hat{\tau}_i}_{=1}} \cdot \hat{\tau}_i = (\mathbb{F}_i \cdot \hat{\tau}) \cdot \hat{\tau} \tag{2.53}$$

Once the parallel component is known, the perpendicular component is obtained by subtract-

ing the parallel component from the true force (equation 2.54).

$$\mathbb{F}_{i,\perp} = \mathbb{F}_i - \mathbb{F}_{i,\parallel} \tag{2.54}$$

A downside of the tangent estimation is that the perpendicular component of the force is not perfectly perpendicular to the true MEP. Consequently, it contains also contributions parallel to the MEP, which will push the image steadily in the direction of the closest minimum. To avoid all images collapsing into a minimum, an additional force is introduced. This spring force $\mathbb{F}^S_{i,\parallel}$ aims at distributing the images equally along the MEP and is defined in equation 2.55. The direction of this force depends on the tangent and the structural differences to the previous and next image, while the magnitude of the force is controlled by the spring force constant k.

$$\mathbb{F}^S_{i,\parallel} = k \cdot (|\mathbb{R}_{i+1} - \mathbb{R}_i| - |\mathbb{R}_i - \mathbb{R}_{i-1}|) \cdot \hat{\tau} \tag{2.55}$$

Finally, the force of image i is defined as the sum of the perpendicular and the spring force (equation 2.56) in case the image is not an extrema. Otherwise the force is defined as in equation 2.57 in case the so called climbing image (CI) is activated. In the latter case the image is forced to move in the opposite direction of the parallel force, i.e. the image is moving in the direction of the TS.

$$\mathbb{F}^{\mathsf{NEB}}_i = \mathbb{F}_{i,\perp} + \mathbb{F}^S_{i,\parallel} \tag{2.56}$$

$$\mathbb{F}^{\mathsf{NEB}}_i = \mathbb{F}_i - 2\mathbb{F}_{i,\parallel} \tag{2.57}$$

3

PESE - A Potential Energy Surface Explorer

The motivation to model CVD and ALD processes is founded in the pursuit of understanding material growth at an atomistic level. Such a deep understanding enables to identify important elemental steps and to tune them seeking for optimal growth processes and ultimately improve devices. As a first step, the gas phase reactivity of the used precursors can be modelled. In an ideal ALD process the precursors reach the surface intact wherefore gas phase reactivity is usually not of interest. However, for CVD processes gas phase decomposition reactions of the precursors are of interest due to the elevated process temperatures. Consequently, gas phase decomposition of TEGa[65], TBP[65,67,68], tert-butylarsine (TBAs)[68], di-tertiary-butyl-arsano-amine (DTBAA)[69] or boron containing precursors[66,70] were studied for example.

As a result of these seemingly simple decomposition processes already complex reaction networks with a variety of elemental steps and consecutive reactions were observed. However, a complete and realistic model of growth processes must include the description of decomposition reactions on the substrate surface in addition to the gas phase reactivity. Unfortunately, additional layers of complexity are added to the computations by incorporating a surface: Before any decomposition reaction is studied, different adsorption sites must be identified to understand their influence on the decomposition reactions. Furthermore, a tedious side effect of using the adsorption structure as a starting point for decomposition reactions is that the reactions are often not invariant with respect to the precursor orientation. Only for high symmetry adsorption sites the orientation of the reactive group might be ignored. These two effects already result in considerably larger reaction networks. In addition, plenty of surfaces and reconstructions for every surface are relevant even for the growth of a single material as nicely shown with the III-V compound semiconductors[59].

For the study of molecular reactivity, a huge complexity of reaction networks stimulated the development of tools and software for the automated exploration of a reaction network.

These approaches have in common that they have to find a way to optimize intermediate structures, reaction paths and the corresponding TSs. Furthermore, every approach must define an exploration procedure, which defines the direction of the exploration. Here, plenty of solutions to these two tasks are possible as presented by extensive overviews in the literature[101–108].

The simplest exploration procedures are based on chemical transformation rules to guide the exploration. Furthermore, thermodynamic data are stored in a database and used to predict the stability of intermediates and reaction barriers. An example for this approach is the reaction mechanism generator (RMG)[109–111] developed by Green and coworkers. Here, all allowed reactions are stated as graph-based rules. In addition, thermodynamic and kinetic data are transferred to the system of interest by linear scaling relations[112,113]. Within this approach surfaces can be included as a node in the graph-based reaction rules[110]. Also, in an extension to the RMG, thermodynamic data of surface reactions were derived from quantum mechanical calculations[114–116]. Furthermore, an automated TS generation by using group contributions and the distance geometry approach[112,114] was added to improve the estimation of reaction barriers. Here, the final reaction path can be obtained by internal reaction coordinate (IRC)[117] calculations after the optimization of the TS guess.

Similar rule-based approaches where also developed by other groups[118–123]. Furthermore, Nørskov and coworkers proposed the idea to select interesting elemental steps for a refinement by DFT calculations based on machine learning (ML) approaches[124]. However, all approaches combining a rule-based exploration with a reaction energy and structure database have in common that they rely on their predefined data whereby the successful study of new systems depends on these data.

Instead of using predefined data to estimate reaction barriers another class of approaches aims at observing the reaction event directly. The group of Martínez-Núñez developed an approach based on molecular dynamic (MD) simulations[125–128]. Here, the MD simulations are performed with semiempirical methods. Furthermore, all vibrational modes are populated to increase the rate at which TSs and reaction events are observed. The TS region is then identified by the observation of connectivity changes during a MD simulation. Several structures within the TS region are used in individual attempts to preoptimize the TS before DFT is used with a TS optimizer in a final optimization. Similar MD simulation based approaches used a high temperature or pressure[129–131] to accelerate the observation of reaction events. Another related approach is the exploration of the PES based on metadynamic calculations[132–137]. Here, an additional bias potential or force is recursively added to increase

the sampling speed of the MD calculation. The bias potential is controlled by so called collective variables, which are predefined by the user and for example based on bond lengths or angles. However, a reasonable definition of the collective variables is challenging. In case the definition of the collective variables is too strict only a few types of reactions can be observed. In contrast, a huge number of sampling steps are necessary for the exploration of reaction events in case the definition of the collective variables is too general.

A common downside of MD based approaches is that they have to rely on semiempirical methods to run a large number of MD trajectories. However, semiempirical methods are based on empirical parameters and offer a good accuracy only for systems reasonably represented in the datasets, which are used to fit the parameters. Consequently, computationally less demanding exploration approaches enabling the usage of DFT as the underlying theory are developed, too. As an example, Liu and coworkers developed the stochastic surface walking (SSW) method[138–140] as a combination of structural optimizations guided by a bias-potential and Metropolis Monte Carlo sampling. The idea is to start at a user defined minimum structure and follow random mode directions by the bias-potential driven constrained-Broyden-dimer (BP-CBD) method[141,142]. The acceptance of new intermediates is then based on a Metropolis Monte Carlo algorithm using DFT energies. In case the intermediate is accepted, TS calculations are performed with the double-ended surface walking method[143]. This approach was used to study the reactivity of water and CO on Cu(111)[144] or to sample crystal structures and phase transitions[145,146].

Another approach called ZStruct[147–149] is developed by the group of Zimmerman. This approach utilizes connectivity graphs to identify potential intermediates which are then optimized by semiempirical methods and refined by DFT. TS are subsequently located by the double-ended growing string method (GSM)[150–155]. More interestingly, with S-ZStruct[156] an extension to surfaces was implemented and used to study the decomposition of propanoic acid on Pd(111) and the ALD of TiN on Si(100). For the initial structure generation S-Zstruct relies on predefined definitions of binding sites as a user input. In the initial structure generation different orientations are considered and after the optimization screened by a root mean square deviation (RMSD) criterium. As for ZStruct a graph-based creation of possible reaction products is used while these information are encoded in the search direction of the single-ended GSM.

A similar approach based on connectivity graphs was also developed by Green and coworkers[103,157]. In this approach the freezing string method[158] is used to obtain a TS guess. Subsequently, the Berny TS finder[159,160] is used to optimize the TS and IRC[117] calculations to obtain

the reaction path. In contrast to ZStruct also a sampling of conformers is conducted for every intermediate. Further contributions following this approach originate from the groups of Habershon[161,162] and Kim[163] by combining graph-based explorations approaches with MD simulations or basin-hopping methods.

Instead of using the general connectivity of molecules for the exploration of reaction networks, the exploration can be restricted to interesting transformations by detecting the reactive sites as proposed by Reiher and coworkers[164]. Their approach uses chemical descriptors[164,165] to select possible reactive centers. By overlapping two reactive sites followed by structural optimizations new intermediates are observed. In a subsequent step the TS can be optimized with any available TS optimization algorithm.

These exploration approaches, which generate the intermediates before the TS optimization usually rely on double-ended TS optimizer. However, in the exploration of a reaction network complex reaction paths as indicated by the presence of multiple elemental steps are found. Here, double-ended methods are usually not well suited. Consequently, single ended TS optimizer can be more efficient. Furthermore, single ended methods can also be used in case only the starting point of the reaction path is known. Consequently, also exploration approaches based on single-ended TS finder are developed.

A notable example is the approach by Maeda and coworkers implemented in the global reaction route mapping (GRRM)[166,167] strategy. This strategy is based on two methods namely the anharmonic downward distortion following (ADDF)[168–170] and the artificial force induced reaction (AFIR)[101,171–175] method. The idea of the ADDF method is that an anharmonic curvature of the PES in comparison to an harmonic potential is observed by moving from a minimum to a transition state. The ADDF method follows this curvature to obtain a TS. While the ADDF method is well suited for unimolecular reactions, an extension to biomolecular reactions is obtained with the AFIR method. Here, an artificial force is added between molecules or selected atoms to observe a reaction during a structural optimization. The thereby obtained reaction path is also close to the true MEP and is used to refine the TS by for example the locally updated planes[176] method. Within the GRRM strategy a complex reaction network can be obtained based on an initial structure provided by the user. However, also the artificial forces must be defined in advance. This approach was applied to several surface systems as for example the decomposition of methanol on Pt(111)[177], H_2O decomposition on Cu(111)[178], O-atom diffusion on Si(001)[179] and CO oxidation on Pt(111)[180].

All presented approaches for the exploration of surface reaction networks have in common that they are often restricted to simple metal surface where high symmetry points are known.

Otherwise, as for Si(001) a user input defining the adsorption site is required as a starting point. Consequently, for a completely automated exploration of surface reaction networks a systematic approach to identify possible adsorption sites is needed. In addition, the identification of adsorption sites has to include an investigation of the adsorption process. Here, possible adsorption paths have to be calculated and the probability to adsorb at a certain adsorption site has to be estimated. Until now, these important steps for the description of surface reaction networks are neglected in all exploration approaches.

Furthermore, computational performance has to be a crucial aspect of every exploration algorithm focusing on surface reaction networks since accurate models containing a substrate surface are computationally significantly more demanding than gas phase calculations. Consequently, a high intrinsic parallelization is seeked for the exploration algorithm enabling the investigation of large molecules on complex surface reconstructions with a multitude of adsorption sites. In addition, it is desirable to entirely get rid of manual evaluation steps so that the exploration of a surface reaction network can be performed in a single continuous calculation.

Also, to improve the algorithms for the exploration of surface reaction networks, the most critical issue of double-ended methods namely dealing with multistep pathways[102] should be addressed. Solving this shortcoming would enable the usage of these otherwise stable and reliable methods for complex surface reaction paths.

Consequently, within this chapter my software PESE - a Potential Energy Surface Explorer for thin film growth (PESE)[181] is introduced in detail addressing these needs. In section 3.1 the programming and parallelization concept is introduced, while the individual routines are explained in section 3.2.

3.1. Concept of PESE

The basic idea of PESE is to mimic the steps a researcher would perform to study the chemical reactions within MOVPE. This means, that PESE is not a software development in the sense that a completely new algorithm to explore the PES is derived. Instead PESE relies on established approaches to find and track structures on the PES and replaces the manual evaluations and processing steps. Furthermore, PESE is also taking care of organizing and starting all calculations by accessing a common quantum chemistry software. This is necessary to reach a high degree of automatization and to reduce human intervention to a minimum. Thereby,

the exploration of a reaction network in a single calculation on high performance computing resources is enabled.

Types of Calculations

For the calculation of the introduced elemental steps (adsorption, diffusion, decomposition) in Figure 1.3 a common approach is to obtain important minima on the PES in a first step by optimizing (Section 2.2.3) several initial guesses. This includes adsorption and decomposition minima. In a second step these minima can be linked by calculating the corresponding reaction paths using the double-ended NEB method (Section 2.2.4). The NEB method is established as a fast and reliable method for periodic systems since the computation of the second derivative to the energy is not needed. In addition, the complete reaction path linking two selected minima is obtained with the corresponding TS. Furthermore, additional minima along a reaction path prior or after the main reaction event can be identified. However, as discussed in the introduction, the standard version of the NEB is not able to optimize several TSs or minima along a reaction path.

Based on in this common approach, the exact sequence of all calculations performed by PESE is shown in Figure 3.1a. In a first step, PESE is optimizing the isolated systems of the adsorbate and surface to obtain their total energies, which are necessary to derive adsorption and bonding energies for every minimum. In a second step, all adsorption minima are calculated by PESE. The calculation of decomposition minima is performed subsequently since PESE derives the initial guess for decomposition minima based on the adsorption minima. Once

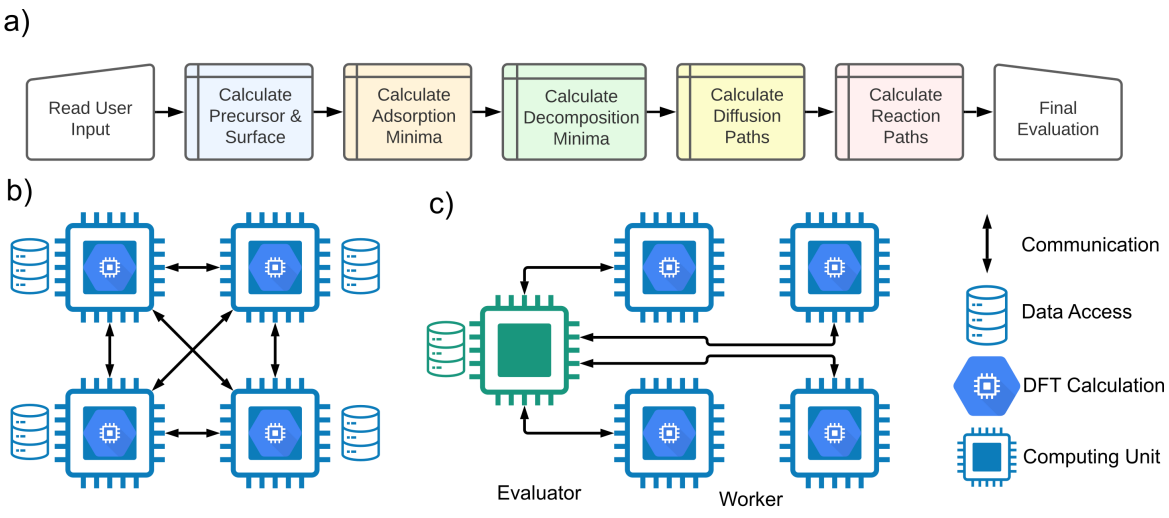


Figure 3.1.: Ordering of tasks to calculate the reaction network shown in (a). Possible parallelization layouts shown in (b) and (c) with a legend explaining the pictograms.

all computationally less demanding structure optimizations are completed, PESE is investing the computational resources in the calculation of reaction paths. Here, diffusion (adsorption) paths based on the adsorption minima are calculated at first while decomposition paths based on the adsorption and decomposition minima are calculated in a final part.

Parallel Programming

Before the implementation of the calculation sequence is introduced, a brief discussion of the concepts behind parallel programming is helpful. In general, a software is parallelized with the desire to reduce the necessary time to process a fixed task or solve a fixed problem. This is achieved by distributing the work on several computing cores. In this context, a core is the smallest computing unit and a building block of every CPU. Distributing the work on several cores comes with the additional cost of communication between the cores. This communication is for example necessary to share data or to synchronize the performed tasks. Here, the programmer is responsible to define the necessary communication by using the open multi-processing (OpenMP) or message passing interface (MPI) standard. Especially the MPI standard is important for PESE since it is used to define the communication between cores on separate nodes of a computer cluster.

For parallel programming it is important to understand that the amount of communication scales with the number of used cores. Therefore, it is not possible to increase the efficiency of a parallel program towards infinity by simply using more and more cores. To obtain the best efficiency and to reduce the communication the parallelization concept has to match the addressed problem.

Here, two general concepts should be discussed: The concept of data parallelism means that every core performs the same task but only on its own small chunk of data. This concept is used in every quantum chemistry software since huge matrices have to be handled. In this example, every core contributes a small part to these matrices. In this way a parallelization over k-points, bands or basis functions is achieved. The second concept is the task parallelism. Here, every core processes its own task with its own data. This concept is in particularly advantageous if several independent tasks can be defined, but horribly inefficient if tasks have to be processed successively. In practice, a parallel program is often a mixture of both concepts. Nonetheless, these concepts help to understand the implementation of PESE.

Parallelization Concept of PESE

For PESE an efficient parallelization is necessitated due to the high number of individual calculations in every step of Figure 3.1a. In principle, two different parallelization concepts would be possible: PESE could either perform one calculation at a time with as many computing resources as possible (data parallelism) or perform several calculations in parallel each with a fraction of the computing resources (task parallelism). Performing only a single calculation at once would suffer from the communication overhead as soon as a huge number of cores is utilized. In contrast, the concept of task parallelism is easily applied since the individual calculations within every step are independent. Here, the parallelization over calculations is even a so called embarrassingly parallel problem since no communication between cores working on different calculations is necessary at all. Consequently, for PESE the task parallelism concept was selected as the underlying programming model. In a rigorous implementation this would result in every core performing its own calculation. However, to be able to use more cores than calculations are to be performed and to finish any individual calculation in a reasonable time several cores are grouped to process a single calculation. Overall, this results in PESE following the task parallelism on a high level and the data parallelism on the low level.

For the development of PESE this concept had to be turned into an implementation strategy. Here, I could either follow a symmetric or asymmetric parallelization as shown in Figure 3.1. In a symmetric parallelization (Figure 3.1b) all groups of cores processing their individual calculation are equal. This requires again some communication overhead between these groups to synchronize the work balance, i.e. which group has to process which calculation. The advantage of this implementation would be that all cores are used for processing calculations. However, a proper implementation of the communication is difficult. The communication would need to take place in between two calculations since the individual calculation must not be interrupted. This would result in idle times for every group of cores finishing their calculation before the group processing the most demanding calculation.

The alternative implementation of an asymmetric parallelization is shown in Figure 3.1c. Here, a single group of cores in the following called *evaluator* is responsible to organize and schedule all calculations. All other groups of cores in the following called *workers* would process the calculations. In this implementation again communication overhead arises due to the synchronization of workers with the evaluator. However, the communication overhead for the workers is minimized. This is only valid in case the communication between the workers and evaluator is negligible so that a worker is never waiting to be able to communicate with the evaluator. The disadvantage of this implementation is that the evaluator is not processing any calculations. However, this disadvantage vanishes the more workers are used. Since

the computationally most expensive part for PESE is performing the DFT calculations, I want to minimize every communication and synchronization for the workers. Consequently, the *Evaluator-Worker* model is used for PESE.

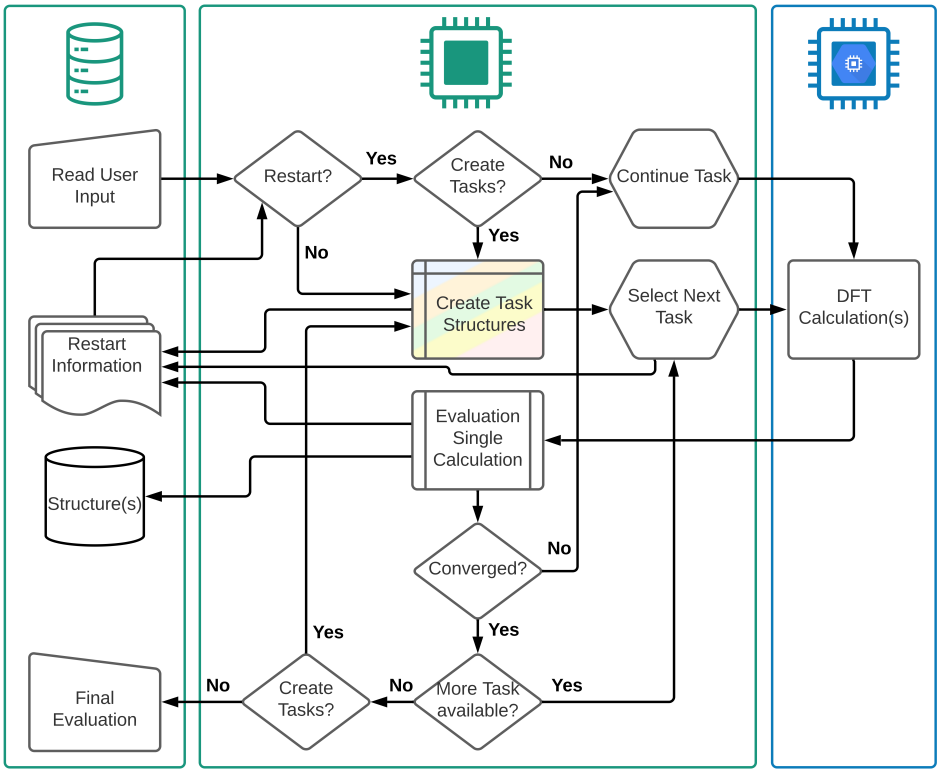


Figure 3.2.: Final routine layout of PESE. The necessary data management is shown on the left, the additional work performed by the evaluator in the middle and the work performed by the workers on the right. The connection to Figure 3.1a is the highlighted ”Create Task Structures” field, which is following the sequence of Figure 3.1a. Here, a task represents any DFT calculation.

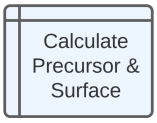
Continuation of Calculations

So far I discussed the sequence of calculations PESE is performing and the parallelization concept. Another important aspect is the possibility to restart a PESE calculation in case an error occurred or not all calculations were processed. Here, it is crucial to prepare for a restart in a sense that none or only a minor part of a calculation has to be redone. This is achieved by storing crucial information on the hard drive. For all workers the information which calculations are processed has to be stored. For this a unique label is added to every calculation and stored with the information of the worker. In case PESE is restarted every worker can detect based on the label which calculations it has to continue. Furthermore, for the evaluator the

information of every processed or evaluated calculation has to be stored. With this information the evaluator is, in case of a restart, able to detect all calculations, which still must be executed. Besides the information necessary to continue or schedule the next calculations, all results of already processed calculations have to be stored too. This is achieved by storing information of the atomic positions, forces and energies for every obtained structure.

By combining all discussed aspects of PESE, namely the types of calculation, the parallelization concept and the continuation of calculations the overview in Figure 3.2 is obtained.

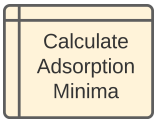
3.2. The Algorithm



The previous section introduced the parallelization concept and data management of PESE. In this section the individual routines to create and evaluate a calculation are introduced in detail. Here, we follow the sequence of Figure 3.1a. All routines introduced here are handled by the evaluator since the workers are, as discussed above, only responsible to perform the DFT calculation.

The first step of Figure 3.1a to optimize the structure of the studied precursor and surface is a trivial step since both structures are provided by the user. Therefore, despite creating a request to optimize these structures by a worker no additional action is necessary for the evaluator. The creation of initial guesses for the adsorption minima is the first non-trivial step for the evaluator.

3.2.1. Obtaining Adsorption Minima



The structure of adsorption minima depends on the structure and reactivity of the surface and precursor. Initial guesses for these structures could be obtained based on reactivity descriptors[164]. This would require initial DFT calculations and for plane wave codes often a projection to atom centered basis functions to derive descriptors as for example partial charges[182]. PESE is following a simpler approach by searching for adsorption minima in a brute force manner. The general idea is to divide the surface based on a regular grid, which is then used to position the studied precursor above the surface. In this way a large number of initial guesses is obtained, which have to be optimized by the workers. This approach does not require a detailed knowledge of the reactivity of the surface or precursor but is computationally more

expensive due to the large number of optimizations. Nevertheless, this approach is applicable to any surface without the need for any user input and does not overlook structures due to user bias.

Symmetry Detection

To minimize the number of grid points and thereby the number of initial guesses, the symmetry of the studied surface is exploited. For this PESE is detecting the translational and mirror symmetry of the surface. In the current implementation these symmetry elements can only be detected for a rectangular unit cell with the lattice vectors parallel to the cartesian coordinate axes. However, this approach is sufficient to handle the common III/V semiconductor surface reconstructions[59].

In general, a symmetry element is identified in case it can be successfully applied to every atom. To obtain a possible set of symmetry operations for the surface, PESE is calculating all possible translational and mirror symmetry elements for a single atom first. In Figure 3.3 the rules for the symmetry detection are sketched. Here, a translational symmetry in x-direction is the distance r_x between two atoms of the same element (Figure 3.3a,d). Furthermore, the displacement between these atoms in y- and z-direction $r_{y,z}$ has to be smaller than a preset threshold of 0.1 Å (Figure 3.3c). This threshold ensures that the symmetry is detected even if the atoms are not perfectly aligned. The value of the threshold was obtained by applying the approach to a surface of interest. Translational symmetries in y-direction are identified in the same fashion as the symmetry elements in x-direction.

Mirror planes perpendicular to the x-direction are identified in the center of a translation in

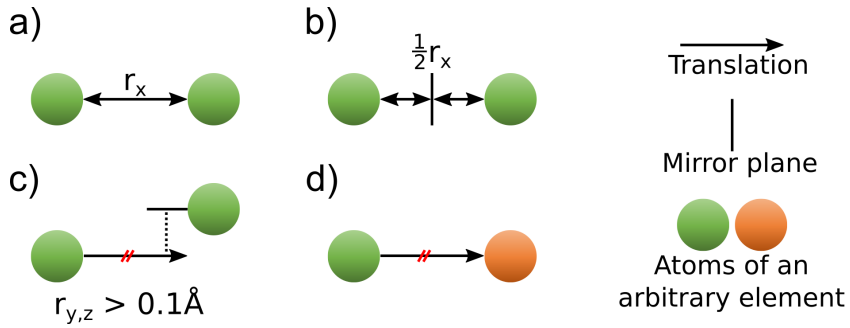


Figure 3.3.: Rules for the symmetry detection of the studied surface. Detection of valid symmetry operations shown in (a) for translation and (b) for mirroring. Rejection of a symmetry operation in x-direction in case the displacement of the atoms in y- or z-direction is too large (c) or the symmetry operation would link two atoms of a different element (d).

x-direction (Figure 3.3b). Again, both atoms have to be of the same element and must not be displaced in y- and z-direction by more than 0.1 Å. Mirror planes perpendicular to the y-direction are identified in an analogous manner.

Based on this set of symmetry elements for a single atom, the symmetry elements of the surface are identified. This is achieved by stepwise applying every obtained symmetry operation to every atom of the surface model. In case a symmetry operation is not mapping an atom on another atom of the same element, this symmetry operation is rejected from the pool of possible operations. Again, a mapping is only valid in case the displacement between the atoms is smaller than 0.1 Å. Finally, symmetry operations which are successfully applied to every atom represent valid symmetry operations of the surface.

With the obtained symmetry elements of the surface a minimal area is defined as sketched in Figure 3.4. The starting point is the whole surface area (Figure 3.4a). This area is reduced in case a translational symmetry in x- or y-direction was identified (Figure 3.4b). In this case the distance between the borders of the minimal area matches the length of the translation vector in this direction. The position of a minimal area reduced by a translational symmetry element can be chosen completely arbitrary. In a second step the distance between two parallel mirror planes is compared to the size of the current minimal area perpendicular to the mirror planes. In case the distance between the mirror planes is smaller than the width of the minimal area, the minimal area can be reduced (Figure 3.4c). This is achieved by adjusting the position of the minimal area to the position of the mirror planes. With the resulting minimal area a grid dividing the surface is generated.

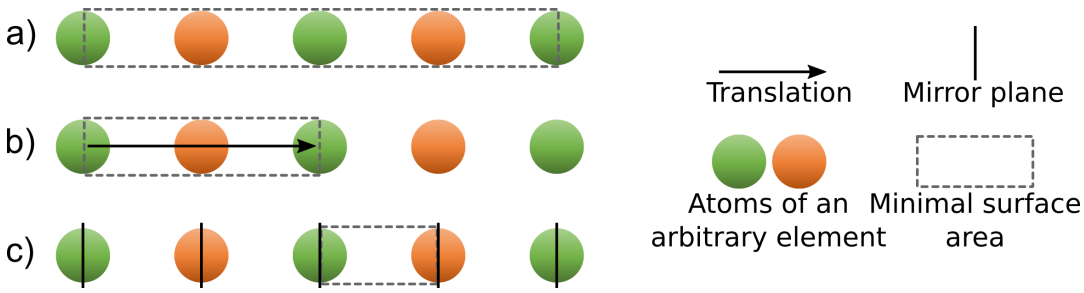


Figure 3.4.: Rules to detect the minimal area of the studied surface. The starting point is always the complete area of the surface (a). This area is minimized by stepwise considering the translational (b) and mirror symmetry (c). The minimal area is positioned in the center of the surface where possible.

Introduction of a Grid

As sketched in Figure 3.5a the grid points are equally distributed in x- and y-direction while their density is controlled by a user specified keyword. The z-position of every grid point is optimized to maintain a certain distance between the adsorbate and surface instead of keeping the grid point at a fixed z-position. In this way the position of the grid points is accounting for different orientations of the adsorbate and different structural motifs of the surface.

To adjust the z-position of the grid points every orientation of the studied adsorbate is generated first. All orientations are obtained by applying rotation matrices to the atomic coordinates of the adsorbate. Here, the exact number of used rotation matrices and thereby orientations depend on a user defined number of rotations around every coordinate axis. The exact structure of the rotation matrices follows the z-y-x convention and is explained in detail in the appendix A.3. However, not every possible rotation matrix is used to generate an initial guess for adsorption minima. Instead, unique orientations are identified to reduce the number of initial guesses for adsorption minima. This is achieved by creating all possible orientations once and comparing the structural similarity of all obtained orientations. Here, the structural similarity is judged based on the common criterium of the RMSD (appendix A.1).

However, the RMSD has the crucial downside that the obtained values depend on the ordering of the atoms in both orientations. As shown in Figure 3.6b a non ideal ordering leads to a larger RMSD value. This is especially a problem for rotations since the rotational symmetry of the adsorbate is not accounted for by the RMSD itself. A solution to this obstacle is to use a RMSD value minimized with respect to the ordering of the atoms as shown in Figure 3.6c. For such an assignment problem the Hungarian algorithm[183,184] can be used as discussed in the literature[185–187]. In the following a corrected RMSD value is referred as H-RMSD.

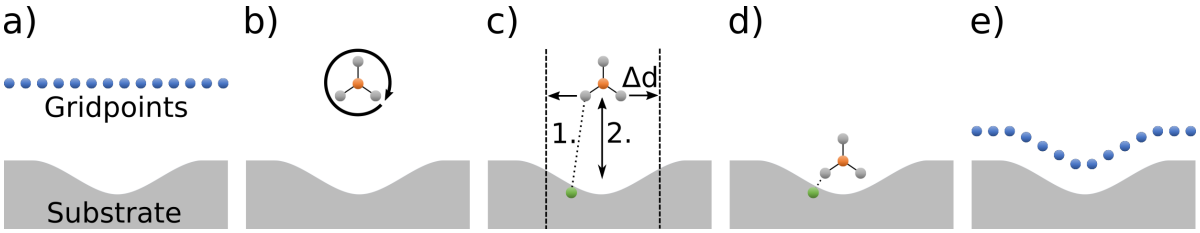


Figure 3.5.: Introduction of a grid for the initial guesses of adsorption minima. (a) The grid equally divides the surface in x- and y-direction. (b) The adsorbate is placed with different orientations by its center of mass on every grid point. (c) The height of the adsorbate is adjusted to a selected distance to the surface. (d) The final guess for an adsorption minimum. (e) Adjusting the adsorbate height results in a grid following the surface structure.

In Figure 3.6 the basic idea of the Hungarian algorithm is sketched while an extensive explanation is found in the appendix B.1. The basic idea of the Hungarian algorithm is to start with the best assignment for every atom (Figure 3.6d). In our case the best assignment is the smallest distance between the assigned atoms and by this the smallest contribution to the RMSD. In case this initial assignment leads to one atom being assigned to several other atoms - D1 to C2 and D1 to C3 in Figure 3.6d - the algorithm resolves this double assignment with the smallest possible cost. In the example of Figure 3.6 this results in D1 being assigned to C2 and D2 to C3 (Figure 3.6e).

By using the root mean square deviation minimized by the Hungarian algorithm (H-RMSD) value identical orientations are identified in case the H-RMSD is falling below a user defined threshold. Consequently, only one set of rotation angles leading to this orientation is kept for the following steps. With the unique orientations the initial guesses for the adsorption minima can be generated. Following Figure 3.5 for every combination of grid point and unique rotation an initial guess is generated. For this the molecule is placed with its center of mass on the x- and y-position of the grid point (Figure 3.5b). For the z-position the upper limit of the unit cell is chosen to adjust the exact position in the next steps.

To adjust the z-position and iterative procedure is applied (Figure 3.5c). First, the smallest distance between an atom of the adsorbate and surface is searched. In this step PBC are only

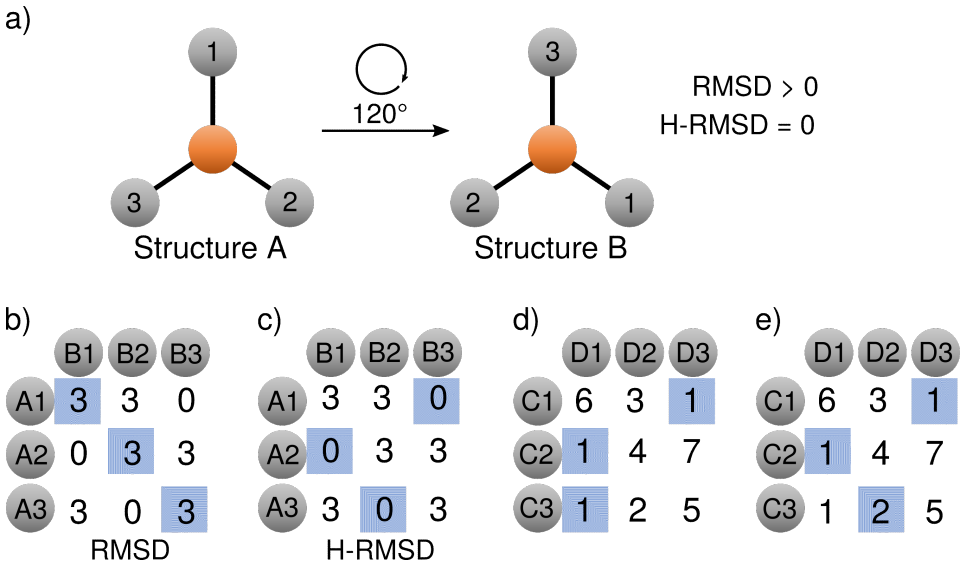


Figure 3.6.: Structural similarity judgement based on the RMSD and the H-RMSD. Consequence of a rotation to the atom labels sketched in (a). Arbitrary distance matrices for (a) with pairs of atoms selected by the RMSD (b) and by the H-RMSD (c) highlighted in blue. Example for the initial (d) and final (e) selection of the Hungarian algorithm with an arbitrary structure.

considered for the x- and y-direction. However, not every atom of the surface is considered for this step. Only surface atoms within a displacement Δd in the xy-plane to any adsorbate atom are considered. This restriction can be used to allow the adsorbate to fall into a deep trench at the surface in case the selected displacement cutoff Δd is smaller than the desired distance between the adsorbate and the surface. In case no surface atom at all is found within Δd, the cutoff is automatically increased by $10\,\%$. However, it is necessary to make sure that the seeked distance between the adsorbate and surface has to be at least equally or larger than Δd. Otherwise the derived height of the adsorbate might be ill defined in equation 3.1.

Once a surface and adsorbate atom was selected the height of the adsorbate is adjusted based on the wanted distance r^W and the coordinates of the surface r^S and adsorbate r^A atom as shown in equation 3.1. However, in case the wanted distance r^W is smaller than the in-plane displacement Δd the square root term results in an imaginary number and with this in an imaginary height for the adsorbate.

$$r_z^A = \sqrt{(r^W)^2 - (r_x^S - r_x^A)^2 - (r_y^S - r_y^A)^2} + r_z^S \tag{3.1}$$

With the known z-position r_z^A of the adsorbate atom the whole molecule is shifted. These steps are repeated until no further change in the height of the adsorbate is observed and the final initial structure (Figure 3.5d) is obtained. Overall, this whole procedure of adjusting the height of the adsorbate to the surface represents a grid following the silhouette of the surface as sketched in 3.5e.

3.2.2. Obtaining the Adsorption Basin

So far, the creation of initial guesses to obtain the adsorption minima was introduced. With the adsorption minima also the adsorption energy is known, which can serve as a guess for the likelihood to observe this structure. Especially in case a thermodynamic equilibrium is present, the most stable adsorption structure is the most important one. Here, the molecule might adsorb in a less stable minimum but is then able to diffuse to the most stable adsorption minimum assuming that the diffusion is not hindered by large reaction barriers. However, in case consecutive reactions like decomposition reaction have a similar or even smaller energy barrier as diffusion reactions a thermodynamic equilibrium between the adsorption structures is no longer established. In this case, differences in the reaction barriers for different adsorption minima can be crucial to understand the reactivity of an adsorbate.

To link the reaction network of an adsorbate to experimental observations by for example kinetic Monte Carlo (kMC) simulations[78] an initial guess for the population of the adsorption

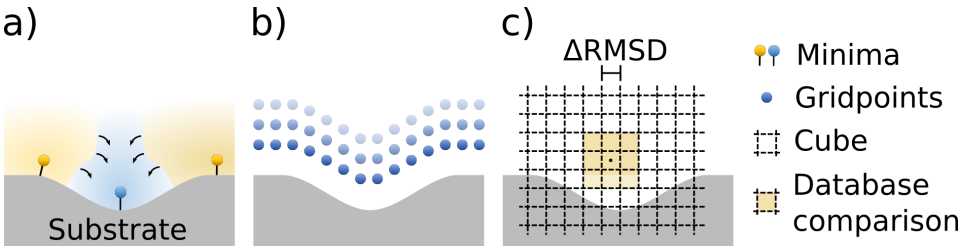


Figure 3.7.: (a) Introduction of the adsorption basin for the observed adsorption minima. All structures leading to the same adsorption minimum correspond to the same adsorption basin. (b) To obtain the basin the grid is extended in the z-direction following the previously introduced rules. (c) Cubes dividing the unit cell for improving the database comparison. Marked cubes indicate an example for the considered area for a database comparison since only neighboring cubes to the optimized structure are considered.

minima is necessary. Here, a random population of the adsorption minima or a population based on the adsorption energies would be possibe. However, it is more common to estimate the adsorption probability based on the size of the adsorption site and the local sticking coefficient[188]. Here, the size of the adsorption site is for example roughly estimated by the density of the adsorption sites[79,189]. To obtain a more accurate criterium based on the actual structure of the adsorption site, I would like to estimate the relative volume of the basin of the PES for every adsorption minimum. Unfortunately, sampling the whole multidimensional space of the PES even by enhanced sampling methods like metadynamic calculations[133–137] would be computationally too expensive. Therefore, with PESE, I obtain a rough estimate of the basin by sampling the 3-dimensional space of the adsorbate positioned above the surface. In this sampling all structures above a surface leading to the same adsorption minimum are considered as part of the same adsorption basin as sketched in Figure 3.7a. I want to use the relative size of these adsorption basins as a geometric criterium to judge the adsorption probability of the adsorption minima. To calculate the adsorption basin, the grid introduced in the previous section is extended in the z-direction by forming several layers of grid points (Figure 3.7b). As for the initial guesses for the adsorption minima, for every grid point several orientations of the adsorbate are used and the actual z-position is adjusted based on the distance between adsorbate and surface.

Once all initial structures are generated the trivial approach to obtain the basin would be to optimize every generated structure and combine those resulting in the same adsorption minimum. However, this would be computationally very expensive, since the more distant an adsorbate is to the surface, the more optimization steps are necessary to reach the structure of the adsorption minimum. Also, this approach would be highly inefficient since in the

optimizations of structures where the adsorbate is further away from the surface most likely already calculated structures would be observed. For the sake of efficiency, PESE is optimizing every initial structure for the adsorption basins until a minimum structure or an already known structure is obtained. This requires that all structures obtained in the structural optimizations are stored. A minimum structure is detected as usual by the forces being smaller than the selected convergence threshold. To judge whether a known structure is reached three criteria based on the available information of the energy, structure and force are introduced: The first criterium demands that the total energy of the optimized and known structure differs by less than a user defined threshold. The structural similarity is again judged based on the previously introduced H-RMSD value. For simplicity, this value is calculated solely with the adsorbate atoms. The last criterium is based on the forces. However, PESE is not directly comparing the forces of the optimized and known structure. Instead, the atoms of the optimized structure are moved along the forces acting on them. Then, the H-RMSD value to the known structure is calculated again. The third criterium is fulfilled if the H-RMSD is decreasing after applying the forces. This means that the currently optimized structure is moving in the direction of the known structure.

Following these rules, the optimizations of an initial structure can be terminated as early as possible. However, to ensure a high efficiency of the basin calculation, the comparison of an optimized structure with known structures has to be performed within an neglectable amount of time. Consequently, a sophisticated implementation of these rules in PESE was seeked. As a first aspect of a performant implementation, the storing of the known structures was addressed. An efficient storing of the known structures is obtained by dividing the whole unit cell in small cubes as shown in Figure 3.7c. Here, structures are sorted into the cubes by their centroid. Furthermore, the size of the cubes is selected to be equal to the H-RMSD threshold of the structural comparison. Consequently, structures fulfilling the H-RMSD criterium are mostly found in the same or a neighboring cube. Therefore an optimized structure is compared only to the structures in neighboring cubes as sketched in Figure 3.7c. In this way, the number of necessary comparisons is minimized. Furthermore, the known structures are stored on the hard drive sorted by their cube to minimize the access and reading time on the hard drives. Here, a redundant array of inexpensive disks (RAID)[190] system based on hard disk drives (HDDs) should be used as the storage system to enable small access times and high data transfer rates. A replacement of the HDDs by more efficient solid-state drives (SSDs) would presumably facilitate only a small increase in performance due to the small size of the structure files. However, the usage of SSDs would come along with a significant increase in costs in comparison to HDDs. Overall, only the information about the total number of structures

and structures per cube has to be stored in the main memory and restored every time PESE is restarted.

As a second part of a performant implementation, an efficient comparison of the optimized to the known structures had to be developed. This is achieved by parallelizing the comparison over all cores of the evaluator. Here, I follow the data parallelism since every core of the evaluator performs the same comparison but with different known structures within the defined cubes. In case an agreement is found the comparison is terminated for all cores. Furthermore, only the information whether and to which known structure the optimized structure is equal to is stored. The final assignment of all structures to a minimum is done in a subsequent evaluation step. However, a minor check of the assignment is necessary to prevent an indirect self-assignment of known structures, which would result in structures being assigned to no minimum at all.

3.2.3. Evaluation of Adsorption Minima and Adsorption Basins

The previous sections introduced the generation of initial guesses for adsorption minima and the adsorption basins. As discussed in section 3.1, these structures are optimized by workers. Once, these optimizations are completed all adsorption minima are obtained. In principle the initial guesses for decomposition products could be generated based on this set of adsorption minima. However, as a consequence of the used brute force approach several structures for the same adsorption minima are obtained. Therefore, PESE checks all adsorption minima for duplicates to continue the creation of initial guesses for decomposition structures with the smallest possible set of adsorption minima. The removal of duplicates is achieved by judging the similarity of adsorption minima by a total energy and a H-RMSD threshold. Here, the H-RMSD value is further extended to obtain a root mean square deviation minimized by the Hungarian algorithm under consideration of the surface symmetry (SH-RMSD). The SH-RMSD value is determined by calculating a H-RMSD value for every possible symmetry operation and selecting the smallest obtained H-RMSD value. This is necessary to identify adsorption minima located at the edges of the minimal surface area as duplicates. With this unique set of adsorption minima the initial structures for decomposition products are generated as described in section 3.2.4.

For the subsequent calculations, the removal of duplicates is the only necessary evaluation step. To estimate the volume of the adsorption basin for every adsorption minimum additional evaluation steps are necessary. However, these steps are computationally expen-

sive. Therefore, the final evaluation of the adsorption basins described here is separately performed once all quantum chemistry calculations are completed by the workers.

The first step to evaluate the adsorption basins is to read all structures stored on the hard drive and to extract all information of interest. Here, the total energy to calculate the adsorption energy and the dispersion contributions to the total energy are relevant. Also, the centroid of the adsorbate and the orientation of the adsorbate is stored. Furthermore, for every structure the value of the largest force is searched to judge the convergence behavior. Finally, the most crucial information of the current assignment of every structure is saved. It is important to note that the coordinates of every structure are actually not needed in every of the following steps and therefore also not permanently stored. In this way the memory demands are kept as small as possible.

While reading these information the assignments are processed simultaneously. Here, the assignments obtained during the optimizations are the starting point and a recursive assignment is performed to assign every structure to an adsorption minimum. Furthermore, the removal of duplicates has to be considered in the sense that all structures previously assigned to a duplicate have to be reassigned to a unique minimum. In addition to processing the assignment, the adsorption type is defined for every structure.

Here, every structure is considered to be chemisorbed, physisorbed or a gas phase structure. The gas phase structure is easily identified by a lack of interaction between the adsorbate and the surface. Consequently, this adsorption type could be recognized by a large distance between the adsorbate and the surface and an adsorption energy close to zero. The distinction between a chemisorbed and a physisorbed structure is however more complex. Ideally, a chemisorbed structure would be identified by a strong bonding of the adsorbate to the surface mainly due to electronic and orbital based interactions while a physisorbed structure would show interactions mainly based on van der Waals interactions. However, in both cases a large adsorption energy can be observed. Therefore, the adsorption energy itself is not a meaningful criterium. Instead, a decomposition of the adsorption energy is needed for an interpretation of the individual contributions[191]. However, even in case the individual contributions are known, a straightforward interpretation, in the sense that a larger contribution to the total energy due to van der Waals interactions represents a physisorbed structure, might not be possible[191].

Since PESE is neither running additional calculations to derive the individual contributions to the adsorption energy in detail nor able to automatically derive reasonable interpretations of

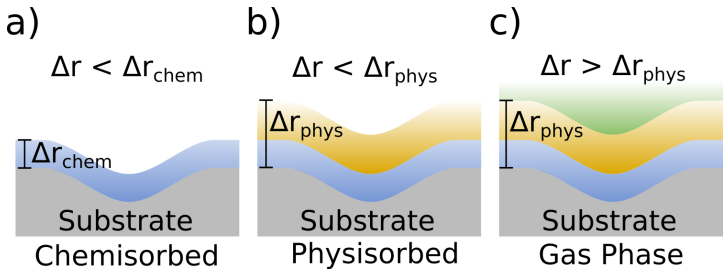


Figure 3.8.: Definition of adsorption types based on the distance Δr between the adsorbate and the surface. (a) In case the distance is smaller than Δr_{chem} the structure is considered as chemisorbed. (b) In case the distance is only smaller than Δr_{phys} the structure is considered as physisorbed. (c) Otherwise the structure is considered to be in the gas phase. All distances are defined as multiples of the covalent radii.

the energy decomposition, a simplification is used to identify chemisorbed and physisorbed structure based on the distance between the adsorbate to the surface. A shortcoming of this approach is that physisorbed structures showing strong van der Waals interactions with the surface and consequently a small distance to the surface might be identified as chemisorbed structures. However, these physisorbed structures might be interesting as a starting point for decomposition reactions and the incorrect identification thereby less problematic. The final approach used in PESE is sketched in figure 3.8. Here, the adsorption type is defined based on the smallest distance between the adsorbate and the surface. However, the smallest distance is not an absolute value in Å. Instead, every distance between two atoms is weighted by the sum of the covalent radii of the two atoms. Thereby, the absolute value of the distance depends on the type of atoms in the adsorbate and surface. This simplifies the selection of a reasonable threshold since this value is less dependent on the system. PESE also allows to select an additional adsorption energy criterium to define the adsorption type although it is problematic as previously discussed. Furthermore, this criterium is strongly dependent on the system and therefore less handy. As discussed in the next paragraphs the detection of the adsorption type is also used to reduce the memory requirements.

Once all information is obtained the size of every basin should be estimated. However, the dataset derived so far has two shortcomings. First, the structural data are arbitrarily distributed in space. This makes an estimation of the basin volume difficult. The second shortcoming is that not the whole space above the surface is sampled since the grid to generate the initial structures for the adsorption basin can not be chosen to be extremely dense. So, we are interested in an easy approach to estimate the basin volume while smearing the data to fill the whole space. This can be achieved by mapping the dataset on an equally spaced grid while every datapoint is smeared by a Gaussian function (GF). The equally spaced grid

enables us to estimate the volume by simply counting the grid points, while the smearing by a Gaussian function ensures that the whole space is covered. The usage of a Gaussian function as the smearing function is not mandatory here but simply convenient.

Following this idea two steps are necessary to estimate the size of the adsorption basins. First a projection on the equally spaced grid and then an evaluation of this grid. Here, the number of grid points is again defined by a user specified keyword. The projection or smearing of the data points is done by positioning the Gaussian function at the centroid μ of the adsorbate. An unnormalized Gaussian function shown in equation 3.2 is used to obtain a value of one exactly at the centroid of the structure. For performance reasons the smearing by the Gaussian function is capped at a certain radius around the centroid of the structure. Finally, the contributions of the minima to every grid point are normalized if necessary since the sum of all relative contribution must not be larger than $100\,\%$. Due to the huge number of grid points and structures to be mapped PESE is exploiting the data parallelism for every core of the evaluator by parallelizing the mapping over the stored structures.

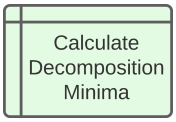
$$\mathrm{GF}(\mathbf{r}) = e^{-\frac{1}{2}\frac{(\mathbf{r}-\mu)}{\sigma}^2} \tag{3.2}$$

Once the mapping is completed the size of a basin can be estimated based on the equally spaced grid. The contributions of every minimum to a certain grid point is stored in a matrix. Due to the huge number of grid points and minima this matrix can be large. Therefore, all minima assigned to be in the gas phase are combined before the projection and treated as a single minimum. This drastically reduces the size of this matrix and by this the memory requirements during the projection. With this matrix, different distribution schemes are possible to estimate the volume. Here, three different approaches are implemented in PESE: In a first scheme every grid point is counted only for the minimum, which has the largest relative contribution to this grid point. This scheme is in spirit similar to the calculation of Mulliken partial charges. The second scheme is a slight variation to the first scheme. Here, every grid point is counted to a minimum which is dominant as indicated by a relative contribution larger than a certain percentage. In a third scheme the grid points are partially counted based on the contribution of every minimum. Independent, from the selected partitioning an estimation for the adsorption basin is obtained and the evaluation thereby completed.

A comparison of the adsorption probability derived based on the basin size to other computational or experimental approaches is possible. Here, a high number of MD simulations can be performed to estimate the adsorption probability based on the observed adsorption

trajectories[192]. However, performing a statistical meaningful number of MD simulations is a computationally demanding task. Experimentally, the adsorption probability is accessible by dosing a small number of molecules on a surface and counting the occupied adsorption sites on scanning tunneling microscope images[93,193]. Here, the concentration of molecules has to be small so that the molecules can be assumed to be independent in their adsorption behavior.

3.2.4. Obtaining Decomposition Products



In the previous sections, the creation of initial guesses for adsorption minima and the selection of unique minima were introduced. Based on these unique adsorption minima the initial guesses for decomposition structures are generated. Also, the adsorption type can be used to exclude minima corresponding to gas phase or physisorbed structures. In this way only decomposition reactions starting from chemisorbed minima are derived. As for the generation of initial guesses for the adsorption minima, PESE is not following the approach of using reactivity descriptors[164] for the decomposition products. Instead, PESE is following the idea of representing any reaction based on connectivity descriptors[147]. The connectivity descriptors are derived based on the atomic distances. Consequently, no additional calculations to estimate the reactivity of a functional group for example are needed. Decomposition reactions are then described by manipulating the connectivity descriptors. This approach is a straightforward way to create initial guesses for decomposition structures. Nevertheless, it is sufficient to describe every decomposition reaction of interest. Also, additional types of reactions can easily be included. However, the disadvantage of this approach is that every individual decomposition reaction of interest has to be implemented in advance. Also, a decomposition reaction, which is not described by any of the implemented rules, can not be studied with PESE. Consequently, the selection of rules in the implementation as well as in the application of PESE is the most crucial point to miss potentially important reactions due to a developer or user bias.

In PESE only the most simple decomposition reactions shown in Figure 3.9 are yet included. Here, the used connectivity descriptor is the presence of bonds between atoms. As a criterium to decide whether a bond is present or not the common comparison to the sum of the covalent radii[194] is performed: In case the sum of the covalent radii times 1.3[194] is larger than the interatomic distance a bond is identified. The three rules of Figure 3.9 are based on the manipulation of bonds. By breaking a single bond two fragments are obtained (Rule

1, Figure 3.9a). This rule corresponds to homolytic cleaving of bonds and the formation of radicals. For the second decomposition type (Rule 2, Figure 3.9b) two bonds are broken and a new bond is formed between the former unbound fragments. By this rule elimination reactions like the important β-H-elimination reaction[67,68] are described. In addition to these two decomposition reaction types a third reaction is implemented in PESE. Following Figure 3.9c two fragments obtained by the first rule are combined to one molecule. This rule describes radical recombination reactions and is in particular useful to obtain a set of new molecules, which might be formed by the decomposition of the studied molecule.

The rules introduced here have the limitation that they cannot handle cyclic compounds. For these a cleaving of a single bond does not necessarily result in two fragments, which is conditional to decomposition reactions. However, this connectivity descriptor based approach enables to include more complex types of reactions like reactions with the surface as in ALD or alkyl-exchange reactions[195] by simply adding the corresponding rules. To fine tune the studied decomposition reactions additional functionalities are implemented in PESE . Here, the generation of decomposition products can be restricted to the generation of user defined fragments. By this, the number of fragments can be reduced to relevant fragments only.

All introduced decomposition schemes can be applied to the initial adsorbate structure in the

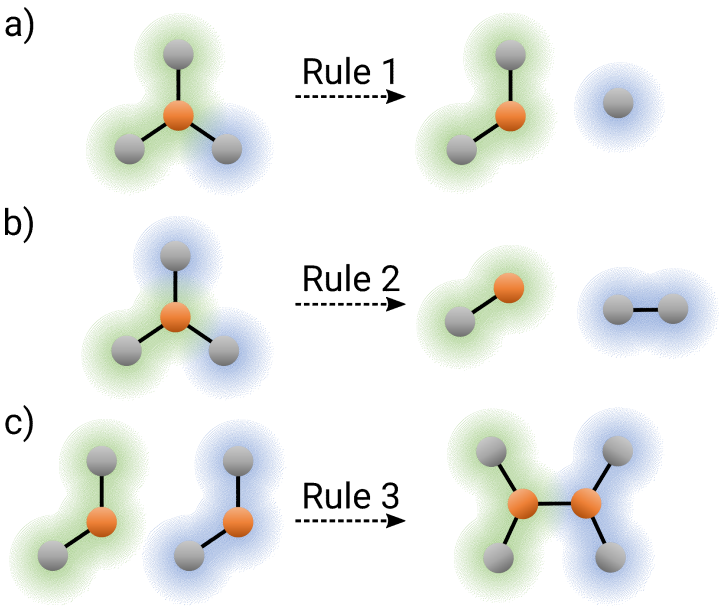


Figure 3.9.: Rules to derive decomposition structures. (a) Homolytic cleaving of a single bond results in two radicals. (b) Breaking of two bonds and forming a new bond between the former unbound fragments describes elimination reactions. (c) Radical recombination by combining two fragments obtained by rule 1 in (a).

gas phase. By this a set of possible decomposition structures is generated, which could then be investigated in separate calculations. To obtain initial guesses for decomposition structures adsorbed on the studied surface the unique adsorption minima are used. Here, the rules of Figure 3.9 only introduce the fragmentation scheme, while the creation of the actual structure is sketched in Figure 3.10: By applying any of the first two decomposition rules to the adsorption minimum (Figure 3.10a) two fragments are obtained (Figure 3.10b). For the next step, a main fragment has to be detected. For this the number of bonds between the surface and the fragment, the number of atoms per fragment and the mass of the fragment is used. The fragment with the largest number of bonds to the surface, the largest number of atoms or the largest mass is considered as the main fragment. In case both fragments are identical the first fragment is the main fragment. The main fragment is used to create an equally spaced grid. As sketched in Figure 3.10c the grid is oriented at the center of mass of the main fragment. Furthermore, the grid is restricted to a certain radius r around the main fragment. In addition, all grid points have to rest within a certain angle φ to the vector connecting both centers of mass of the fragments. This definition ensures that the second fragment is neither located far away from the main fragment nor moving to the other side of the main fragment. In a next step the second fragment is rotated so that the dangling bond is pointing towards the surface (Figure 3.10d). This rotation is only necessary in case the fragment was derived by the first decomposition rule. Then the fragment is placed on all grid points by adjusting the height to the surface as discussed for the adsorption minima. For the height adjustment, the atoms of the main fragment are ignored to enable a bond formation to the surface. However, to maintain a certain distance between both fragments the second fragment is shifted along the center of mass vector to increase the in-plane distance between the fragments to 1.4 covalent radii (Figure 3.10e).

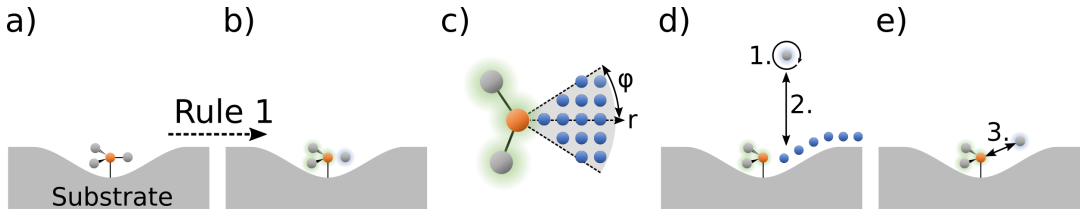


Figure 3.10.: Generation of initial guesses for decomposition structures. (a) The starting point is a previously obtained minimum structure. (b) A decomposition rule is applied to the adsorbate. (c) The main fragment is detected as described in the text and a regular grid is generated. The grid is defined by a radius r and angle φ to the main fragment. (d) The second fragment is rotated if necessary and placed on the grid points ignoring the main fragment. (e) The distance between the second and main fragment is adjusted if necessary.

Equation 3.3 shows the starting point to calculate the magnitude d of this displacement. Here, $\mathbf{r}_2$ and $\mathbf{r}_1$ are the x- and y-coordinates of atoms in fragment two and one, respectively. These atoms are selected based on the smallest distance between the fragments. After a displacement $\Delta\mathbf{r}$ of the second fragment a distance of $1.4r_{cov}$ should be obtained.

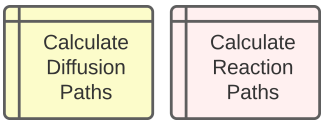
$$1.4r_{cov} = \sqrt{\left[\mathbf{r}_2 + d \cdot \Delta\mathbf{r} - \mathbf{r}_1\right]^2} \tag{3.3}$$

Transforming equation 3.3 results in equation 3.4. Here, two solutions to d can be obtained by solving this quadratic equation by for example the pq formula. For the displacement the largest, positive solution is then used. Since this displacement is calculated by only considering the xy-plane, the actual height of the second fragment must be adjusted again.

$$\begin{aligned} 0 = d^2 & \\ + d \cdot & \underbrace{\frac{2\Delta r_x(r_{1,x} - r_{2,x}) + 2\Delta r_y(r_{1,y} - r_{2,y})}{\Delta r_x^2 + \Delta r_y^2}}_{p} \\ + & \underbrace{\frac{(r_{1,x} - r_{2,x})^2 + (r_{1,y} - r_{2,y})^2 - (1.4r_{cov})^2}{\Delta r_x^2 + \Delta r_y^2}}_{q} \end{aligned} \tag{3.4}$$

Following the described steps, the initial guesses for decomposition structures are generated. As the initial guesses for the adsorption minima, these structures are optimized by the workers. Once all decomposition structures for one adsorption minimum and one type of decomposition are optimized, duplicates are removed by the introduced SH-RMSD values.

3.2.5. Obtaining Reaction Paths



The previous sections introduced the generation of adsorption and decomposition structures and their optimization to obtain the corresponding minima on the PES. Based on all obtained minima diffusion and decomposition paths are proposed. This section introduces how the starting point and endpoint of every reaction is selected, how the initial guesses of the reaction paths are derived and how the paths are then finally optimized by the workers.

Reaction Path Selection

The selection of the starting point and endpoint for decomposition reactions is obvious. Here, the adsorption minima used to create the decomposition products are always the starting point while all unique decomposition structures are the endpoint for their individual decomposition path.

For diffusion paths the selection is less obvious. The simplest approach would be to select diffusion paths based on their length or change in energy[147,164]. The length of the diffusion path could be judged by the RMSD value between the starting point and endpoint. However, the selection of a suitable RMSD threshold to accept or decline a diffusion path is complicated. In Figure 3.11 the problem is illustrated. If the threshold is set too small (Figure 3.11a) only minima close to each other are connected while no complete network is obtained. If the threshold is set too large (Figure 3.11b) too many diffusion paths are calculated, which is computationally expensive.

A solution to this problem is the introduction of a relative length threshold. Here, within PESE, I follow the approach to discard a direct diffusion from i to j with the length r_{ij} in case there is a two-step diffusion process over structure k. This is achieved by the following rules whereby the direct diffusion is discarded in case all rules are fulfilled:

- $r_{ik} < r_{ij}$
- $r_{kj} < r_{ij}$
- $r_{ik} + r_{kj} < \lambda \times r_{ij}$

The first and second rule state that both steps r_{ik} and r_{kj} of the two-step process are required to be smaller than the direct diffusion r_{ij}. In addition, the third rule controls the additional distance acceptable to discard a direct for two shorter diffusions. Here, the parameter λ is only in the range 1 to 2 meaningful. Setting λ to 1 would result in every path being selected since it is impossible to find two paths, which are in their sum smaller than the direct diffusion path. Setting λ to 2 results in a minimal diffusion network since only two paths, which are individually shorter than the direct path, have to be present. This case is also sketched in Figure 3.11c. The introduction of this relative length threshold λ is also from an energetic point of view meaningful. Here, I follow the assumption that a short diffusion path with small structural changes corresponds to a small reaction barrier. This assumption is justified by the so called principle of least motion[196–198], which is discussed in detail in the next section. Consequently, by following the introduced rules long diffusion path with likely a large reaction

barrier are replaced by two short diffusion paths with likely smaller reaction barriers. In PESE both approaches of a relative and absolute length are combined: The relative length is used to derive a complete diffusion network while an absolute length and energy threshold can be used in addition to discard too long or endothermic diffusion paths. It is also possible to stepwise select diffusion paths by decreasing the value for λ. In this case, PESE performs a comparison to the previous value to identify the newly selected diffusion paths.

For the selection of diffusion paths to obtain a complete diffusion network it is not sufficient to consider only the unique minima. In PESE all minima are duplicated based on the symmetry rules. This is necessary to also obtain diffusion paths across the border of the minimal area. Furthermore, all minima in neighboring unit cells are created by using PBC. In this way, also diffusion paths passing the unit cell border are included. However, duplicating all minima using PBC results in $9N$ minima in total with N being the number of minima per unit cell. Consequently the selection of the diffusion paths based on the relative length requires to perform, in the worst case, $(9N-1)\cdot(9N-2)$ comparisons per possible path. Therefore, an efficient selection of diffusion paths by parallelization is implemented in PESE. Furthermore, a performance optimization is possible by prescreening all diffusion paths: A complete set of reaction paths is already obtained by only considering those paths where the adsorbate is starting within the minimal area defined by the deteced surface symmetry. Therefore, the position of the center of mass for every adsorbate is detected and only diffusion paths starting in the used minimal area are considered. However, prior to this step the atoms of the adsorbate have to be normalized: All atoms are shifted to minimize the distance to the

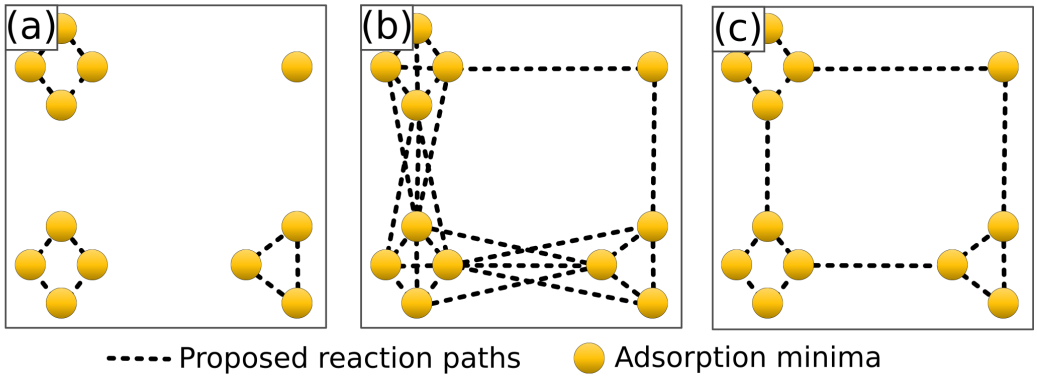


Figure 3.11.: Selection of diffusion paths based on their length. (a) A small RMSD value was selected to calculate only short diffusion paths. (b) A larger RMSD value was proposed in comparison to (a) to also include diffusion paths linking the previously isolated groups. (c) A relative RMSD threshold is used for the path selection with the corresponding rules described in the text and $\lambda = 2$.

first adsorbate atom as the fix point. This is necessary since structures close to the edge of the unit cell might be split due to the PBC. This would result in a wrong center of mass value. In addition to this prescreening, in PESE the calculation of diffusion paths can be restricted to certain types of diffusion paths as for example diffusion paths linking chemisorbed minima only.

Reaction Path Interpolation

Once a set of decomposition and diffusion paths is proposed, an initial interpolation for each reaction path has to be created. The reaction path interpolation is decisive for the later performed reaction path calculation. A good interpolation is necessary to obtain a quickly converging reaction path with the lowest transition state. The interpolation itself consists of two crucial steps. Firstly, the assignment of atoms in the initial and final structure of the path and secondly the actual interpolation. Here, the assignment defines the structural changes observed along the reaction path, i.e. bond breaking or formation events or conformer changes, while the interpolation is the initial guess how everything happens, i.e. smooth or abrupt changes and simultaneous or successive reaction events.

For every decomposition and diffusion path it is useful to perform the assignment in a way that a reaction path with the smallest possible reaction barrier is obtained. Reaction paths with small reaction barriers are of special interest since they mainly define the kinetics of the system[188]. However, at this point the reaction barrier is unknown while only the initial and final structure of the reaction path is known. Therefore, an estimate linking the change in structure to the reaction barrier is needed. Here, the principle of least motion[196–198] was introduced as a possible approach. This principle states that the smallest activation energy is observed in case the motion of atoms and changes in the electronic configuration are minimized since both processes would require energy. However, also reactions violating this principle are known[199,200]. Still, since some guidance is needed for an automated assignment of atoms, PESE is following the principle of least motion while future extensions to PESE will address alternative assignments and reaction paths. An exception to this approach is made for bond breaking and formation events since these events mostly likely come along with a larger change in energy in comparison to translation and rotation events or minor changes of internal coordinates. Therefore, the side condition is introduced that the number of formed and broken bonds should be minimal in the assignment. Unfortunately, this side condition is non-trivial since a direct combination with the Hungarian algorithm could not be achieved. Therefore, a more sophisticated algorithm shown in Figure 3.12 was developed.

For this algorithm two concepts have to be introduced first:

- Local equality: Local equality is present in case one atom of a first structure and one atom of a second structure have the same element and number of bonds to other elements. As an example consider the C^1 atom in $H_3C^1-CH_3$ and $H_3C^1-CH_2-OH$. Here, the atom has three C–H bonds and a single C–C bond. Both C^1 atoms are locally equal.
- Global equality: Global equality is present in case two structures have the same bonding pattern. An equivalent formulation is that two atoms of the same element have the

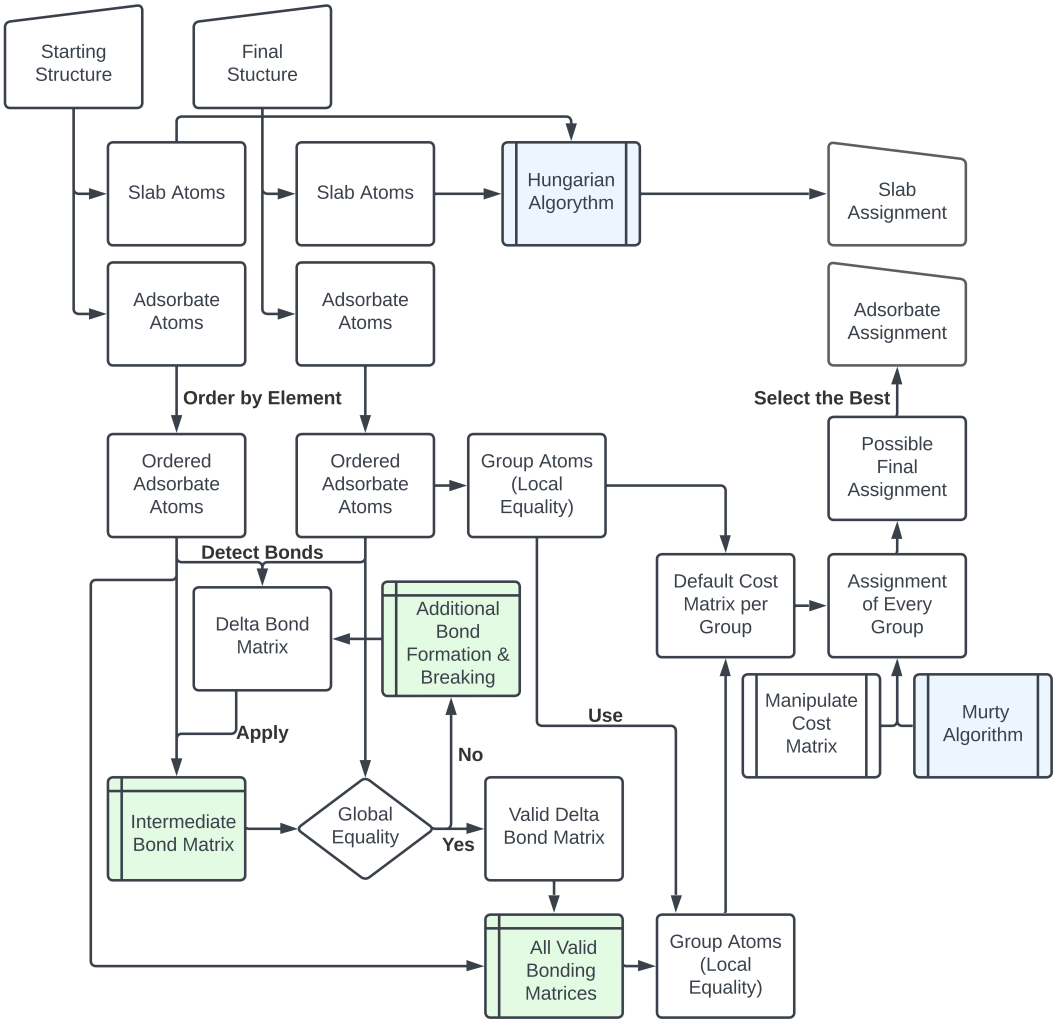


Figure 3.12.: Flowchart of the enhanced assignment algorithm. The details of this algorithm are explained in the text. Blue marked elements indicate an utilization of a literature known assignment algorithm. Green marked elements indicate recursive routines. All steps after a recursive routine are applied to every outcome of this routine even if the subsequent steps are recursive routines themself.

same recursive bonding pattern. Here, the two structures $H_3C^1 - CH_3$ and $H_3C^1 - CH_2 - OH$ are an example for the absence of global equality. To determine whether two structures are globally equal the number of elements, number of bonds and also the local equality for every atom is used as prescreening.

The algorithm shown in Figure 3.12 starts by dividing every structure in the atoms of the slab model and the adsorbate. For the atoms of the slab model an assignment by the Hungarian algorithm is performed. For these atoms an enhanced assignment scheme is not necessary since they mostly are unchanged in the decomposition reactions of the adsorbate. However, a simple assignment is still necessary since the ordering of atoms might be changed due to the application of the detected symmetry elements in previous steps. In a second step, the adsorbate atoms are sorted by their element if necessary. Based on the bonding matrix for every adsorbate structure in the initial and final structure the so called delta bond matrix is obtained. This is a difference matrix and displays the number and type of bonds, which have to be formed or broken to change from the initial to the final structure. This delta bond matrix gives a first idea of the necessary transformations. However, in case a type of bond is formed and broken this transformation is not contained in the delta bond matrix. Therefore, the delta bond matrix has to be verified. This is done by applying the matrix to the bonding matrix of the first adsorbate. Here, all combinatorial possible bonding matrices have to be generated in a recursive manner since it is not known which bond should be formed or be broken. All obtained bonding matrices are compared to the bonding matrix of the second adsorbate structure. In case global equality is observed in a single comparison, a valid delta bond matrix describing the reaction is obtained. In case none of the created bond matrices matches the bond matrix of the second adsorbate structure, more changes are necessary. In this case one additional bond is formed and the same type is broken. Following this approach more and more delta bond matrices are used. This approach corresponds to the idea of minimizing bond breaking and formation events for the reaction path.

While the delta bond matrix shows how many bonds of each type are formed or broken, it does not contain the information to which part of the bonding matrix this change should be applied. Therefore, the information from the delta bond matrix has to be translated into bonding matrices. In the previous step, the delta bond matrix was applied to the bonding matrix of the first adsorbate see whether a valid delta bond matrix is found. In the next step all allowed mutations are detected. This is again achieved by applying the delta bond matrix to the bonding matrix of the starting adsorbate, but this time the process is not stopped as soon as one valid bonding matrix is obtained instead all valid matrices are stored.

For every valid bonding matrix an assignment is generated. Here, the atoms are ordered

in groups based on their local bonding at first. Then the Hungarian and Murty algorithms are used to obtain the i-th best pairings for every group. The usage of the Murty algorithm becomes necessary since the atoms are grouped based on their local equality, which could introduce insufficient assignments. A crucial part of the assignment of every group is that the initial cost matrix of every group is manipulated. Here, already assigned atoms are considered to reduce the number of possible assignments for the currently assigned atoms. As an example we assume that all carbon atoms in two structures of TBP were already assigned. For the assignment of the hydrogen atoms we can use this information so that all hydrogen atoms belonging to the same CH_3 group in the initial structure also remain at the same CH_3 group in the final structure. In this example the problem of assigning hydrogen atoms does not contain nine hydrogen atoms connected to carbon atoms, but three groups of three hydrogen atoms each connected to the same carbon atom. In a last step, the total cost of every valid bonding matrix is compared and the one with the smallest cost values is selected for the initial interpolation of the reaction path.

The simplest choice for an interpolation is the linear interpolation in cartesian coordinates. As shown in equation 3.5 the N structures $\mathbb{R}_i$ along the reaction path are obtained by adding a fraction of $\Delta\mathbb{R}$ to the initial structure $\mathbb{R}_0$. Here, $\Delta\mathbb{R}$ is simply the change of the cartesian coordinates from the initial to the final structure of the path.

$$\mathbb{R}_i = \mathbb{R}_0 + \frac{i}{N+1}\Delta\mathbb{R} \tag{3.5}$$

However, any non-linear movement in the cartesian space like for example rotations of the adsorbate or changes in bond and dihedral angles are insufficiently described. Therefore, the linear interpolation in cartesian coordinates is only used for the mostly static part of the system, which is the surface. For the adsorbate PESE is using internal coordinates. Here, in comparison to the cartesian coordinates, internal coordinates have the advantage that changes in bond length, bond angles and dihedral angles can smoothly be interpolated. In addition, every internal coordinate can be changed in the interpolation without affecting all other internal coordinates. The transformation between cartesian and internal coordinates is described in the appendix A.4. As for the cartesian coordinates the interpolation in internal coordinates is performed linearly from the starting to the final structure.

Since the interpolation in internal coordinates represents only changes of the intramolecular structure, additional interpolations considering changes relative to the surface have to be added as sketched in Figure 3.13. Here, two slightly different approaches are implemented in PESE.

In the first approach the adsorbate is assumed to be independent from the surface. This case is detected by the absence of any covalent bond between the adsorbate and the surface. In this context the covalent bond is again detected as atomic distance being smaller than 1.3 times the sum of the covalent radii. Consequently, this approach can also be applied to physisorbed structures in case this is requested by the user. Since the adsorbate is assumed to be independent from the surface, the interpolation is performed as a simultaneous interpolation of changes of the center of mass, of the orientation and of the internal coordinates of the adsorbate. The best rotation matrix describing the change of the orientation of the adsorbate between the start and final structure is obtained by the Kabsch algorithm (appendix B.2). For the interpolation the obtained rotation matrix is decomposed into individual rotation around the coordinate axes (appendix A.3), since these rotations can be partitioned. In principle, a successive interpolation of the center of mass, orientation and internal coordinates would be possible in case any knowledge about what should happen first is present.

The second approach is used if a covalent bond between the adsorbate and the surface is detected. In this case, the surface atoms bonded to the adsorbate are added to the adsorbate structure as pseudo atoms. Then the change of the orientation and internal coordinates is interpolated as in the first approach. However, instead of also interpolating the center of mass, the surface atoms of the adsorbate structure are mapped on their corresponding linear interpolation in cartesian coordinates. This approach ensures, that a reasonable interpolation of the orientation and internal coordinate of the adsorbate with respect to the surface are obtained.

Reaction Path Calculation

Once a reaction path interpolation is obtained, the reaction paths can be optimized by workers. Here, the NEB method introduced in section 2.2.4 is used. Several changes are introduced to improve the performance and tailor this method for PESE. These changes address either

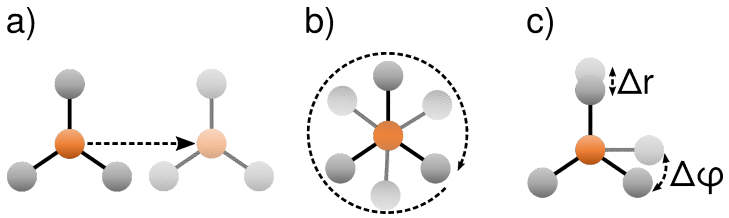


Figure 3.13.: Improved interpolation of reaction paths. The interpolation is obtained by a linear interpolation of (a) the center of mass, (b) the orientation of the adsorbate and (c) the internal coordinates of the adsorbate.

the parallelization or the NEB algorithm itself.

In the standard NEB, a parallelization over the images along the reaction path is performed to optimize a single path as fast as possible. However, this comes at the cost of significant communication overhead as already discussed in the parallelization concept of PESE. Furthermore, a fixed number of images has to be used since the number of cores used for this calculation cannot be changed during the calculation. This parallelization has another shortcoming as discussed in the literature[201]: Since the slowest converging image is determining the convergence of the path a lot of computational overhead is generated by performing optimization steps for already converged images. Therefore, for PESE I follow the approach of switching to a serial implementation of the NEB as discussed by Lindgren and coworkers[201]. An ostensible disadvantage of this approach is that a calculation of a single path is several factors slower. However, since PESE is calculating several reaction paths in parallel with restricted computing power, getting rid of the computational overhead is actually an advantage. In addition, a serial NEB is more flexible regarding the number of images and PESE is able to focus all resources on unconverged images.

Despite switching to a serial implementation of the NEB several improvements sketched in Figure 3.14 were implemented: Following the work of Lindgren and coworkers[201] the convergence criterium of every image i with the coordinates $\mathbb{R}_i$ is adjusted based on its distance to the next TS with the coordinates $\mathbb{R}_{i_{max}}$. Here, the convergence criterium f^i_{max} of image

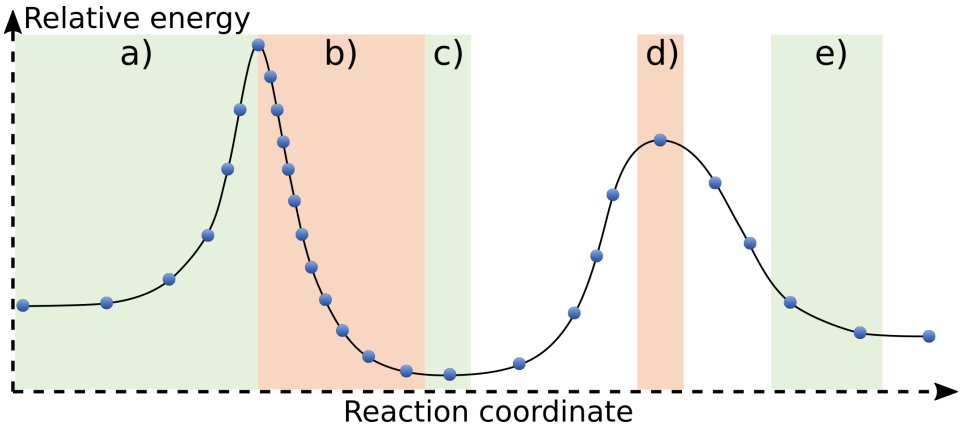


Figure 3.14.: Schematic reaction path optimized by the NEB method. Blue dots represent a structure along the reaction path connecting the starting and final structure. (a) A section of the path with a low resolution of images. (b) A section of the path with high resolution of images. (c) Minima and (d) TSs are detected based on their energy and optimized on request. (e) The convergence criterium for images distant to a TS is lowered.

i is adjusted based on the convergence criterium $f_{max}^{i_{max}}$ of the TS as shown in equation 3.6. The scaling factor α_s could be freely adjusted but is usually set to the recommended value of 2[201].

$$f_{max}^{i} = f_{max}^{i_{max}} \cdot [1 + \|\mathbb{R}_{i_{max}} - \mathbb{R}_i\| \cdot \alpha_s] \tag{3.6}$$

Furthermore, I followed the idea of Kolsbjerg and coworkers[202] to change the number of images during the calculation. However, instead of adding new images only in the area of the TS I follow for PESE the approach to define an energy and spatial resolution based on which images are introduced or removed. Since the number of images is changing during a calculation the spring force constant has to be adjusted accordingly. This is necessary to ensure that untouched images experience no change in their spring force. By using the definition of the spring force shown in equation 3.7 the two spring force constants k_{left} and k_{right} can be adjusted individually. This enables PESE to automatically increase the spring force constant in case new images are added or to decrease the constant in case images are removed.

$$\mathbb{F}_{i,\|}^{S} = [(k_{left} \cdot |\mathbb{R}_{i+1} - \mathbb{R}_i|) - (k_{right} \cdot |\mathbb{R}_i - \mathbb{R}_{i-1}|)] \cdot \hat{\tau} \tag{3.7}$$

As a final improvement, minima along the reaction path are detected based on their energy and optimized. All these extensions are in practice activated once a user defined convergence for the whole path is reached. In this way more crucial manipulations like the detection and optimization of TS and minima is only performed once the path is already decently converged. With this enhanced NEB implementation PESE can calculate the proposed diffusion and decomposition paths to obtain the entire reaction network.

4

Reactivity of Atoms and Small Molecules on Gallium Phosphide

Gallium phosphide GaP, a III-V compound semiconductor, is used as the substrate for the first applications of PESE. GaP itself is studied for applications in the field of water splitting reactions[203] or window layers for solar cells[204]. Also, due to a band gap of 2.27 eV[205] GaP is of interest as a component in LEDs. However, for optoelectronic applications the indirect bandgap of GaP[206] is limiting device performance and thereby the possible application range.

Still, GaP is heavily studied for (opto-) electronic applications whereby special attention is given to the direct integration of III-V compound semiconductors with silicon[33,48,207–211]. Here, the main hindrances for the direct combination of III-V compound semiconductors with silicon are the differences in the lattice parameters[212] and the growth of a polar material (III-V materials) on a nonpolar material (silicon)[213]. These obstacles result in dislocations[212,214] and antiphase domains[211,215], respectively. The idea to use GaP as a buffer layer originates from the property to be nearly lattice matched to silicon[206]. Therefore, for the growth of GaP on silicon only the problem of the ”polar-on-nonpolar epitaxy”[213] remains. By carefully tuning growth conditions for GaP on silicon a defect free GaP layer can be achieved[48,216–218]. Consequently, for the integration of another III-V material on top of the GaP buffer layer only the differences in lattice parameters have to be addressed by strain management[219].

As introduced, the decisive step for the integration of III-V compound semiconductors with silicon is the controlled formation of a GaP layer on silicon. However, a complex interface formation between GaP and Si(001) is observed[63] and furthermore demanding computations are already necessary to understand the complex adsorption process of GaP precursors on hydrogen passivated Si(001)[72]. Consequently, instead of focusing on the growth of GaP on a silicon substrate, the homoepitaxial growth of GaP was studied by applying PESE to the decomposition of GaH_3 and PH_3 on GaP as the first model systems. In addition, also the mobility of bismuth on GaP was studied with PESE primarily for debugging purposes. The motivation to study these systems will be introduced in the corresponding sections.

4.1. Computational Methods

Selecting the Gallium Phosphide Surface Reconstruction

For the gallium phosphide surface several reconstructions are known based on a plethora of calculations and experiments[59,220]. Based on the reaction conditions, two reconstructions for the (001) surface are most common: On the one hand, a P-rich (2×2) reconstruction is obtained by P-precursor stabilized growth of GaP. On the other hand, a Ga-rich (2×4) reconstruction is obtained without P-stabilization where phosphorus atoms are desorbing at temperatures above $490\,^{\circ}$C[220–222]. The P-rich reconstruction is made of a layer with buckled P-dimers whereby one P-atom of every dimer is bonded to an additional hydrogen atom[60]. In comparison, the Ga-rich reconstruction is made of a Ga-layer with a Ga-P dimer on top of it[221,223–227]. Therefore, this reconstruction is also called mixed-dimer reconstruction. In addition to these two common reconstructions, intermediate surface reconstructions are experimentally known and even more were theoretically predicted, although not observed under experimental conditions[58,61,221,222,224,225,227].

As a first test case for PESE the two stable surface reconstructions under either P- or Ga-rich conditions are of interest. The P-rich surface is due to the additional hydrogen atoms more challenging for PESE. These hydrogen atoms are known to be quite mobile at room temperature[228] indicating a high reactivity of these atoms. This is problematic since reactions between a precursor molecule and the hydrogen atoms are likely but not yet described by PESE. Therefore, for all studies, the simpler Ga-rich mixed-dimer reconstruction was used to study the decomposition of precursors.

The mixed-dimer reconstruction shows an interesting electronic structure[225,227,229,230]. The highest occupied surface state is located at the P-atom of the mixed-dimer and attributed to the p-orbital of the electron pair. In addition, an unoccupied surface state attributed to the empty Ga p-orbital is located at the mixed-dimer. Thereby, the electronic structure of the mixed-dimer is very similar to the electronic structure of a Si-Si dimer on a Si(001) surface[231,232]. However, the surface state based on the empty p-orbital at the mixed dimer is not the lowest unoccupied surface state. Instead, the lowest unoccupied surface state arises from dangling bonds at every Ga-atom of the topmost Ga-layer. In addition, occupied surface states are observed as σ type Ga-Ga bonds between Ga-atoms of the topmost Ga-layer as well as for Ga-Ga bonds to the mixed dimer.

Periodic DFT Calculations

Within these first applications of PESE, all periodic DFT calculations were performed with the Vienna Ab initio Simulation Package (VASP, version 5.4.4)[233–236]. In all calculations the functional by Perdew, Burke and Ernzerhof (PBE)[237,238] and the DFT-D3 dispersion correction[90,91] were used. SCF calculations were converged up to changes in electronic energy of less than 10^{-6} eV, while minima and transition states were converged up to forces smaller than 10^{-2} eV·Å^{-1}. The plane wave cutoff within the PAW approach[86,87] and the k-point grid[239] were converged based on the total energy of the GaP bulk structure, the GaP(001) slab model and the studied molecules and atoms Bi, GaH_3 and PH_3.

The Gallium Phosphide Bulk Model

For the bulk structure of GaP a plane wave basis of 400 eV and a Γ-centered Γ(555) k-point grid were chosen ensuring a residual error of less than 1.0 kJ·mol^{-1} in the total energy. Based on these settings the optimized bulk structure of GaP with a lattice parameter of 5.478 Å ($+0.5$ % with respect to the experimental parameter of 5.451 Å[240]) shown in Figure 4.1 was obtained.

Typical for III-V semiconductors, GaP shows a zinc blende crystal structure. This structure corresponds to two face-centered cubic (fcc) structures - one with Ga placed at the origin $(0, 0, 0)$ and the second with P shifted by a quarter of every lattice vector $(0.25, 0.25, 0.25)$. In Figure 4.1a the cubic unit cell comprised of the 4 gallium and 4 phosphorus atoms is shown. As indicated by the side view in Figure 4.1b, every atom within this crystal structure is showing a tetrahedral coordination by atoms of the other element. The common perspective on the

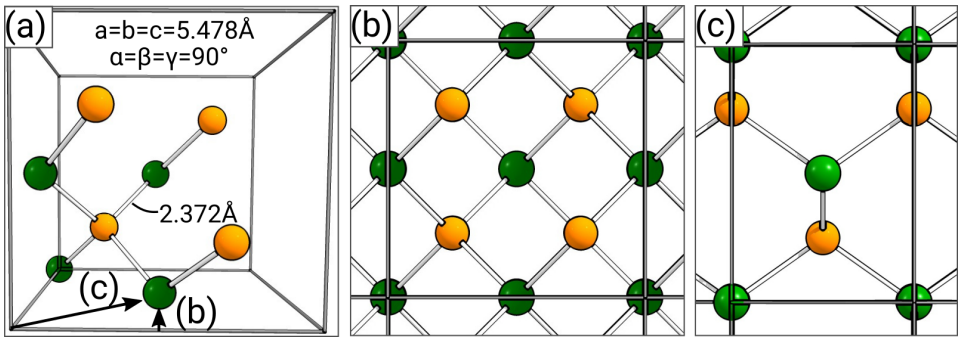


Figure 4.1.: Optimized gallium phosphide bulk structure. Optimized cubic unit cell of GaP shown in (a). Different perspectives on the cubic unit cell with PBC in (b) and (c). The perspectives of (b) and (c) are indicated in (a) by arrows. Color code: Ga (green), P (orange).

GaP structure used within this thesis is shown in Figure 4.1c. This perspective is obtained along the (-110) direction or by rotating the bulk by $45°$ around the z-axis.

The Gallium Phosphide Slab Model

Based on this optimized GaP bulk a 4×4 GaP(001) slab model in the (2×4) mixed-dimer reconstruction was created as shown in Figure 4.2. For this, the bulk structure was rotated by $45°$, accordingly duplicated in the in-plane directions and cut along the (001) direction. For this slab model the number of layers was benchmarked in the range of 3 to 13 layers. The change in total energy per added atom was compared to the most accurate estimate of adding a bulk-like layer - 13th layer (Ga) and 12th layer (P). At 5 layers the change in total energy per added atom dropped below $1\,\mathrm{kJ{\cdot}mol^{-1}}$. Consequently, the used GaP slab consists of 3 Ga-layers and 2 P-layers. In all calculations the two bottommost layers were frozen in their bulk position (Figure 4.2c, d). Furthermore, the bottommost Ga-layer was saturated

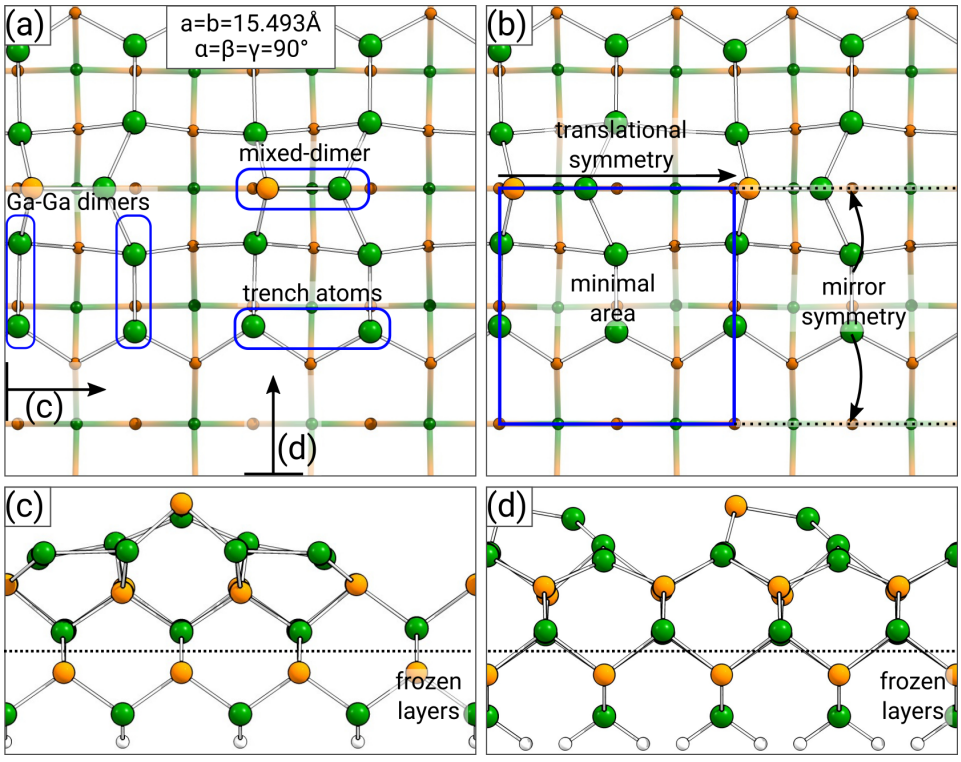


Figure 4.2.: Optimized gallium phosphide surface structure with a (2×4) mixed-dimer surface reconstruction. Top view shown in (a) and (b) and side view shown in (c) and (d). The common adsorption sites of the mixed-dimer, Ga-Ga dimers and Ga trench atoms are shown in (a). The minimal area detected by the indicated symmetry elements is shown in (b). Color code: Ga (green), P (orange).

by fractionally charged ($+1.25$ e) pseudo H atoms[241] restoring the tetrahedral coordination of the Ga atoms. The Ga-H bond lengths were set to that of gallium hydride (1.573 Å[242]). The used plane wave cutoff and k-point grid were converged to a residual error of less than 0.2 % in the total energy. Consequently, a cutoff of 350 eV and a Γ-centered Γ(221) k-point grid were used. Furthermore, 22 Å of vacuum were added to prevent spurious interactions in z-direction due to PBC.

As shown in Figure 4.2b, the GaP slab model used contains a translation symmetry element (translation by 7.7464 Å) due to the usage of a 4×4 instead of the minimal 2×4 slab. Furthermore, two mirror planes indicated by dotted lines in Figure 4.2b are present (position in x-direction: 5.810 Å and 13.556 Å). Therefore, to comprehensively study the reactivity of an adsorbate on this surface, only the marked minimal area has to be considered, as automatically done by PESE. Although only a fraction of the surface is used to investigate the reactivity of an adsorbate, the shown 4×4 slab is used instead of the minimal 2×4 slab. Similar to the introduced vacuum layer, this is motivated by preventing spurious interaction in the x- and y-direction between several adsorbates due to PBC.

Additional Settings for PESE

Here, only a summary of the chosen settings is provided. The justification of the individual settings is discussed in the next sections. The initial grid dividing the minimal area was generated with a spacing between grid points of 1.0 Å. The maximum in-plane displacement Δd to use an surface atom for the height adjustment was set to 2.2 Å while the minimal and maximum distance of the adsorbate in z-direction was set to 1.8 Å and 10.0 Å, respectively. For GaH_3 and PH_3 the number of rotation per principal axis was set to 4 in the generation of orientations while the RMSD value to identify identical orientations was set to 0.5 Å.

In the calculation of the adsorption basin an RMSD of 0.5 Å, energy threshold of 4.0 kJ·mol^{-1} and a factor of 1.0 for the displacement along the forces were used to identify an already known structure. For the final projection, a final grid with a spacing of 0.2 Å was generated and all Gaussian functions were cut at 1.0 % of their maximum value.

For the removal of duplicates an RMSD value of 0.5 Å (1.0 Å) and an energy threshold of 4.0 kJ·mol^{-1} (10 kJ·mol^{-1}) was used for Bi (GaH_3 and PH_3). For Bi (GaH_3 and PH_3) chemisorbed minima were detected based on a distance of less than 1.3 (1.2) times the sum of the covalent radii while physisorbed structures were recognized based on a distance of less than 2.5 times the sum of the covalent radii. In addition, an electronic contribution to the adsorption energy of less than -10 kJ·mol^{-1} was requested for chemisorbed minima.

For the generation of decomposition structures only the rules for homolytic bond cleavage and three center reactions were considered. Furthermore, only chemisorbed structures were considered as starting points. The grid for the decomposition structures was generated within an angle of 45° and a radius of 2.5 Å starting at a minimal radius of 0.0 Å. This grid was also generated with a spacing of 1.0 Å between the grid points.

Diffusion paths were generated only for a minimal network (λ set to 2.0). In case the start and end point differ by more than 30 Å or 400 kJ·mol^{-1} a path was not generated at all. For Bi only diffusion paths between chemisorbed minima were considered, while for GaH_3 and PH_3 also paths from a physisorbed or gas phase minimum to a chemisorbed minimum were considered.

All reaction paths for Bi (GaH_3 and PH_3) were generated with an initial spacing between the images of 1.5 Å (2.0 Å). Three convergence levels of 0.100, 0.055 and 0.010 eV·Å^{-1} were used. The optimization of the reaction paths was performed by PESE's implementation of the NEB[99,100] method. Here, forces were only calculated for unconverged images. Furthermore, the number of images was updated and the optimization of minima was performed once the first convergence level was reached. The convergence criterium was adjusted by $\alpha_s = 2$[201] and transition states were optimized once the second convergence level was reached. The number of images was changed to maintain a resolution of at least 50 kJ·mol^{-1} and 1.5 Å. For reaction paths of Bi unconverged paths were restarted with an initial spacing between the images of 0.75 Å or an even increased resolution of 0.55 Å to identify the influence of the resolution on the convergence. The spring force constant was set to 2.5 eV·Å^{-2} for Bi and to 5.0 eV·Å^{-2} for GaH_3 and PH_3.

4.2. Adsorption and Diffusion of Bismuth on Gallium Phosphide

The adsorption and diffusion of bismuth on GaP was studied with PESE as a first model system. The main motivation to study the most simple system of a single atom was to debug every routine of PESE. However, bismuth is also highly relevant in applications of III-V semiconductors like lasers, solar cells or photodetectors[243–246].

The interest in bismuth for optoelectronic applications is based on the desire to tune the band gap of the composite material. For example in GaAsBi an incorporation of bismuth is reducing the band gap[247–249] due to the interaction of the bismuth defect state and the valence band state[250]. Another interesting effect is the increased spin-orbit splitting[251–253]. This effect can improve the device efficiency since problematic energy loss processes such as auger recombination are suppressed[49,209]. However, to completely suppress this loss channel bismuth incorporations of $10\,\%$ or more are needed[252,253] while only amounts in the range of a few percentage[254–256] and up to $8.0\,\%$[209] are reported.

Within this group of bismuth containing composite materials, Ga(P,Bi) is of interest due to its potential as a laser for telecommunication[209,257–259]. Consequently, theoretical studies addressing the electronic properties exist[257,258,260]. Furthermore, a great advantage of Ga(P,Bi) is that it can be grown on GaP[209], which itself can be grown lattice matched on silicon[261]. Thereby, a direct integration of a Ga(P,Bi) laser on silicon is accessible.

Instead of studying the Ga(P,Bi) composite material, PESE was used to focus on the adsorption structures and mobility of bismuth on GaP. Still, this information is crucial for understanding the behavior of bismuth in the growth of Ga(P,Bi) on top of the GaP layer. In addition, the mobility of bismuth is of great interest to prevent the formation of bismuth clusters[49]. Similar studies addressing the adsorption of group V atoms on III-V surfaces in general[262,263] or bismuth explicitly[262,264–267] are available although commonly the (110) surfaces have been addressed[262–264].

Choosing the Grid Settings

To study the adsorption of bismuth on GaP using PESE several keywords had to be set. As explained in section 3.2.1 a grid dividing the minimal area is used to generate the initial structures. Two important keywords are linked to the process of generating this grid. The first keyword controls the spacing between grid points and thereby also the density of grid points.

In addition, the second keyword defines the maximum in-plane displacement Δd (Figure 3.5c) between adsorbate and surface atoms. This displacement is used for the height adjustment of the adsorbate.

In Figure 4.3 the total number of grid points and their lateral position controlled by the spacing keyword are shown. This keyword expects the desired distance between two grid points in Å. However, to obtain the actual number of grid points and their spacing, the size of the minimal area is divided by the chosen value whereby the result is rounded up and increased by one. The so obtained number of grid points define the actual spacing. In this way, PESE can place a grid point at every corner of the minimal area and equally in between the minimal area. Consequently, it is possible that with a keyword of 2.0 Å (Figure 4.3d) and 2.5 Å (Figure 4.3e) the exact same grid is obtained at the end. Overall, this formula ensures that the actual density is always equal or slightly larger than specified by the keyword, while the corners of the minimal area are exactly sampled by the grid points.

The selection of the grid is a crucial step in the usage of PESE. A too dense grid (Figure 4.3a) would result in a high number of calculations while choosing a too coarse grid (Figure 4.3f) increases the possibility to miss a relevant adsorption structure. Consequently, for the selection of this keyword some basic knowledge concerning the reactivity of the studied surface is

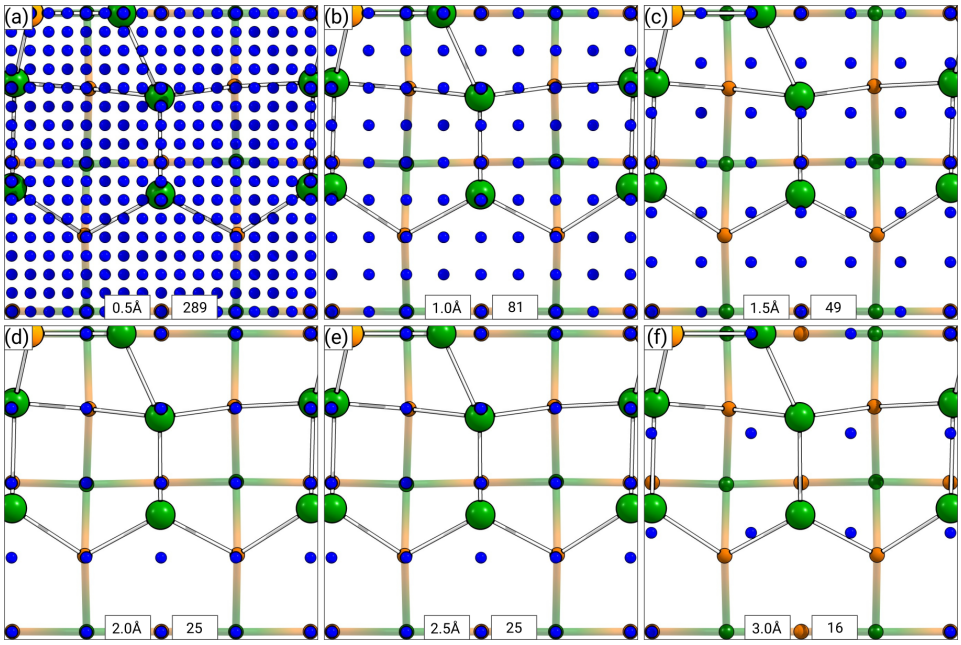


Figure 4.3.: Influence of spacing setting on the number of grid points with Bi as adsorbate. Total number of grid points and their lateral position within the minimal area are shown for several settings of the spacing keyword: (a) 0.5 Å, (b) 1.0 Å, (c) 1.5 Å, (d) 2.0 Å, (e) 2.5 Å, (f) 3.0 Å. Color code: Ga (green), P (orange).

helpful. Alternatively, a dense grid should be selected in case the surface is studied with PESE for the first time.

In adsorption studies of acetylene and ethylene on Si(001), it was found that these molecules do not only interact with the topmost surface layer of the semiconductor, but are also able to insert into a bond to the sublayer[232]. To enable PESE to find similar structural motifs on the studied GaP surface, the spacing keyword was set to 1.0 Å for all investigations of Bi, GaH_3 and PH_3. As shown in Figure 4.3b this setting results in grid points on top of every atom in the topmost layer and sublayer. This ensures that structures described by the formation of dative bonds, a common motif for group 13 and 15 compounds[268,269], can be obtained. Furthermore, grid points on top of all bonds within the top layer and from the top layer to the sublayer are present. This ensures that structures resulting from the bond insertion of the adsorbate can be obtained.

For the calculation of the adsorption basins this layer of 81 grid points is stacked in z-direction. To completely sample the adsorption basins the number of layers should be selected in a way that the highest layer of points results in structures showing minor to no interaction between the adsorbate and the surface. Since this distance is unknown in advance the maximum distance to the surface is set with 10.0 Å to a very large value. As discussed later, this choice

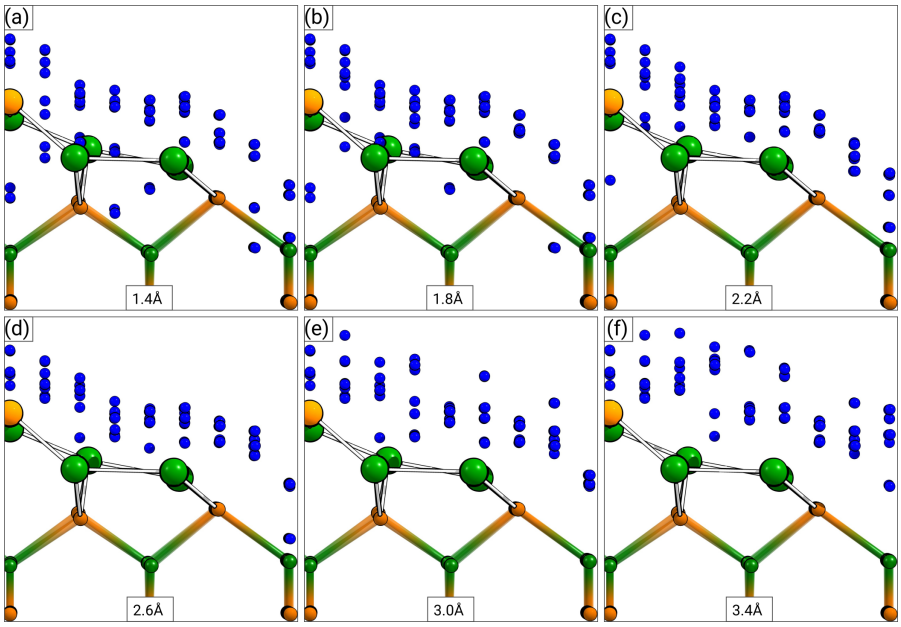


Figure 4.4.: Influence of the displacement keyword on the position of the grid points with Bi as adsorbate. The positions of the grid points are shown for settings of (a) 1.4 Å, (b) 1.8 Å, (c) 2.2 Å, (d) 2.6 Å, (e) 3.0 Å and (f) 3.4 Å. Color code: Ga (green), P (orange).

was sufficient to sample the whole adsorption basin for every studied molecule. Besides the largest distance to the surface also the smallest distance of the grid points to the surface is of relevance. Here, this distance should be smaller than the distance between an actual chemisorbed adsorbate and the surface. The placement of the adsorbate rather close to the surface is necessary since the distance at which a transition from chemisorbed to physisorbed structures takes place is unknown. Therefore, the distance between the first layer of grid points and the surface was set to 1.8 Å, which corresponds to roughly 75 % of the GaP bond length observed at the GaP-dimer. The interlayer distance between the grid points is for simplicity controlled by the same keyword as the intralayer distance and consequently set to 1.0 Å. As a result 10 layers and a uniform distribution of 810 grid points in total is obtained. Still, the small interlayer distance is for the computational performance of PESE unproblematic since the optimization of a structure is terminated as soon as a known structure is observed (see section 3.2.2).

In Figure 4.4 the effect of the second user defined keyword, the displacement in xy-direction between surface and adsorbate atoms, is shown. This keyword is decisive to ensure that the grid points follow the shape of the surface. In case the allowed displacement is too small (Figure 4.4a,b) smaller adsorbates or a single atom like Bi is penetrating the topmost surface layer since no atom is directly beneath the grid point. Especially for a setting of 1.4 Å (Figure 4.4a) this effect is clearly visible. By increasing the value of this keyword PESE can in

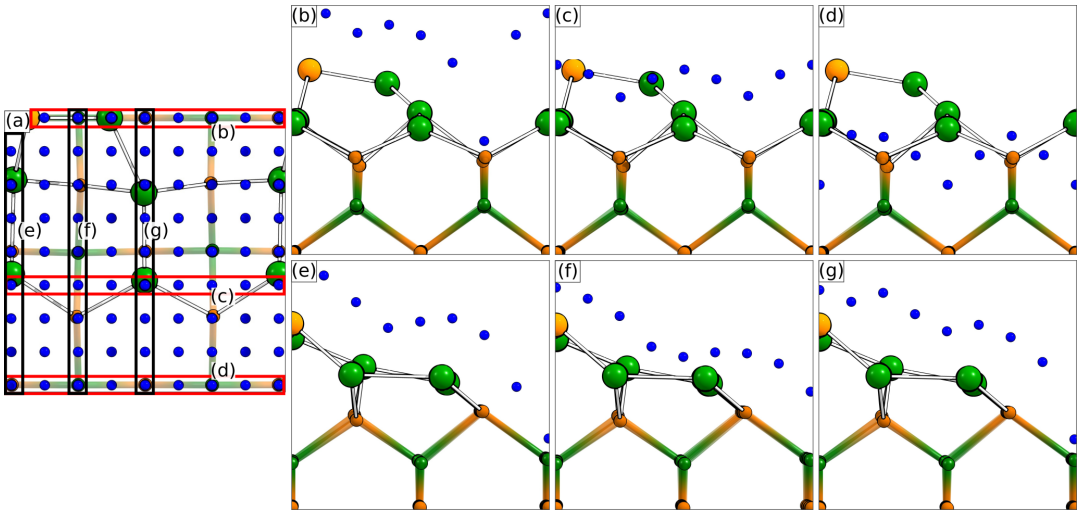
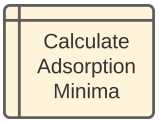


Figure 4.5.: Final grid used in all calculations of PESE with Bi as the adsorbate. A top view on the final grid with horizontal sections marked in red and black in (a). Horizontal red sections are shown in (b) to (d) and vertical black sections are shown in (e) to (g). Color code: Ga (green), P (orange).

principle use more surface atoms to detect the minimal distance between the adsorbate and surface and to adjust this distance to the desired value. In the end this results in a smoother height difference between neighboring grid points and thereby prevents that grid points are penetrating the topmost layer. However, if this value is significantly larger than the chosen minimal height of 1.8 Å (Figure 4.4e, f) PESE has to increase the minimal distance to obtain non-imaginary values for the height (see equation 3.1). Consequently, the grid points are no longer following the shape of the surface.

In the present calculations, a setting of 2.2 Å was chosen as a compromise. As shown in Figure 4.5 it is nicely visible that PESE is able to place the grid points in a way that they resemble the actual surface shape. While choosing the best value for this user defined keyword is nontrivial, the effect of this keyword can easily be tested by the user prior to start computationally demanding DFT calculations since the generation of the initial structures and thereby the grid can be performed even on a single core within a few seconds.

Adsorption Minima



In the optimization of the 810 initial structures, 241 minima structures were observed by a force of less than 0.01 eV·Å^{-1}. These minima were controlled regarding duplicates. For this, an RMSD and energy threshold had to be chosen. To find the optimal value of these settings, PESE allows to perform this step separately from all other calculations. In this way, the user can get an idea of how the keywords influence the selection of unique minima by visually inspecting the outcome. For Bi on GaP, the settings of 0.5 Å and 4.0 kJ·mol^{-1} were selected.

After the removal of duplicates, 204 structures were obtained as unique minima. With RMSDs and energy values of less than 0.075 Å and 0.1 kJ·mol^{-1}, respectively, all duplicates were detected with deviations significantly smaller than the allowed thresholds. This shows that for Bi on GaP the removal of duplicates is for a large range non-sensitive to the selected keywords. Furthermore, the minor deviations to unique minimum structures show that only very similar structures were removed.

Within this set of 204 unique minima 14 were attributed to a chemisorbed structure based on their distance to the surface. All other minima were assigned to gas-phase structures. These minima should not be seen as adsorption minima since the non-existent forces are a consequence of the large distance between the adsorbate and the surface. However, the observation of these structures clearly shows that the initial maximal distance between the topmost grid points and the surface was sufficiently large.

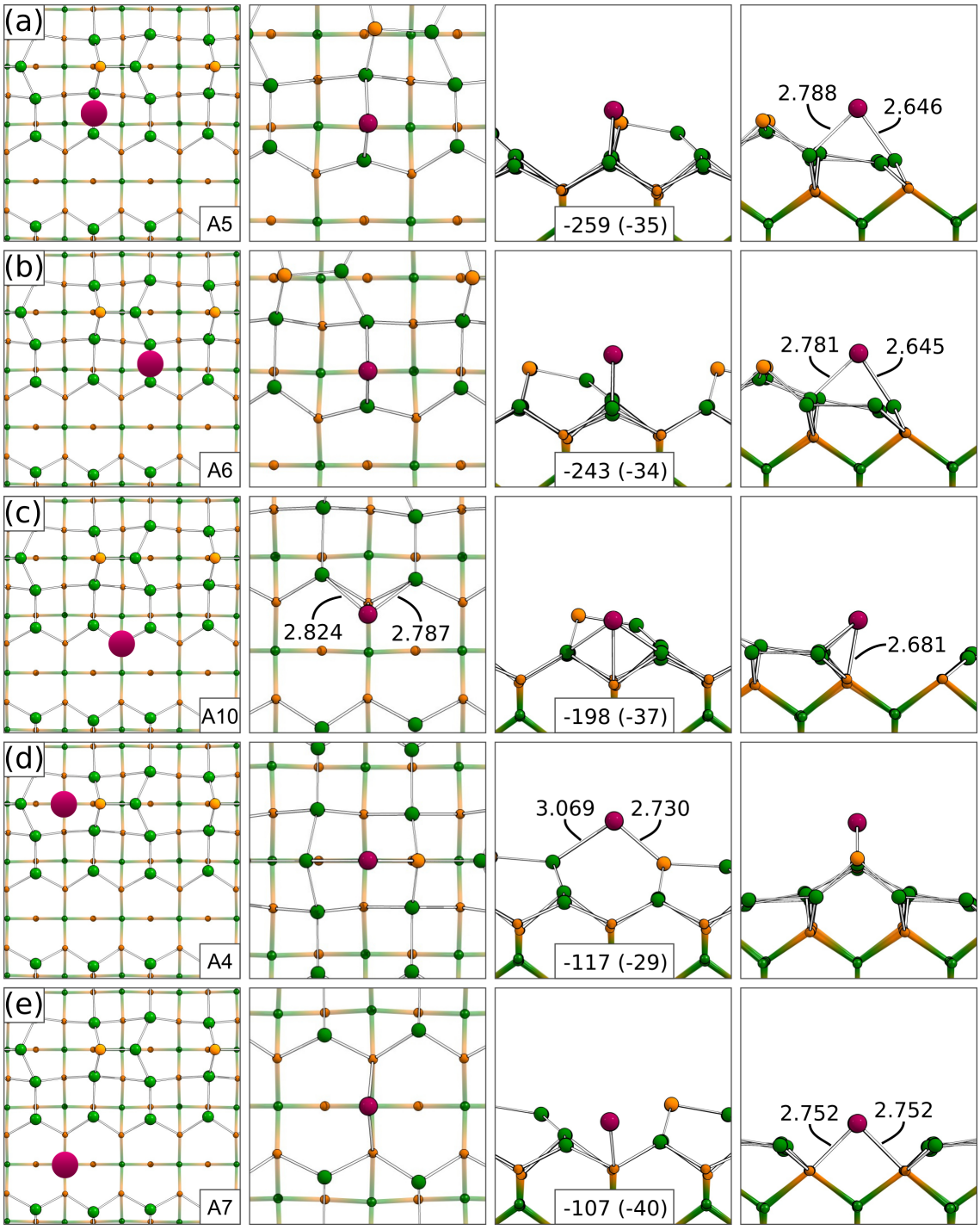


Figure 4.6.: Top and side view on adsorption structures of Bi on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima (”A”) follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), Bi (purple).

A selection of the chemisorbed minima is shown in Figure 4.6 while all other chemisorbed minima are contained in the appendix (Figure C.2 and C.3). As shown in Figure C.2a and C.2b the most stable adsorption structures for bismuth are obtained for an insertion in the Ga-Ga dimer bond. Here, adsorption energies of -259 and -243 kJ·mol^{-1} are observed, respectively. These structures are asymmetric in the sense that bismuth shows slightly longer Bi-Ga bonds towards the mixed-dimer. With adsorption energies of -117 and -107 kJ·mol^{-1}, the least stable structures are observed in case bismuth is bonding to the P-atom of the mixed-dimer (Figure 4.6d) or to P-atoms of the sublayer (Figure 4.6e). Surprisingly, the observed Bi-P bond lengths are with 2.681 Å (Figure 4.6c) to 2.752 Å (Figure 4.6e) in a similar range as the observed Bi-Ga bonds.

With -116 to -253 kJ·mol^{-1}, very similar adsorption energies were also observed for bismuth on Si(001)[270]. These high adsorption energies emphasise the high reactivity of single bismuth atoms, which is attributed to their electronic structure and quartet ground state.

Adsorption Basins

In the calculation of the adsorption basins, $17\,206$ structures were stored in the database. The raw information of these structures is displayed in Figure 4.7. In Figure 4.7a the assignment of every data point to a minimum is shown. Here, it is nicely visible that all data points in close proximity to a final minimum feature the same color as the minimum. Thereby, the adsorption basins are already visible. Also, due to the color distribution it is observable that the adsorption basins clearly differ in size. In Figure 4.7b the adsorption energy of every data point is shown. As intended by the selection of the minimal height, data points most closely to the surface are labeled in red indicating less stable structures. With increasing distance to the surface, data points turn blue showing the favorable adsorption height of bismuth on GaP(001). By further increasing the distance to the surface the data points turn white showing the loss of interaction between the bismuth and the GaP surface. However, at higher distances the data points turn red again. Here, a classic example of the known limitations[271] of DFT is observed. By increasing the distance between Bi and the GaP surface the spin state of the system is not correctly described since fractional spin states are favored[271]. This results in too large total energies and thereby adsorption energies. Nonetheless, since this problem occurs mainly for larger distances[271] and the forces are only in their magnitude overestimated but correct in their direction, the adsorption basins are unaffected as visible in Figure 4.7a. The dispersion contributions of every data point are shown in Figure 4.7c. Here, at very large distances to the surface no interaction (red points) is observed while attractive interaction

(blue points) are visible the closer the Bi gets to the surface. A similar trend is observed for the forces (Figure 4.7d, e). Here, no forces are present at large distances between the Bi and the surface (blue points). By decreasing the distance, the forces increase (red points). However, after reaching an inflection point above the position of the minima the forces decrease again before the minimum structures (blue points) are reached. If the distance is further decreased the forces increase again indicating the repulsion between Bi and the GaP surface.

To estimate the size of every basin, the data of Figure 4.7a is mapped on the final grid. Here, the important keyword is the width of the Gaussian function, which is used to smear the data points. The motivation to smear the data points is to fill the whole space of the unit cell. As shown by Figure 4.7a this is not achieved by using the raw data. A suitable selection of the smearing keyword therefore aims at filling the whole space without going beyond that. In

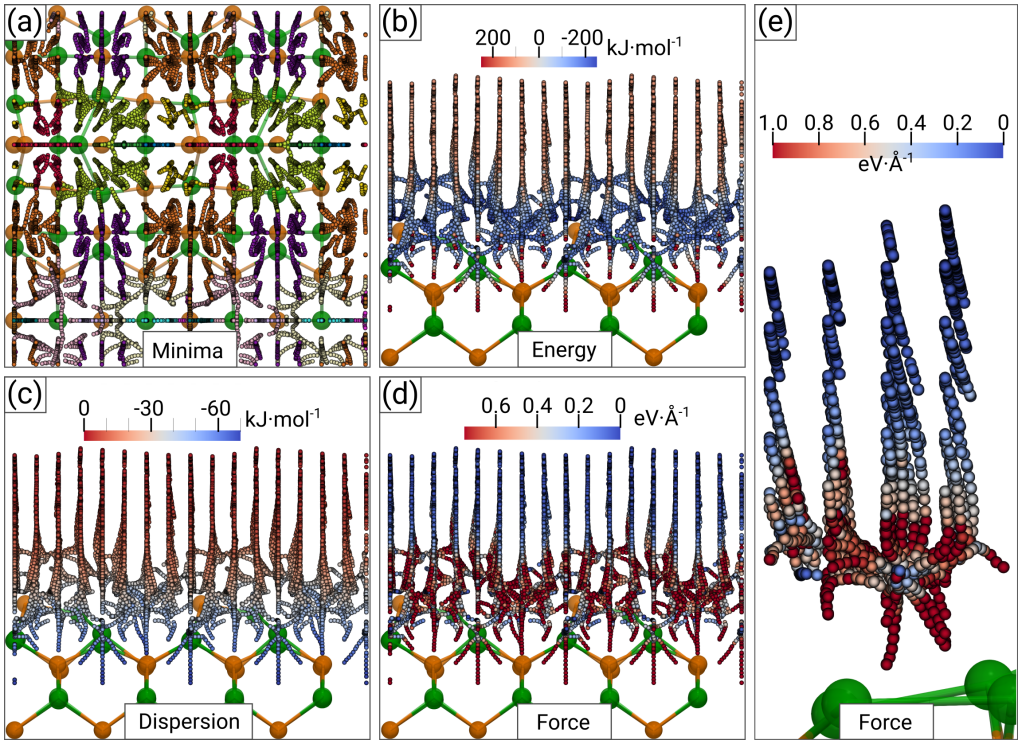


Figure 4.7.: Initial data of the adsorption basins of bismuth on GaP(001). Every small point represents an individual database structure. The position of the point corresponds to the position of the Bi atom in this structure. The assignment of every data point to a minimum is indicated in (a) by the color. The adsorption energy of every data point is shown in (b) while the dispersion interaction is shown in (c). The magnitude of the force acting on the Bi atom is shown in (d) with an enlarged excerpt showing data points corresponding to a single minimum in (e). Color code: Ga (green), P (orange).

Figure 4.8 different settings of the width keyword ω are shown with the range of smearing stated in grid points n. The keyword itself states the fraction of the selected spacing between the initial grid points and is thereby independent from the used initial grid. However, as discussed later, the smearing must be adjusted to the studied system. In all visualizations, the final grid points are only colored if a basin is contributing with at least $1\,\%$ while the actual color is based on the basin contributing the most. As shown in Figure 4.8a, a too small smearing results in the space of the unit cell not being filled. By increasing the keyword, the space is slowly filled as indicated by the coloring of the final grid points. At a value of $\omega = 0.20$ (Figure 4.8d) the whole space of the unit cell is attributed to a basin.

Further increasing the smearing has two effects: Data points at the bottom and top of the sampled area start to smear out in the positive and negative z-direction. Both is unreasonable since Bi is neither penetrating the surface nor being attracted at all to a basin at high distances to the surface. The second effect is that different basins start to override each other in the visualization. However, this is unproblematic since the actual size of the basin is only slightly effected as shown by Figure C.1. Overall, a reasonable value of the smearing keywords ensures that the obtained data points represent the whole area of the unit cell while being as small as possible. Consequently, for the evaluation and visualization of the adsorption basins of Bi on GaP(001) a value of $\omega = 0.20$ was selected.

For the five minima shown in Figure 4.6, the final basins are shown in Figure 4.9. As expected from the raw data, it is visible that grid points in close proximity to a minimum are also attributed to this minimum. Furthermore, the basins are laterally restricted: They compete with all other adsorption basins in the in-plane directions and are in the z-direction enclosed by the surface and the distance, for which the adsorbate-surface interactions vanish. The observation that the basins are reaching to the gas phase indicates that no physisorbed structures are present as expected for bismuth as the adsorbate. This is of interest for the question[192] whether a direct adsorption to a certain minimum is possible. Here, for all basins reaching to the gas phase, a direct adsorption can be expected. However, if a basin is not reaching to the gas phase (Figure C.4d), a barrier in the adsorption is present since the adsorbate has to change the basin which corresponds to moving in a direction not matching the actual forces.

The largest basin with a relative size of $31.1\,\%$ to the cumulated size of all basins was found for the most stable minimum (Figure 4.9a). This is a surprising observation since it means that in $69\,\%$ of all adsorption attempts the adsorbate would initially adsorb in a less stable adsorption minimum. Consequently, only in case the temperature is high enough, enabling the diffusion of Bi, the majority of Bi atoms can be expected in the most stable position. Furthermore,

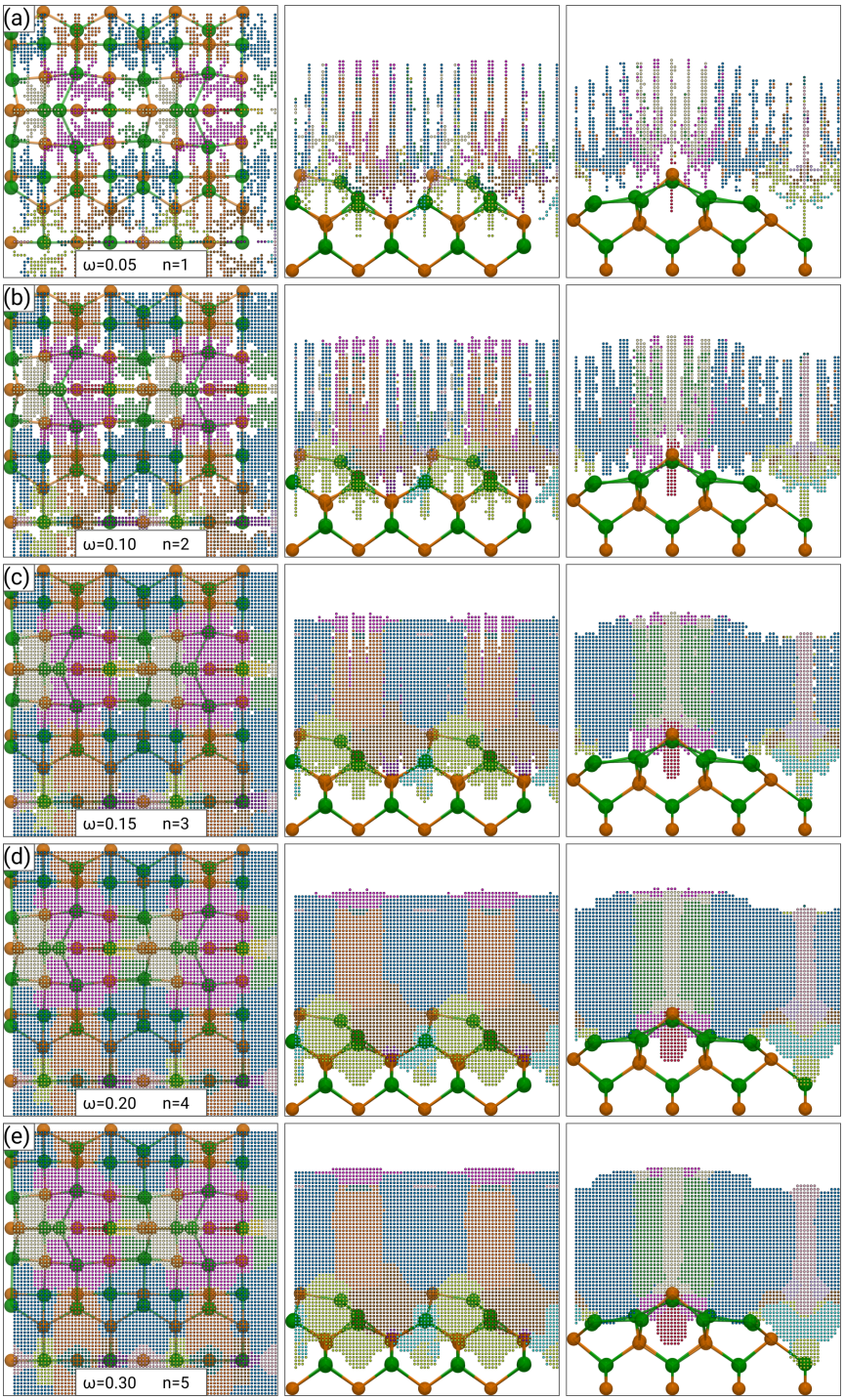


Figure 4.8.: Influence of the smearing keyword ω on the final adsorption basin mapping for Bi on GaP(001). Every final grid point, which is assigned by more than $1\,\%$ to a minimum, is shown as a small point in the color of the minimum. For ω settings of (a) $\omega = 0.05$, (b) $\omega = 0.10$, (c) $\omega = 0.15$, (d) $\omega = 0.20$ and (e) $\omega = 0.30$ are shown. In addition, the range of the smearing is stated in terms of final grid points n. Color code: Ga (green), P (orange).

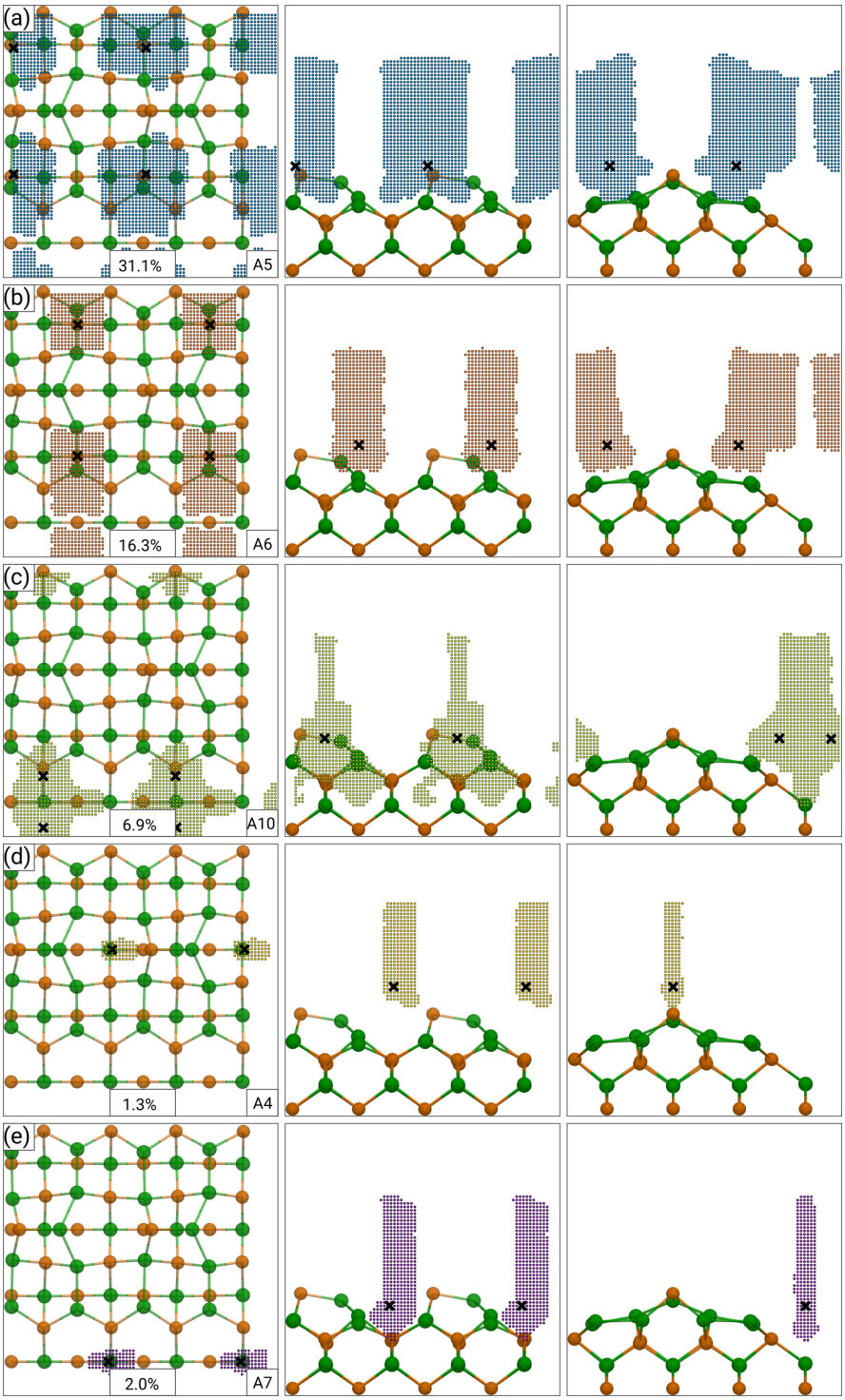


Figure 4.9.: Top and side view on adsorption basins of Bi on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than 1 % assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange).

this observation shows that studying only the most stable structure for the decomposition of larger precursors might not be sufficient and in fact wrong in case the kinetics are mainly defined by a combination of several less stable minima. This will especially hold true if the decomposition reactions are in competition with diffusion reactions.

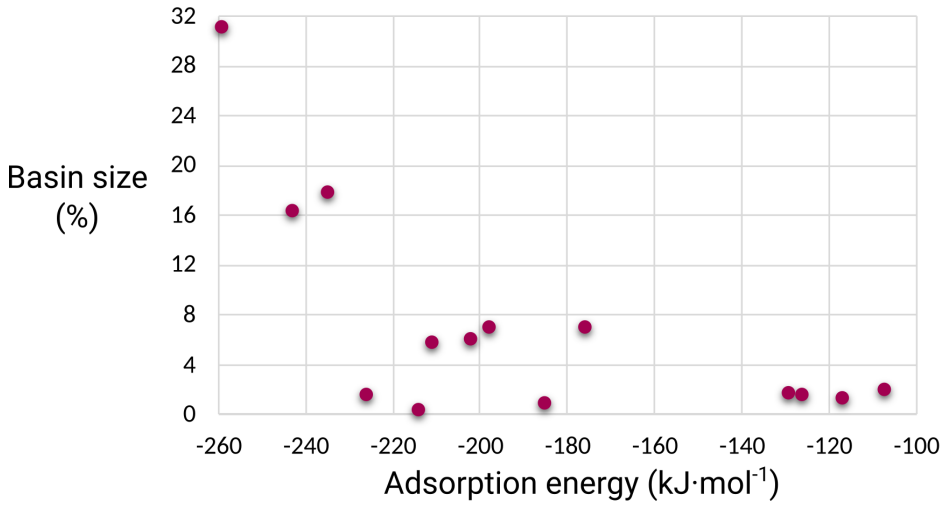
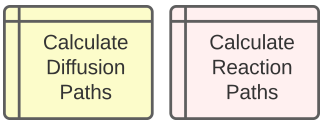


Figure 4.10.: Correlation between the basin size and adsorption energy of Bi on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in % and the adsorption energy of the assigned minimum in kJ·mol^{-1}.

The observation that the largest basin is present for the most stable structure (Figure 4.9a) raises the question whether the size of a basin, and thereby the geometric probability to adsorb in a certain minimum, is related to the adsorption energy. To address this question for the system Bi on GaP, the correlation between basin size and adsorption energy is shown in Figure 4.10. Here, a relation between basin size and adsorption energy could be assumed for the most ($\Delta E_{Ads} < -230$ kJ·mol^{-1}) and least ($\Delta E_{Ads} > -140$ kJ·mol^{-1}) stable structures. However, as indicated by the intermediate structures, no correlation between the basin size and the adsorption energy is present. Since for this system no clear evidence for or against a correlation between basin size and adsorption energy is observed, this question will be addressed in more detail with the larger datasets of GaH_3 and PH_3.

Diffusion Paths



Based on all obtained adsorption minima, PESE proposed 20 diffusion reactions whereby 19 were successfully optimized. A general example path is sketched in Figure 4.11 explaining

the representation of the final network overview. The primary goal of PESE is to find transition states along the reaction path to estimate the reaction barrier. Additional minima are observed in case a direct reaction between the selected starting and endpoint is possible. For the final network overview of Bi on GaP(001), all adsorption minima ("A") obtained by the calculation of the adsorption basin are labeled in blue while additional path minima ("P") in orange. As also indicated by the path in Figure 4.11 it is possible that a former adsorption minimum turns out to be a transition state. This is possible in case the PES is too flat in some regions leading to forces dropping below the convergence threshold. However, since the NEB exhibits a higher resolution than the optimization of structures, these wrong minima can be detected. Such minima are indicated by the red border in the final representation. Furthermore, the size of the nodes represents their adsorption energy whereby larger nodes represent more favorable structures. The activation energy for a reaction step is encoded in the size of the arrow connecting two nodes by increasing the thickness for decreasing barriers.

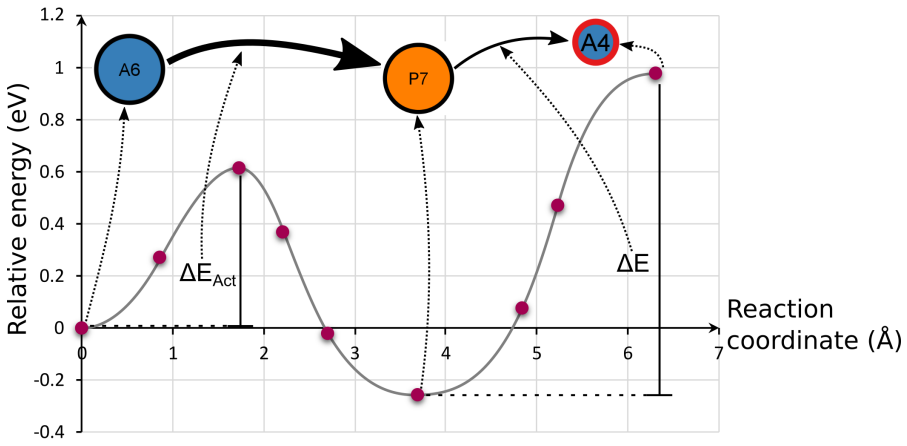


Figure 4.11.: Example diffusion path of Bi on GaP(001). The representation of the final reaction network by nodes (circles) and edges (arrows) is sketched. Minima are shown as circles whereby the adsorption energies define the size of the circles. Reaction steps are shown as arrows whereby the activation energies define the thickness of the arrows. TSs, which were initially identified as minima, are labeled by a red border.

The final data of the diffusion network for Bi on GaP are shown in Figure 4.12 and Table 4.1. In this visualization the internal nomenclature of PESE is used, which is also shown in the corresponding structures of Figures 4.6. In Figure 4.12 the most stable structures A5, A6 and A9 are identified by the largest nodes. As indicated by the arrows, all three structures are favored in terms of kinetics: Thicker arrows are pointing to them while only thin arrows are pointing in an outgoing direction. This indicates that smaller barriers must be overcome to reach these

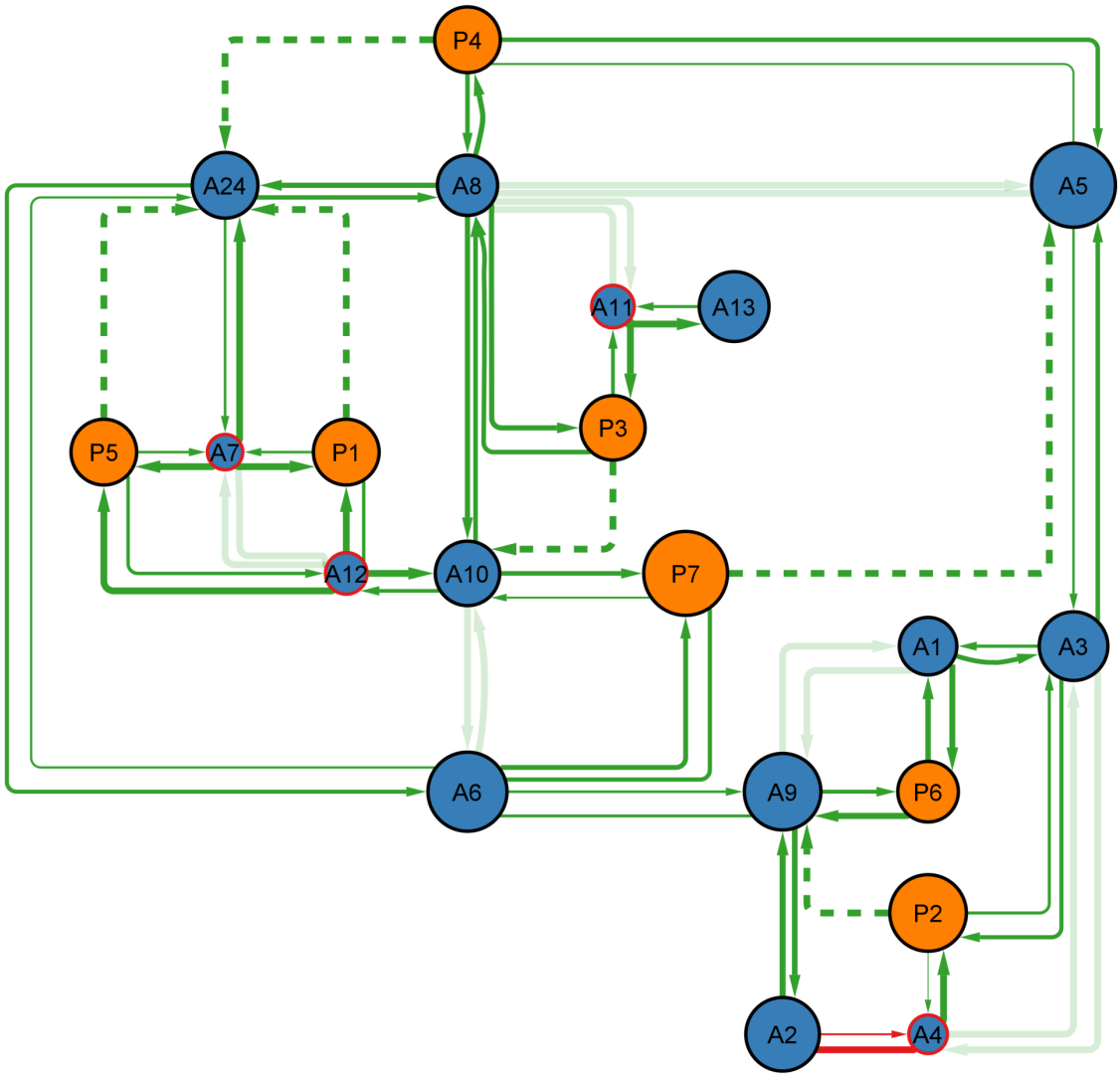


Figure 4.12.: Diffusion network of Bi on GaP(001). Representation of the minima as nodes (circles; blue: adsorption minimum, orange: path minimum) and reactions as edges (arrows). Size of the nodes represents the adsorption energy while thickness of the edges represents the activation energy. Larger nodes and thicker edges represent more favorable structures and reactions. Initial paths, which contained path minima, are shown slightly transparent. Unconverged reaction paths are shown in red. A dotted line represents a path minimum, which is a duplicate of an already known minimum.

Table 4.1.: All values of the reaction network. For every path the source node (S) is stated with its adsorption energy ΔE_{Ads} and basin size (Basin) while the target node (T) is state with its adsorption energy. Furthermore, the reaction energy ΔE_{React} and the reaction barrier for the forward ΔE_{Act}^{for} and reversed ΔE_{Act}^{rev} reaction is stated. In addition, information whether the reaction path calculation is converged (Conv., T: True or F: False) and whether no additional minima were found and the reaction path is therefore a final path (FP) are stated. Activation energies are only stated for converged and final paths.

S	ΔE_{Ads}	Basin	T	ΔE_{Ads}	ΔE_{React}	Conv.	FP	ΔE_{Act}^{for}	ΔE_{Act}^{rev}
A1	−175.9	6.9	A3	−210.8	−34.9	T	T	45.2	80.0
A1	−175.9	6.9	A9	−235.4	−59.5	T	F	—	—
A1	−175.9	6.9	P6	−185.4	−9.5	T	T	20.0	29.5
A2	−226.2	1.5	A4	−116.6	109.6	F	T	109.6	0.0
A2	−226.2	1.5	A9	−235.4	−9.2	T	T	17.3	26.5
A3	−210.8	5.7	A1	−175.9	34.9	T	T	80.0	45.2
A3	−210.8	5.7	A4	−116.6	94.2	T	F	—	—
A3	−210.8	5.7	A5	−258.5	−47.7	T	T	50.5	98.2
A3	−210.8	5.7	P2	−235.4	−24.7	T	T	59.4	84.0
A5	−258.5	31.1	A3	−210.8	47.7	T	T	98.2	50.5
A5	−258.5	31.1	A8	−184.5	74.0	T	F	—	—
A5	−258.5	31.1	P4	−201.5	56.9	T	T	108.3	51.4
A6	−243.4	16.3	A9	−235.4	8.0	T	T	94.6	86.6
A6	−243.4	16.3	A10	−197.5	45.8	T	F	—	—
A6	−243.4	16.3	A24	−201.5	41.8	T	T	102.3	60.5
A6	−243.4	16.3	P7	−258.5	−15.1	T	T	42.6	57.7
A7	−106.6	2.0	A12	−126.1	−19.4	T	F	—	—
A7	−106.6	2.0	A24	−201.5	−94.9	T	T	0.0	94.9
A7	−106.6	2.0	P1	−201.5	−94.9	T	T	0.0	94.9
A7	−106.6	2.0	P5	−201.5	−94.9	T	T	0.0	94.9
A8	−184.5	0.9	A10	−197.5	−13.1	T	T	36.5	49.5
A8	−184.5	0.9	A11	−129.0	55.5	T	F	—	—
A8	−184.5	0.9	A24	−201.5	−17.1	T	T	35.0	52.1
A8	−184.5	0.9	P3	−197.5	−13.0	T	T	36.5	49.5
A10	−197.5	6.9	A12	−126.1	71.5	T	T	71.5	0.0
A11	−129.0	1.7	A13	−214.3	−85.3	T	T	0.0	85.3
P1	−201.5	0.0	A12	−126.1	75.5	T	T	75.5	0.0
P1	−201.5	0.0	A24	−201.5	0.0	T	T	0.0	0.0
P2	−235.4	0.0	A4	−116.6	118.8	T	T	118.8	0.0
P2	−235.4	0.0	A9	−235.4	0.0	T	T	0.0	0.0
P3	−197.5	0.0	A10	−197.5	0.0	T	T	0.0	0.0
P3	−197.5	0.0	A11	−129.0	68.5	T	T	68.5	0.0
P4	−201.5	0.0	A8	−184.5	17.1	T	T	52.1	35.0
P4	−201.5	0.0	A24	−201.5	0.0	T	T	0.0	0.0
P5	−201.5	0.0	A12	−126.1	75.5	T	T	75.5	0.0
P5	−201.5	0.0	A24	−201.5	0.0	T	T	0.0	0.0
P6	−185.4	0.0	A9	−235.4	−50.0	T	T	13.8	63.8
P7	−258.5	0.0	A5	−258.5	0.0	T	T	0.0	0.0
P7	−258.5	0.0	A10	−197.5	60.9	T	T	109.8	48.9

structures while larger barriers are needed to leave them. All three structures have the structural motif of Bi forming two bonds to Ga atoms in common. Structures, in which Bi is forming bonds to a P atom, are now identified as transition states (A4, A7, A11) or tend to transform to another minimum directly or over several steps (A1, A3, A8, A10, P6). Interestingly, with structure A13 (Figure C.5c) an isolated structure is found since a barrier of $+85\,\mathrm{kJ{\cdot}mol^{-1}}$ has to be overcome to leave this structure. As shown by Figure C.5c, this structure represents Bi adsorbing in the trench between Ga-atoms of the topmost layer. Consequently, some activation energy is necessary to hop back on top of the Ga-layer.

Overall, all diffusion barriers are in a range of $+14\,\mathrm{kJ{\cdot}mol^{-1}}$ to $+110\,\mathrm{kJ{\cdot}mol^{-1}}$. These barriers are in a very similar range as for the diffusion of Bi on Si(001) where barriers of up to $+127\,\mathrm{kJ{\cdot}mol^{-1}}$ were found.[270] The barriers can be used to estimate a temperature at which the diffusion can be observed by assuming a prefactor of the Arrhenius equation of $10^{13}\,\mathrm{s^{-1}}$[270]. For Bi on GaP(001), a temperature of $301\,^{\circ}\mathrm{C}$ already results in reaction rates of at least $10^{3}\,\mathrm{s^{-1}}$. Therefore, the diffusion of Bi on the GaP surface is promoted at lower temperature as the diffusion within the GaP bulk. Here, temperatures of $600\,^{\circ}\mathrm{C}$[49] are necessary. This observation is also in line with studies to the diffusion of Bi within the silicon bulk[272].

Conclusion

In the first study using PESE the adsorption and diffusion of Bi on GaP(001) was investigated. In total, 11 chemisorbed minima with adsorption energies of up to $-259\,\mathrm{kJ{\cdot}mol^{-1}}$ were found. The adsorption basins were constructed based on $17\,206$ data points. Here, the largest basin was obtained with $31\,\%$ for the most stable structure showing the prominent structural motif of bond formation between Bi and Ga atoms. Based on all observed adsorption minima PESE generated 20 reaction paths whereby 19 reaction paths were successfully optimized. Here, 7 minima and 17 transition states were found in the NEB calculations. While 6 of the path minima were identified as duplicates to already known structures, one was identified as an so far unknown minimum. Based on the transition states diffusion barriers in the range of $+7\,\mathrm{kJ{\cdot}mol^{-1}}$ to $+127\,\mathrm{kJ{\cdot}mol^{-1}}$ were calculated underlining the high mobility of Bi on a GaP surface. Based on the reaction network a preference to diffuse to one of the most stable structures with the formation of Bi – Ga bonds was observed.

The obtained adsorption and diffusion network of Bi on GaP(001) nicely shows that PESE can be used to study surface reaction networks with minimal human interventions. Still, it can be expected that the obtained adsorption structures and diffusion paths are also found by a thorough manual exploration with considerably higher effort. However, the calculation

of the adsorption basins, especially the construction of the structure database and efficient database queries, can not be replaced by manual effort due to the high number of queries. Consequently, the calculation of the adsorption basins is uniquely attributed to the usage of PESE even for the smallest system of a single atom as adsorbate.

4.3. Reactivity of Gallane on Gallium Phosphide

The adsorption and decomposition of gallane (GaH_3) on GaP is studied with PESE as the second system. GaH_3 itself is not a commonly used Ga-precursors since it is highly unstable[273,274] independent from its occurrence as monomer or dimer[275]. For the growth of GaP, GaH_3 is only used in case it is synthesized directly in the vaporization of the Ga-precursor[276] or present as a gallane adduct[277]. More common is the usage of alkyl containing Ga-precursors like trimethylgallium (TMGa), TEGa or even Ga(bdma)H_2[195,278–280]. Here, GaH_3 is obtained as a gas phase decomposition product of TEGa[65]. Since common Ga-precursors as TEGa already show a large variety of decomposition reactions[65], GaH_3 is the smallest reasonable test system for PESE allowing the investigation of decomposition reactions.

Adsorption Minima

Calculate Adsorption Minima

For the calculation of the adsorption basin and the adsorption minima, similar PESE settings were used for GaH_3 as for Bi. Again, based on 810 grid points the initial structures were generated. In contrast to Bi, 12 different orientations of GaH_3 were used at every grid point. The different orientations for GaH_3 were obtained by combining rotations of $0°$, $90°$, $180°$ and $270°$ around every principal axis. From the 64 combinatorial possible orientations only 12 are unique as identified by an RMSD threshold of 0.5 Å. Due to the number of orientations and grid points 9720 initial structures were generated in total.

After the optimization of all initial structures, 5376 structures fulfilled the criterium of a minimum structure. By the removal of duplicates 1549 structures were then identified as unique. Within this step RMSD differences of up to 1.0 Å were observed between duplicate and a unique structures while the differences in energy were always smaller than 2.5 kJ·mol^{-1}. Consequently, the selected RMSD threshold of 1.0 Å was the decisive factor for the detection of duplicates and not the energy threshold of 10 kJ·mol^{-1}. This shows that structural similarity is harder to achieve and thereby the more strict criterium. Out of all unique minima only the small number of 18 minima were identified as chemisorbed structures whereby 14 structures were identified as physisorbed structures. Therefore, the majority of all minima belongs to the group of gas phase structures, which are defined by a large distance and vanishing interactions between the adsorbate and surface.

A selection of chemisorbed minima is shown in Figure 4.13 while all other chemisorbed minima are shown in the appendix (Figure C.7 to C.9). With an adsorption energy of -148 kJ·mol^{-1},

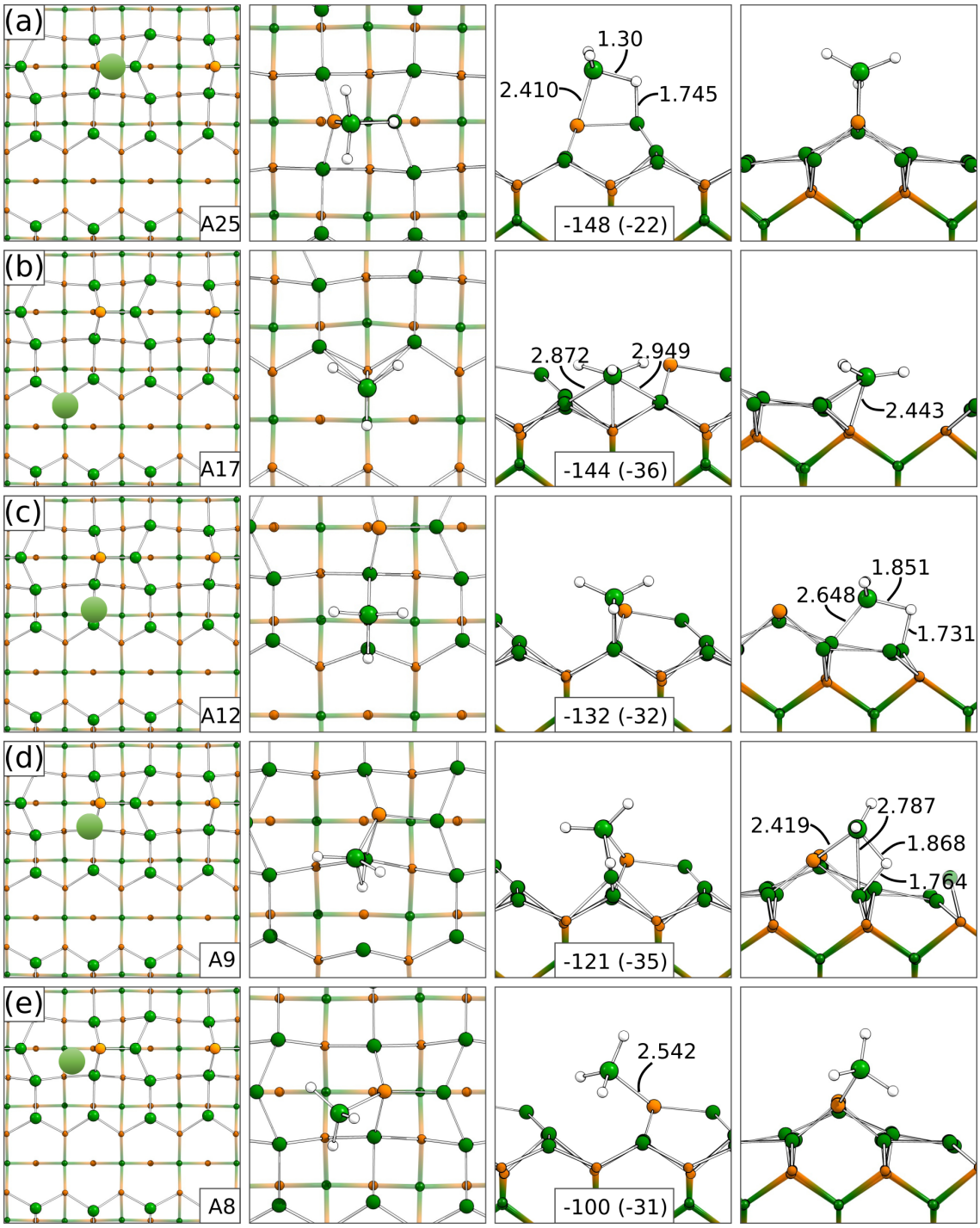


Figure 4.13.: Top and side view on adsorption structures of GaH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in $kJ \cdot mol^{-1}$ with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima ("A") follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

the most stable structures for GaH_3 on GaP(001) is observed in case the adsorption takes place at the mixed-dimer (Figure 4.13a). In this structure a Ga-P and a H-Ga bond is formed. Interestingly, a very similar structure is shown in Figure 4.13e where only the Ga-P bond is formed between GaH_3 and the mixed-dimer. However, this structure shows with $-100\,\text{kJ·mol}^{-1}$ one of the largest adsorption energies for all obtained chemisorbed minima. This shows the importance of the common structural motif of H-Ga bond formation to obtain stable adsorption structures (Figure 4.13c, d). The participation of the hydrogen atoms in bond formation is not unexpected due to the larger electronegativity for hydrogen $\chi_H = 2.20$ in comparison to gallium $\chi_{Ga} = 1.81$. Consequently, the hydrogen atoms in GaH_3 are carrying a negative partial charge and are preferably interacting with the Ga-atoms on the GaP surface. Besides the adsorption on top of the mixed-dimer (Figure 4.13a, e), common structural motifs for GaH_3 on GaP are the adsorption in the trench between two topmost Ga-layers (Figure 4.13b), the adsorption on top of the Ga-Ga dimer (Figure 4.13c) or on top of a Ga-Ga or Ga-P bond to the mixed-dimer (Figure 4.13d).

In comparison to Bi on GaP(001), all observed adsorption energies for GaH_3 are in their range significantly higher by up to $+111\,\text{kJ·mol}^{-1}$ for the most stable structures. However, this difference can be attributed to the unsaturated nature and thereby high reactivity of the Bi atom.

Adsorption Basins

Within the calculation of the adsorption basins a dataset containing $287\,392$ structures was obtained. This is a significant increase by a factor of 16.7 in comparison to the Bi dataset. The larger number of database structures stems from the presence of different orientations for GaH_3 (D_{3h} symmetry) and furthermore from a more complex PES due to the variability of the Ga-H bond lengths. The screening of the smearing parameters is shown in the appendix (Figure C.6). Here, due to the larger dataset a parameter of $\omega = 0.15$ is sufficient to fill the whole space of the unit cell.

In Figure 4.14 and also in the appendix (Figure C.10 to C.12) the adsorption basins of the chemisorbed minima are shown. For the most stable structure (Figure 4.14a) only a basin of $12.2\,\%$ is observed while the largest basin (Figure C.11b) comprises only $15.5\,\%$ of the mapped area within the unit cell. In comparison to Bi with up to $31.1\,\%$ the basins for GaH_3 are drastically decreased in size. This is a consequence of the larger number of chemisorbed and physisorbed structures which compete for the size of their basins. In the dataset of GaH_3 on GaP(001), a clear preference for a single adsorption structure based on the basins is not

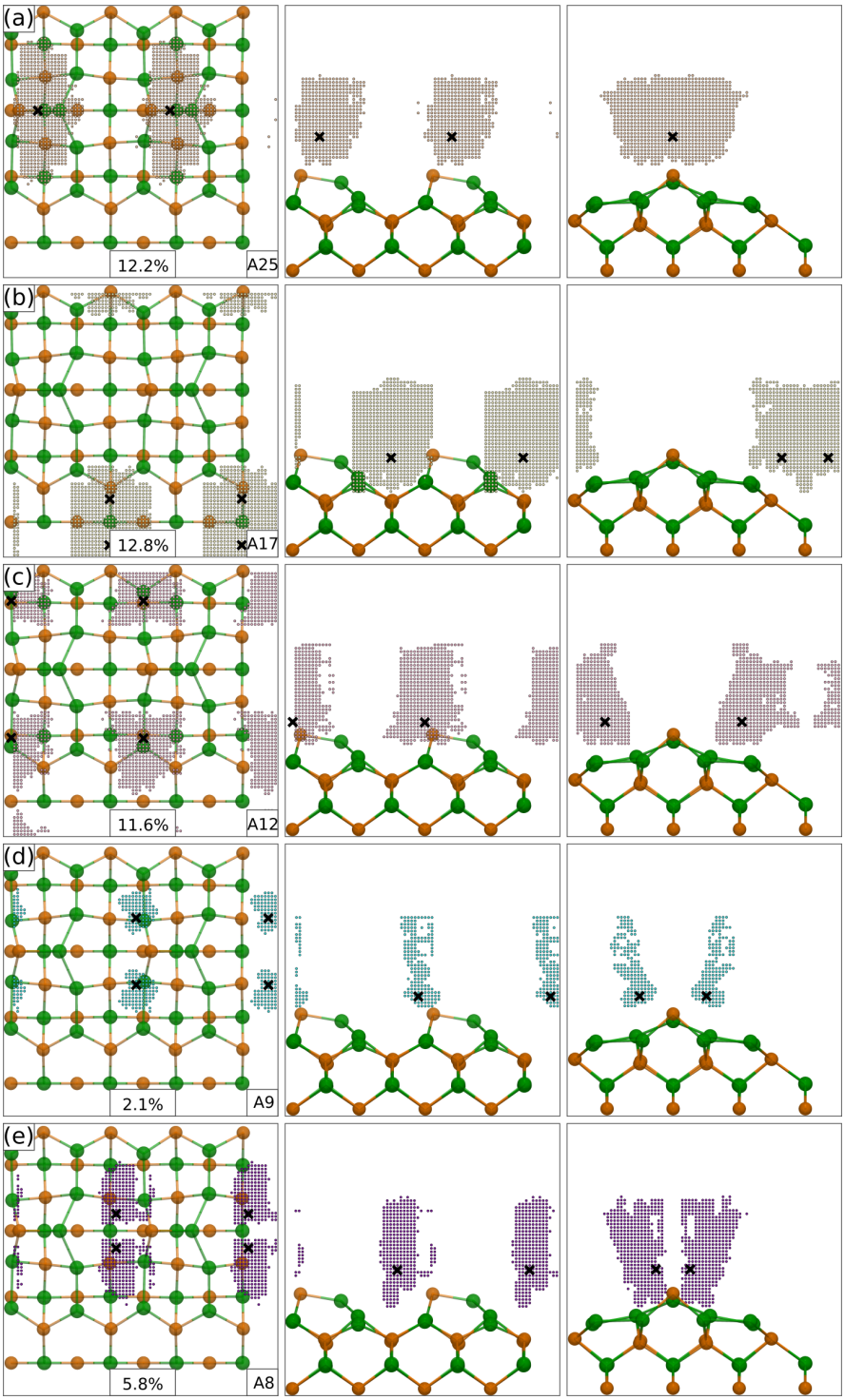


Figure 4.14.: Top and side view on adsorption basins of GaH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

present since the structures with the largest basins range from adsorptions at the mixed-dimer ($12.2\,\%$, Figure 4.14a), structures adsorbed on top of the Ga-Ga dimer ($11.1\,\%$ Figure C.10e, $15.5\,\%$ Figure C.11b, $11.6\,\%$ Figure C.11c) or structures in the trench between two top-most Ga-layers ($11.1\,\%$ Figure C.11e, $12.8\,\%$ Figure 4.14b).

In the basins shown another effect is present as shown in Figure 4.14d, C.10c and C.11a: The adsorption basin of a single minimum is in this representation no longer continuous. This is a shortcoming of the present representation being unable to contain all information of the dataset. Here, a point is colored by the minimum with the largest contribution. Consequently, only the dominant minimum is visible for every grid point. However, GaH_3 on GaP(001) is comprising of different orientations with the same location of the centroid. As a consequence adjacent grid points are assigned to different minima and the basins in this representation are fragmented. Here, a more sophisticated postprocessing and analysis regarding the basin of a single minimum or the effect of the adsorbate orientation has to be developed in the future to access all the information contained in this dataset. For this, the corresponding routines of PESE would need some rework to enable the projection of the database structures on the final grid for only a certain minimum or orientation.

In Figure 4.15 the correlation between the adsorption energy and the basin size for GaH_3 on GaP(001) is shown. Similar to Bi on GaP(001), the largest values for the basins (basin size$>10\,\%$) are often found for the more stable adsorption structures ($\Delta E_{Ads} < -130$ kJ·mol^{-1}). The occurrence of a larger basin with a more attractive adsorption energy could be explained

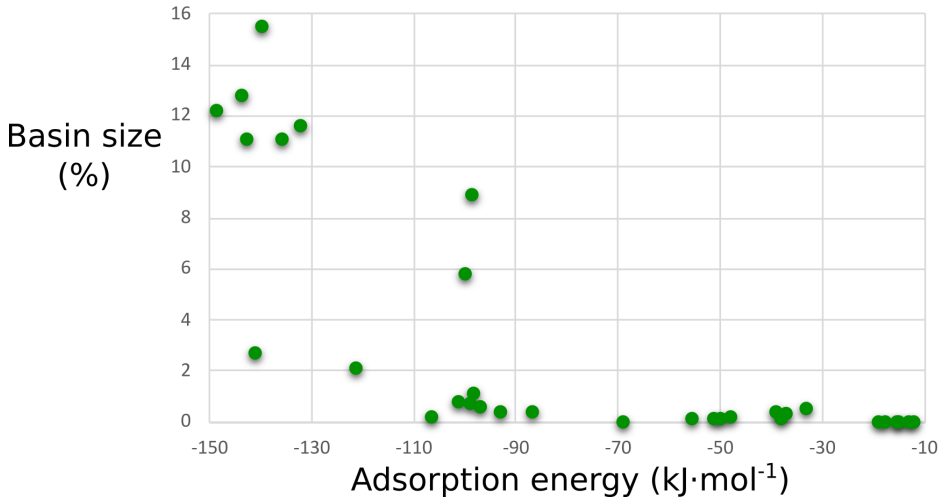
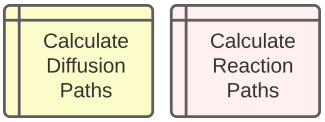


Figure 4.15.: Correlation between the basin size and adsorption energy of GaH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in % and the adsorption energy of the assigned minimum in kJ·mol^{-1}.

by the deepening of the PES due to the attractive forces at the minimum structure. This could also explain the presence of very small basins for weakly bonded structures. However, this effect alone is not able to explain the distribution in Figure 4.15, especially why favorable adsorption energies of $-100\,\mathrm{kJ{\cdot}mol^{-1}}$ or $-141\,\mathrm{kJ{\cdot}mol^{-1}}$ coincide with small basin sizes. Here, a possible explanation is that these structures are hard to reach since they are in high competition with adjacent minima. This could for example explain why for the adsorption on top of the Ga-P bond to the mixed dimer (Figure 4.13d, ΔE_{Ads}: $-121\,\mathrm{kJ{\cdot}mol^{-1}}$) a small adsorption basin of $2.1\,\%$ is observed. Here, this adsorption site is in competition with the adjacent site of an adsorption on top of the mixed-dimer (Figure 4.13e & 4.14e, $5.8\,\%$) and on top of the Ga-Ga dimer (Figure 4.13c & 4.14c, $11.6\,\%$). Following this argumentation, the presence of larger adsorption basins for less strongly bonded structures is then a consequence of the relatively good accessibility of these structures. However, the observation that all physisorbed structures come along with minor basins ($< 1.0\,\%$) is not indicating that physisorbed structures are hard to reach. It should be understood as a proof for the ease of GaH_3 to reach chemisorbed structures due to direct adsorption paths and thereby a very confined and small area in which physisorbed structures are observed.

Overall, a correlation between the size of a adsorption basin and the adsorption energy of the minimum is only present in the sense that a larger (smaller) absolute value of the adsorption energies most likely come along with with a larger (smaller) basin size. Nevertheless, the presented adsorption basins show the importance to not only rely on the adsorption energy but also on the accessibility of an adsorption site and the competition between adjacent adsorption sites to judge the adsorption probability to a certain minimum.

Reaction Paths



Besides diffusion paths of GaH_3, PESE derived possible decomposition structures and proposed the corresponding reactions paths based on the chemisorbed minima obtained. Here, the focus was on hydrogen transfer and hydrogen elimination reactions as shown in reactions R 4.1 and R 4.2, respectively.

$$GaH_3 \longrightarrow H_2Ga^{\bullet} + H^{\bullet} \qquad \text{(R 4.1)}$$

$$GaH_3 \longrightarrow HGa^{2\bullet} + H_2 \qquad \text{(R 4.2)}$$

However, due to the high reactivity of the GaH_3 fragments, further decomposition reactions were observed, which are shown in the reactions R 4.3 to R 4.5. Here, also reactions in which

GaH_3 is losing two (reaction R 4.3) or all three hydrogen atoms (reaction R 4.4 and R 4.5) are observed. It should be noted that these three reactions were not systematically studied and the corresponding dataset is therefore small. Consequently, the obtained reaction energies and barriers should be understood as a first estimate for the importance of these reactions.

$$H_2Ga^{\bullet} + H^{\bullet} \longrightarrow HGa^{2\bullet} + 2\,H^{\bullet} \tag{R 4.3}$$

$$GaH_3 \longrightarrow Ga^{3\bullet} + 3\,H^{\bullet} \tag{R 4.4}$$

$$GaH_3 \longrightarrow Ga^{3\bullet} + H_2 + H^{\bullet} \tag{R 4.5}$$

In Figures 4.16 to 4.19 four examples for different decomposition reactions are shown. All reactions start at an adsorption minimum of GaH_3. In addition, the reactions contain the smallest observed reaction barrier for the individual decomposition reaction. Consequently, an example for reaction R 4.3 is not shown since this reaction was only observed twice as a consecutive reaction of the hydrogen transfer reaction R 4.1.

As shown in Figure 4.16, even the calculation of a simple hydrogen transfer reaction can re-

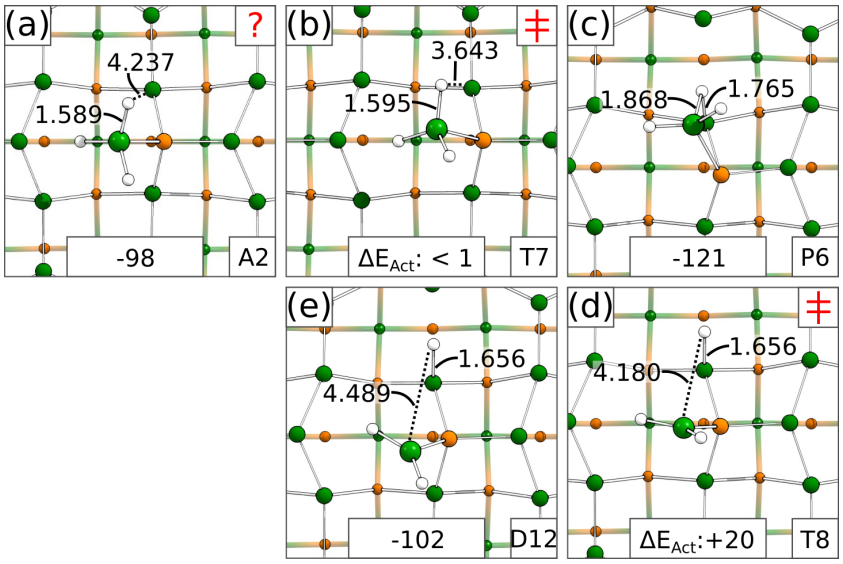


Figure 4.16.: Decomposition path for GaH_3 on GaP(001). This example shows a hydrogen transfer reaction as shown by reaction R 4.1. The transition state structures are marked by a double dagger sign (‡) while the false minimum structure is marked with an question mark (?). Adsorption energies are stated for every minimum structure while activation energies ΔE_{Act} are stated for every transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

sult in complex reaction paths with minima and several TSs along the reaction path. Starting from the initial structure shown in Figure 4.16a a minimum (Figure 4.16c) along the path is observed representing a change in the adsorption site. However, the corresponding TS shows a negligible reaction barrier with less than $+1$ kJ·mol^{-1} (Figure 4.16b). Therefore, the initial starting structure in Figure 4.16a should not be considered a minimum structure since it can be converted to a minimum without the presence of a reaction barrier. The presence of such false minima can not be excluded by using common optimization algorithms alone since a convergence of the optimization is also reached for a plateau on the PES. Frequency calculations could solve the false detection of minima but are too expensive to be performed for every minima PESE is obtaining. A partial calculation of the hessian[281] might be a suitable extension to PESE to solve this problem by calculating only frequencies attributed to the adsorbate.

Continuing from the structure in Figure 4.16c the hydrogen transfer reaction is observed by moving the GaH_2 fragment on the other side of the mixed-dimer. During this reaction the Ga-H distance is increased from 1.868 Å (Figure 4.16c) to 4.489 Å (Figure 4.16e). With $+20$ kJ·mol^{-1} a small reaction barrier (Figure 4.16d) is observed. However, this reaction is not only en-

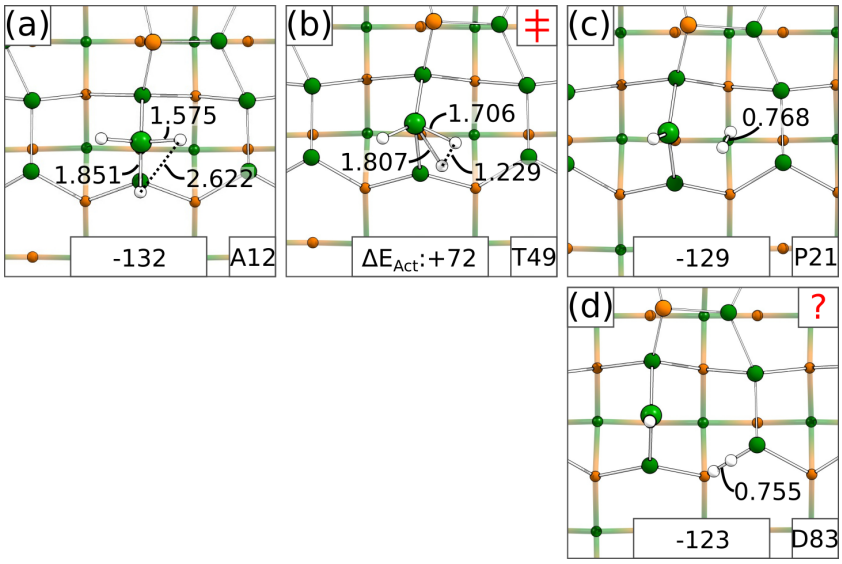


Figure 4.17.: Decomposition path for GaH_3 on GaP(001). This example shows a hydrogen elimination reaction as shown by reaction R 4.2. The transition state structure is marked by a double dagger sign (‡) while a false minimum structure is marked with an question mark (?). Adsorption energies are stated for every minimum structure while the activation energy ΔE_{Act} is stated for the transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

dothermic but the barrier for the backreaction is with $+1$ kJ·mol^{-1} small. Consequently, for this adsorption site the back reaction is more likely as long as none of the fragments are diffusing along the surface.

In Figure 4.17 an example for a hydrogen elimination reaction is shown. As the starting point the GaH_3 is adsorbed on top of a Ga-Ga dimer (Figure 4.17a). The TS (Figure 4.17b) is reached by shortening the H-H distance to 1.229 Å. This change in structure is responsible for a reaction barrier of $+72$ kJ·mol^{-1}. With a bond length of 0.768 Å the final H_2 molecule is present in Figure 4.17c. However, Figure 4.17d was actually used as the final point of the reaction path in the NEB calculation. This structure can be reached by a diffusion of the H_2 molecule but without a TS. Therefore, this time the final decomposition minimum was falsely detected as a minimum structure. Still, this example shows that the implemented NEB algorithm of PESE is stable enough to deal with these false minima.

An example for the interesting reaction R 4.4 is shown in Figure 4.18. This reaction was only observed for structures, in which GaH_3 is adsorbing in the trench between two topmost Ga-layers. Surprisingly, for the cleavage of all three Ga-H bonds only a single TS is observed with Figure 4.18b. Furthermore, the reaction barrier for this reaction is with $+63$ kJ·mol^{-1} even smaller than the corresponding barrier for the shown hydrogen elimination reaction.

The last example in Figure 4.19 shows the hydrogen elimination simultaneously to a hydrogen transfer reaction. The starting point is again an adsorption structure of GaH_3 in the trench between the top-most Ga-layers (Figure 4.19a). The TS (Figure 4.19b) is reached by initiating the formation of the H_2 molecule. This is achieved by reducing the H-H distance to 1.345 Å. Interestingly, the two hydrogen atoms of the H_2 molecule show with 2.633 Å and 1.658 Å distinctly

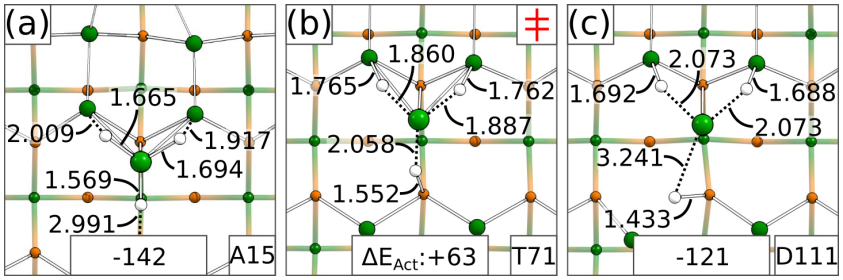


Figure 4.18.: Decomposition path for GaH_3 on GaP(001). This example shows a hydrogen transfer reaction as shown by reaction R 4.4. The transition state structure is marked by a double dagger sign (‡). Adsorption energies are stated for every minimum structure while the activation energy ΔE_{Act} is stated for the transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

different Ga-H distances. Here, the formation of the H_2 molecule and the large Ga-H distances within the former GaH_3 are stabilized by an adjacent surface Ga-atom forming a Ga-H bond of 1.699 Å. In the second part of the reaction the H_2 molecule is released and simultaneously the third H-atom is transferred to a surface Ga-atom as shown by the final bond length of 1.694 Å. Furthermore, with $+111$ kJ·mol^{-1} the largest barrier in the presented examples is observed for this type of reaction.

Within these examples the identification of false minima was introduced. For a general comparison of different decomposition reaction these false minima come along with another problem: For reaction barriers, a false reference point is present in case these false minima are used as the reference since they already include a fraction of the true activation energy. This problem can be solved in case such a false minimum can uniquely be assigned to a true minimum structure. However, this is not always the case for the obtained reaction network of GaH_3 on GaP(001). Furthermore, complexity is added to this problem since even false minima are connected to each other without the presence of a reaction barrier. To solve this assignment problem an iterative algorithm deriving all possible reference points with the correct resulting reaction barriers would be necessary. Since such an algorithm is not implemented yet, the simpler approach of referencing all TSs to the corresponding most stable fragment structure is chosen. Consequently, all reaction barriers stated in the overview in Figure 4.20 include the energy change to a less stable conformer as well as the actual reaction barrier.

For the overview in Figure 4.20, PESE proposed in total 48 diffusion and 169 decomposition paths. In the end, 103 reaction path calculations could be converged while 114 attempts to calculate a reaction path failed. This represents a ratio between converged to unconverged paths of 0.9 (103:114). The observation that PESE manages to converge every second reaction

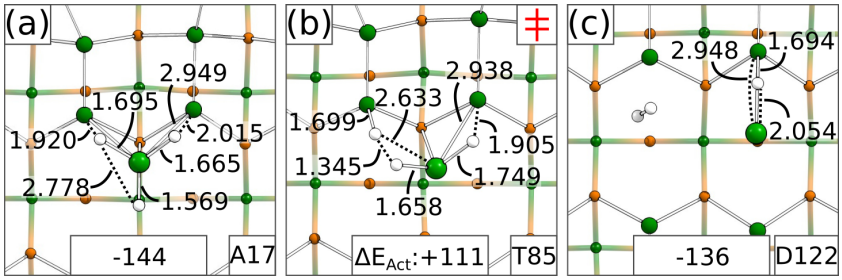


Figure 4.19.: Decomposition path for GaH_3 on GaP(001). This example shows a combination of a hydrogen transfer and hydrogen elimination reaction as shown by reaction R 4.5. The transition state structure is marked by a double dagger sign (‡). Adsorption energies are stated for every minimum structure while the activation energy ΔE_{Act} is stated for the transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

paths shows the stability and reliability of the employed NEB algorithm especially since most of the converged reaction paths obtained several reaction steps and TSs.

As shown in Figure 4.20, for the adsorption of GaH_3 on GaP neither a reaction barrier was found for a path linking a gas phase structure with a physisorbed structure nor for a path linking a physisorbed and chemisorbed structure. This shows the presence of direct adsorption paths for GaH_3 on GaP(001), which is in agreement with the small adsorption basin sizes for physisorbed structures. Furthermore, for GaH_3 a differentiation between the gas phase, physisorbed and chemisorbed structures is also possible based on the adsorption energy: Chemisorbed structures are observed with an adsorption energy of less than $-100\,\mathrm{kJ{\cdot}mol^{-1}}$. A slightly higher bound of $-96\,\mathrm{kJ{\cdot}mol^{-1}}$ for the adsorption energy is obtained in case also conformers observed in the optimization of reaction paths are considered. By increasing the adsorption energy, physisorbed structures are observed with energies in the range of -55 to $-12\,\mathrm{kJ{\cdot}mol^{-1}}$ while the smallest adsorption energies are observed for gas phase structures with an energy of up to $-7\,\mathrm{kJ{\cdot}mol^{-1}}$.

Based on the chemisorbed structures, 46 converged diffusion paths were obtained. Here, the diffusion barriers relative to the most stable GaH_3 adsorption structures are in the range of $+39$ to $+118\,\mathrm{kJ{\cdot}mol^{-1}}$. Therefore, GaH_3 is mobile on the GaP(001) surface under experimen-

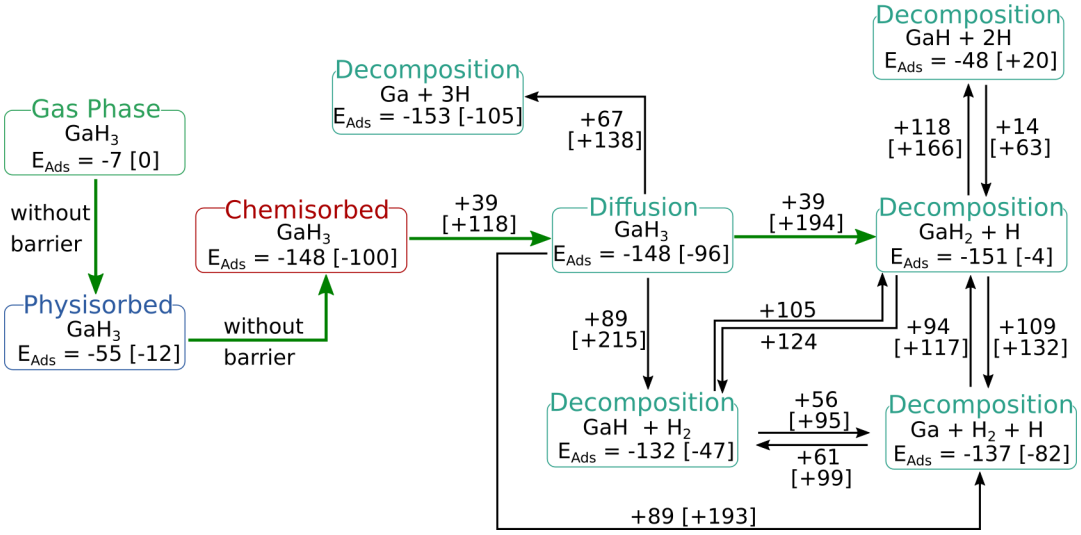


Figure 4.20.: Overview showing the obtained decomposition reactions of GaH_3 on GaP(001). Range of adsorption energies ΔE_{Ads} stated at the fragments while range of activation energies ΔE_{Act} stated at the connections. The highest value is stated in parentheses. Furthermore, the range of activation energies is stated relative to the most stable fragment structure, respectively. Most likely reaction path marked in green. All values in $\mathrm{kJ{\cdot}mol^{-1}}$.

tal conditions of up to $600\,°C$.

The most likely decomposition reaction in Figure 4.20 is the hydrogen transfer reaction (reaction R 4.1) as shown by the lower limit of the reaction barriers for the individual decomposition reactions. Here, barriers in the range of $+39$ to $+194\,kJ{\cdot}mol^{-1}$ are observed based on 27 converged reaction steps. Furthermore, based on the most stable structure for the decomposition product of $-151\,kJ{\cdot}mol^{-1}$, this reaction can be thermodynamically favored. However, with an upper limit of $-4\,kJ{\cdot}mol^{-1}$ also strongly unfavored reaction products were obtained.

The second most likely decomposition reaction of GaH_3 is the reaction R 4.4, in which all three H-atoms are transferred to the surface. With $+67$ to $+138\,kJ{\cdot}mol^{-1}$ a smaller range of reaction barriers was observed. However, as already discussed this reaction is only observed for one type of adsorption structures and only in four converged steps. Still, with an lowest energy of $-153\,kJ{\cdot}mol^{-1}$ this reaction can be thermodynamically favored.

With reaction barriers in the range of $+89$ to $+215\,kJ{\cdot}mol^{-1}$ (hydrogen elimination, 27 converged steps) and $+89$ to $+193\,kJ{\cdot}mol^{-1}$ (hydrogen elimination + hydrogen transfer, 10 converged steps) similar lower limits for the reaction barriers are observed for the reactions R 4.2 and R 4.5. Furthermore, both types of reactions are with their energies of -132 to $-137\,kJ{\cdot}mol^{-1}$ for their most stable structures slightly endothermic in comparison to the most stable GaH_3 structure.

For the decomposition reaction R 4.3 (2 converged steps) as well as reactions linking the discussed decomposition products, only between one and six reaction steps were converged. As already discussed, this is due to PESE focussing on the hydrogen transfer and hydrogen elimination reaction. Therefore, for these decomposition reactions it can be assumed that reaction paths with lower barriers and more favorable thermodynamics are likely. Consequently, the shown examples should be understood as an upper limit to the smallest possible barrier and adsorption energy. Based on the shown numbers in Figure 4.20 a consecutive decomposition is likely in case a hydrogen elimination was observed in the first step due to a small lower limit of the barriers of $+56\,kJ{\cdot}mol^{-1}$. In case a hydrogen transfer reaction was observed in the first step, a consecutive reaction is less likely due to larger barriers of $+94\,kJ{\cdot}mol^{-1}$ or $+118\,kJ{\cdot}mol^{-1}$. However, at experimental conditions of $600\,°C$ all shown reactions are accessible.

In the work of Stegmüller at. al.[65] similar reactions for GaH_3 were calculated in the gas phase. Here, significantly more endothermic reactions are observed with reaction energies of $+338\,kJ{\cdot}mol^{-1}$ for the hydrogen transfer reaction and $+79\,kJ{\cdot}mol^{-1}$ for the hydrogen elimination reaction. Especially, for the hydrogen transfer reaction this can be explained by the

lacking stabilization of the formed radicals due to the presence of the GaP surface. Furthermore, a similar effect is observed for the activation barriers. The barrier for the hydrogen elimination reaction in the gas phase was calculated to be $+212$ kJ·mol^{-1}, which is similar to the upper bound of reaction barriers observed at the GaP surface. This shows that the GaP surface is able to strongly facilitate decomposition reactions of GaH_3.

Conclusion

The adsorption, diffusion and decomposition of GaH_3 on GaP(001) was studied with PESE. Adsorption energies for GaH_3 in the range of -96 to -148 kJ·mol^{-1} with the common structural motif of H-Ga bond formations were observed. Here, the most stable structure is found for an adsorption on top of the mixed-dimer. However, the adsorption basins with up to 15.5 % are small for a single minimum indicating the absence of a preference for a single adsorption site. Instead the adsorption of GaH_3 is expected to take place at the mixed-dimer, in the trench between two topmost Ga-layers or on top of a Ga-Ga or Ga-P bond. Several exemplary reactions paths were presented showing that complex reaction paths are obtained with multiple minima and TS along the reaction paths. Still, PESE was able to converge every second path emphasizing the reliability of the underlying algorithms. This results in the convergence of 113 reaction paths with 87 minima and 142 TSs along the reaction paths. Still, further improvements in the algorithm for the reaction path calculation are desirable to obtain a higher rate of converged reaction paths and to ensure that no possible reaction path is missed. However, it is not reasonable to expect that every reaction path calculation can be converged. Consequently, several attempts have to be performed for every decomposition reaction as currently implemented in PESE.

Based on the obtained data the most likely reaction is a hydrogen transfer reaction for a single hydrogen or all three at once. The formation of H_2 shows larger barriers and less favored reaction energies. Furthermore, in comparison to gas phase calculations more favorable reaction energies and barriers are observed since a major stabilization by the surface is included. Overall, several decomposition channels were observed for GaH_3 which are accessible at experimental conditions.

Similar to Bi on GaP(001), it can be expected that the obtained adsorption structures and simple decomposition paths describing the hydrogen transfer R 4.1 or the hydrogen elimination reaction R 4.2 are found by a thorough manual exploration. However, the calculation of the adsorption basins can only be performed in an automated exploration by PESE. In addition, another advantage of PESE is apparent in the obtained reaction network. Here, unintuitive

reactions as the transfer of all three hydrogen atoms in a single reaction step (reaction R 4.4) are observed since PESE is approaching every minimum and decomposition reactions without bias.

4.4. Reactivity of Phosphine on Gallium Phosphide

The third system studied with PESE is the adsorption and decomposition of phosphine (PH_3) on GaP(001). In contrast to GaH_3, the usage of PH_3 as a precursor is possible[280]. However, due to its toxicity PH_3 is commonly replaced by TBP[280] or more complex precursors as for example di-tert-butylaminophosphane (DTBAP)[282]. Similar to GaH_3, PH_3 can be obtained as a decomposition product of TBP[65].

For PESE this adsorbate is of interest since it is representing the smallest group V precursor. Furthermore, some theoretical studies regarding the adsorption of PH_3 on nanotubes and nanocluster[283,284] as well as the decomposition on some semiconductor surfaces like Si(001)[285–287], Ge(001)[285] and InP(001)[288,289] already exist.

Adsorption Minima

Calculate Adsorption Minima

For the calculation of the adsorption basins and the adsorption minima, similar PESE settings were used for PH_3 as for Bi and GaH_3. Based on 810 grid points and 24 different orientations of PH_3 $19\,440$ initial structures were generated. By the optimization of the initial structures, $11\,143$ minimum structure were obtained. After the removal of duplicates 2199 structures were then identified as unique. Here, RMSD differences of up to 1.0 Å and energy differences of up 10.0 kJ·mol^{-1} were observed between duplicate and a unique structures. However, the upper limit of the energy threshold of 10.0 kJ·mol^{-1} was reached only by a single duplicate indicating that the structural criterium is decisive for PH_3 as well. Within the set of the unique minima only nine minima were identified as chemisorbed structures whereby 118 structures were identified as physisorbed structures. As for GaH_3, the majority of all minima belongs to the group of gas phase structures observed as soon as the distance between the adsorbate and surface is too large for any interaction.

In Figure 4.21 and C.14 the obtained chemisorbed minima are shown. With -66 kJ·mol^{-1}, the most stable adsorption structure is observed in case the adsorption takes place at a Ga-atom at the trench between two topmost Ga-layers (Figure 4.21a). Also, for the adjacent Ga-atom in the trench (Figure 4.21b) a stable adsorption site is present with an adsorption energy of -63 kJ·mol^{-1}. The least stable adsorption site is found in case PH_3 is inserted in a Ga-Ga dimer bond (Figure 4.21d) with -24 kJ·mol^{-1}. In addition, an adsorption structure for PH_3 at the mixed-dimer is observed with an energy of -63 kJ·mol^{-1} as shown in Figure 4.21c. However, these four adsorption sites are the only sites detected for PH_3 on GaP(001). The

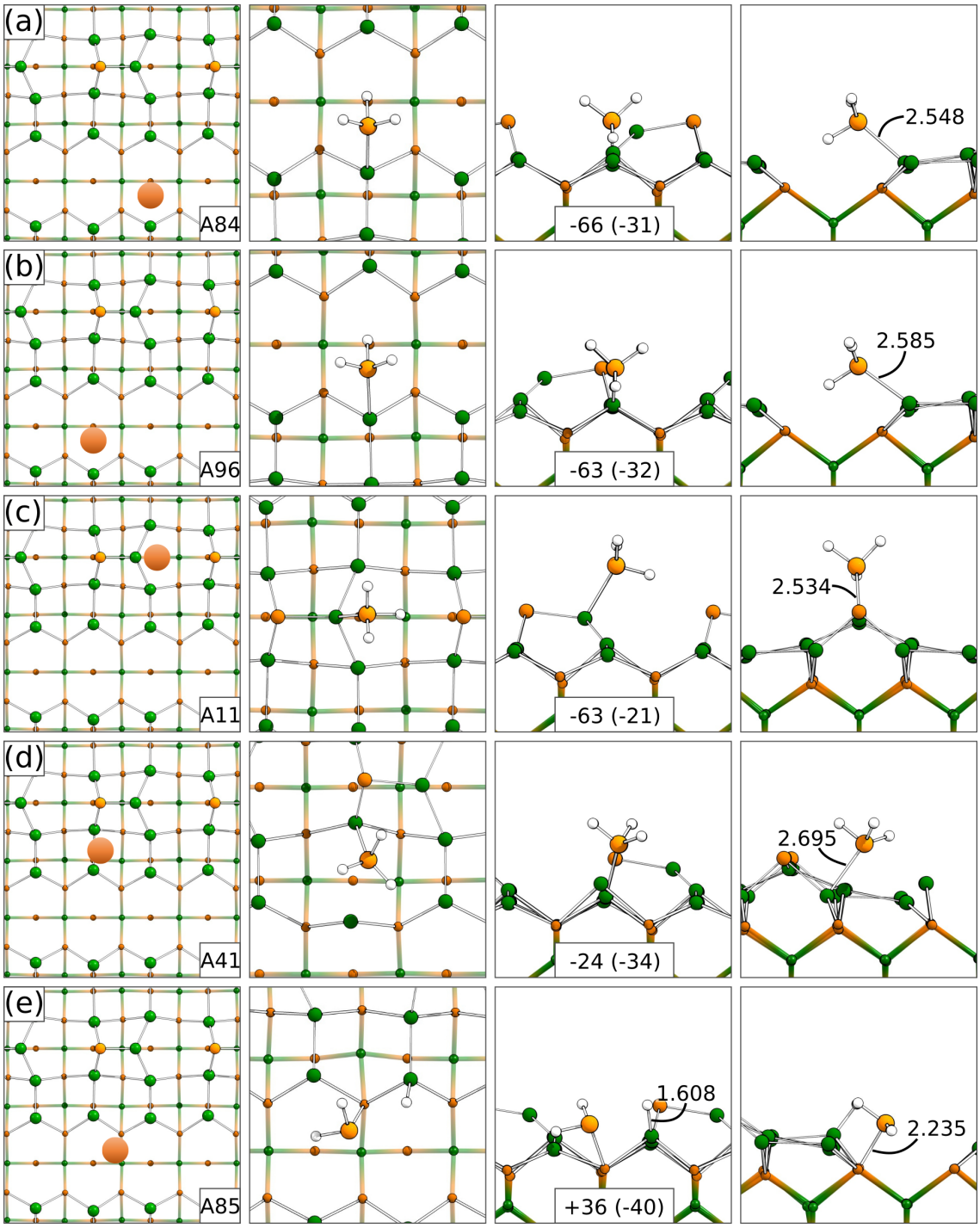


Figure 4.21.: Top and side view on adsorption structures of PH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima (”A”) follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

structure in Figure 4.21e already shows a decomposition structure of PH_3 obtained during the calculation of the adsorption basins. Here, it is unclear whether this structure is observed since the initial distance between the adsorbate and surface was to small or whether a combination of adsorption and decomposition takes place. Furthermore, all structures in Figure C.14 can be obtained by rotating the PH_3 in one of the presented adsorption structures around the formed Ga-P bond. Therefore, they represent conformers and not additional adsorption sites.

The observation that the group 15 precursor shows a smaller variety of adsorption structures on the group 13 rich surface than the group 13 precursor might be surprising at first. However, the common structural motif of the H-Ga bond formation for GaH_3 was attributed to the polarity of the Ga-H bond within the precursor. For PH_3, a strong polarization of the P-H bond is not expected due to the similar electronegativity of the P-atom ($\chi_P = 2.19$) and the H-atom ($\chi_H = 2.20$). This is in agreement to the observation that only Ga-P bonds are formed in the adsorption of PH_3 on GaP(001). Furthermore, the Ga-P bond formation can be attributed to the classical formation of dative bonds[268,269] between elements of group 13 and 15. Consequently, the number of adsorption site is limited by the presence of Ga-atoms in the topmost layer. However, with Figure 4.21d only a single adsorption structure was found in which PH_3 inserts in a former Ga-Ga bond. At this point it is not clear whether similar structures for other Ga-Ga bonds do not exists or are shielded by large, adjacent basins. The latter could be an explanation why PESE was able to observe only one of these structures, even with the systematic sampling of the surface.

Adsorption Basins

For PH_3 on GaP(001) a dataset of $708\,314$ structures was obtained during the calculation of the adsorption basins. Thereby, the dataset again increased by a factor of 2.5 (41.2) in comparison to the GaH_3 (Bi) dataset. The more complex PES is again mainly related to the lower symmetry of the PH_3 molecule (C_{3v} symmetry) resulting in a higher number of orientations. Consequently, twice as much initial structures for every grid point were generated in comparison to GaH_3. Furthermore, as for GaH_3 a second factor in comparison to Bi is the increased complexity due to variable P-H bonds.

As discussed later in a performance section, PESE is able to handle such a large number of structures. However, since small systems like GaH_3 and PH_3 already result in so many database structures it might be necessary for the study of even larger precursors to restrict the sampling of the basins even further. This could be achieved by either focussing on a few

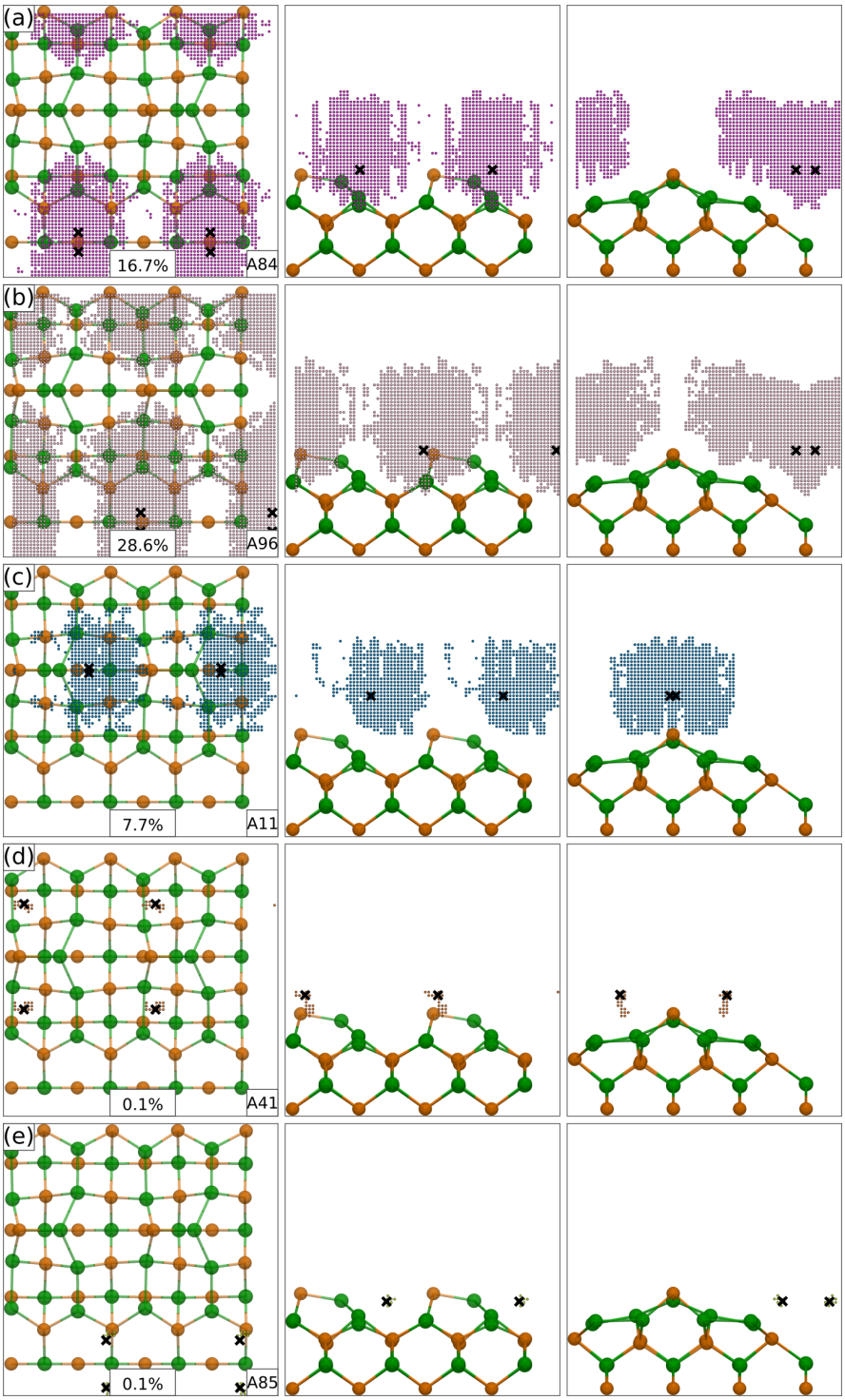


Figure 4.22.: Top and side view on adsorption basins of PH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

selected orientations or by performing a more static sampling. A more static sampling could for example be achieved by enabling PESE to freeze the internal coordinates of the adsorbate as long as a certain distance is present to the surface. This approach of using a static adsorbate structure to study the adsorbate-surface interactions for different orientations and distances was already successfully applied in the study of cyclooctyne on silicon[290,291].

For the evaluation of the PH_3 dataset a projection on the final grid was performed by PESE. The screening of the smearing parameters is shown in the appendix (Figure C.13). Here, due to the size of the dataset, a parameter of $\omega = 0.10$ is sufficient to fill the whole space of the unit cell.

In Figure 4.22 and C.15 the basins of the chemisorbed minima of PH_3 on GaP(001) are shown. Again, the most stable structure in Figure 4.22a is not showing the largest basin with $16.7\,\%$. However, the largest basin of up to $28.6\,\%$ (Figure 4.22b) is observed for the structurally very similar and adjacent adsorption site of PH_3 bonded to a Ga-atom in the trench between the topmost Ga-layers. For the adsorption site at the mixed-dimer (Figure 4.22c) a basin with $7.7\,\%$ is observed while for the two other shown structures (Figure 4.22d and 4.22e) a negligible size for the basin was observed. The negligible basin size for the decomposition structures indicates that this structure is only observed in case the PH_3 is placed in the initial optimization close to the final structure. Therefore, the presence of the decomposition structure in Figure 4.21e is attributed to a too short distance in the generation of the initial structures.

Overall, based on the observed basin sizes, the adsorption of PH_3 at a Ga-atom in the trench

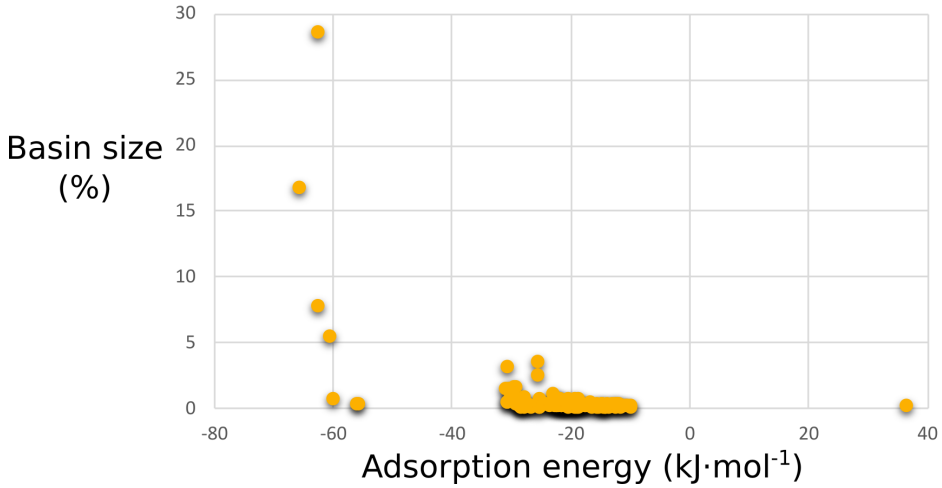
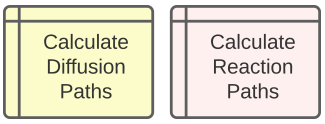


Figure 4.23.: Correlation between the basin size and adsorption energy of PH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in % and the adsorption energy of the assigned minimum in kJ·mol^{-1}.

between the topmost Ga-layers is clearly favored. Furthermore, from the two Ga-atoms in the trench the atom in line with the P-atom of the mixed-dimer is favored due to no competitive adsorption site in close proximity. Here, the Ga-atom trench atom in line with the Ga-atom of the mixed-dimer is competing with the adsorption on top of the mixed-dimer.

In Figure 4.23 the correlation between basin size and adsorption energy is shown for PH_3 on GaP(001). Here, a similar behavior as for Bi and GaH_3 is observed. The largest basins (basin size $>5\,\%$) are observed for the more stable structures with adsorption energies of less than $-40\,\mathrm{kJ{\cdot}mol^{-1}}$. Also, as for GaH_3 a tendency for larger (smaller) adsorption basins with larger (smaller) absolute values for the adsorption energy is present. Still, not every datapoint is strictly following this trend. For PH_3 on GaP(001), the observation of large basins for the chemisorbed structures is surprising since 118 physisorbed structures are present. However, similar to GaH_3 most of the physisorbed structures show negligible basin sizes of less than $0.5\,\%$. Therefore, this indicates that the presented adsorption sites with significant adsorption basin sizes comprise of direct adsorption paths. This coincide with the large extent of the basins in the z-direction.

Reaction Paths



As for GaH_3, PESE was used to study the hydrogen transfer reaction R 4.6 and hydrogen elimination reaction R 4.7 for PH_3 on GaP(001). In contrast to GaH_3, no further decomposition reactions were observed indicating a lower reactivity of the formed fragments. In addition to the decomposition reactions, the connectivity between physisorbed and chemisorbed structures was studied.

$$PH_3 \longrightarrow H_2P^{\bullet} + H^{\bullet} \tag{R 4.6}$$

$$PH_3 \longrightarrow HP^{2\bullet} + H_2 \tag{R 4.7}$$

In Figure 4.24 to 4.26 three example reactions are shown. An example reaction for an adsorption path is shown in Figure 4.24. This example path is important since it shows a crucial shortcoming of the implemented algorithms: In the starting structure (Figure 4.24a) the PH_3 molecule is already placed above the Ga-atom of the mixed-dimer. However, in this structures the hydrogen atoms are pointing toward the surface, which is why a physisorbed structure is observed. As shown by the TS (Figure 4.24b), PESE proposed an inversion of the PH_3 to reach the final adsorption structure (Figure 4.24c). Consequently, this path contains a reaction barrier of $+88\,\mathrm{kJ{\cdot}mol^{-1}}$. However, PESE was not considering the possibility that a rotation of the

PH_3 prior to closing the distance to the surface is energetically favored for the adsorption. The preference of the inversion over the rotation is a consequence of searching for the shortest possible reaction path (see section 3.2.5). Unfortunately, an inversion represents a more complex reaction path than a rotation. As a result, most of the proposed adsorption paths failed to converge. Here, further extensions to PESE are necessary to propose alternative reaction paths even if that results in the selection of longer reaction paths.

The second reaction path in Figure 4.25 shows an example for a hydrogen transfer reaction

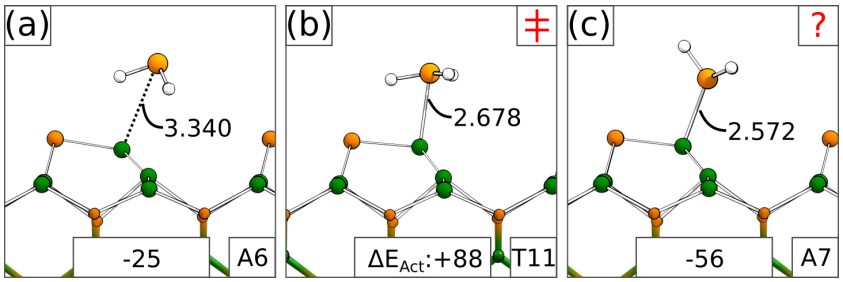


Figure 4.24.: Adsorption path for PH_3 on GaP(001). The transition state structure is marked by a double dagger sign (‡). Adsorption energies are stated for every minimum structure while the activation energy ΔE_{Act} is stated for the transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

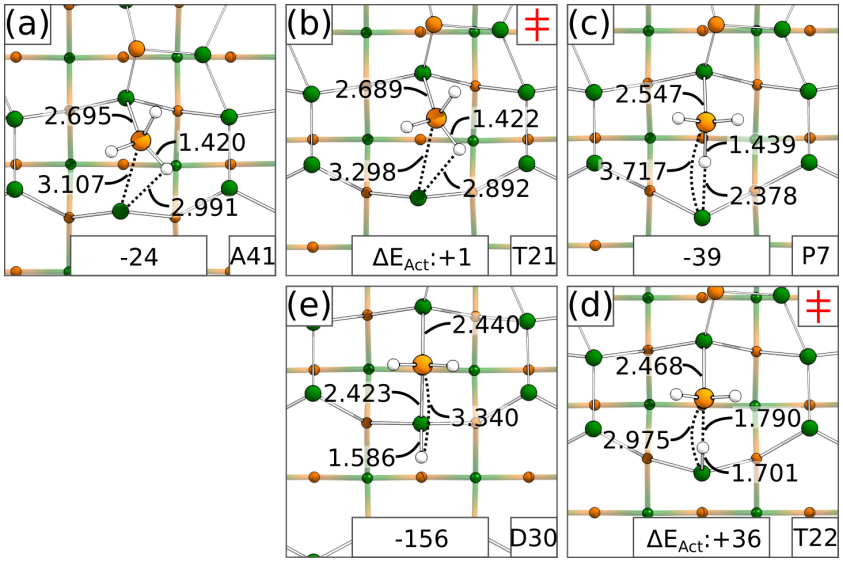


Figure 4.25.: Decomposition path for PH_3 on GaP(001) via a hydrogen transfer reaction R 4.6. The transition state structure is marked by a double dagger sign (‡). Adsorption energies are stated for every minimum structure while the activation energies ΔE_{Act} are stated for every transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Color code: Ga (green), P (orange), H (white).

R 4.6. The starting point of the reaction is the adsorption of PH_3 in the Ga-Ga dimer bond (Figure 4.25a). As a first step, a minor conformer change with a negligible barrier (ΔE_{Act}: $+1$ kJ·mol^{-1}, Figure 4.25b) is observed to reach the actual starting point of the hydrogen transfer reaction (Figure 4.25c). The hydrogen transfer reaction itself shows a barrier of $+36$ kJ·mol^{-1} (Figure 4.25d) necessary to elongate the P-H bond to 1.790 Å. Furthermore, a new Ga-H bond is already visible at the TS as shown by a bond length of 1.701 Å. In addition, the P-Ga distance to the second Ga-atom of the former Ga-Ga dimer is shortened by 0.742 Å. To reach the final structure (Figure 4.25e) the P-H distance is increased to 3.340 Å while at the same time the second Ga-P bond with a length of 2.423 Å is formed for the PH_2 fragment. As shown by the final structure (Figure 4.25e) the PH_2 fragment is now bonded to both Ga atoms of the former Ga-Ga dimer. The whole reaction is exothermic (ΔE_{React}: -117 kJ·mol^{-1}) due to the low energy of the final structure of -156 kJ·mol^{-1}.

The third example for a reaction path is shown in Figure 4.26. This path represents the hydrogen elimination reaction R 4.7. Here, the starting point is the adsorption of PH_3 on top of the mixed-dimer (Figure 4.26a). As for the second shown reaction path, a conformer change is observed as a first part of the reaction. Here, a rotation along the Ga-P bond with a minor barrier of $+7$ kJ·mol^{-1} (Figure 4.26b) is observed to reach the minimum shown in Figure 4.26c.

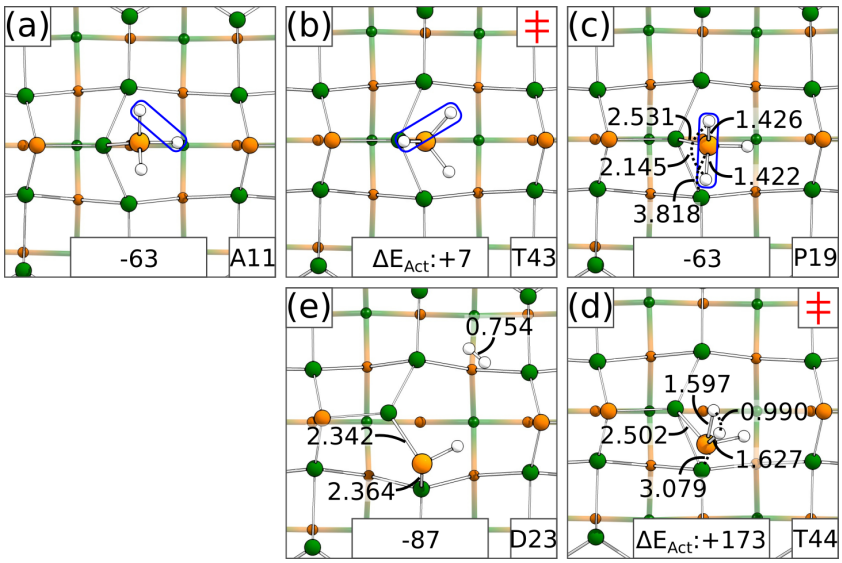


Figure 4.26.: Decomposition path for PH_3 on GaP(001) via a hydrogen elimination reaction R 4.7. The transition state structure is marked by a double dagger sign (‡). Adsorption energies are stated for every minimum structure while the activation energies ΔE_{Act} are stated for every transition state relative to the previous minimum. Energies in kJ·mol^{-1}. Hydrogen atoms selected for the formation of the H_2 molecule marked in blue. Color code: Ga (green), P (orange), H (white).

Interestingly, the structures in Figure 4.26a and 4.26c are identical. The conformer change was only necessary to align the hydrogen atoms, which were selected by PESE to form the H_2 molecule. In particular, the two hydrogens atoms forming the H_2 have to be on opposite sites of the mixed-dimer in the initial structure. Consequently, this additional conformer change can be avoided in case PESE is not restricted to an assignment based on the minimal distance. In the actual elimination reaction the H_2 is formed by closing the distance between the H-atoms to 0.990 Å (Figure 4.26d) while simultaneously moving the PH fragment in the direction of an adjacent Ga-atom (P-Ga distance of 3.079 Å). This reaction step consists of a large barrier with $+173$ kJ·mol^{-1}. By releasing the H_2 molecule the final structure of Figure 4.26e is obtained. Here, the PH fragment is again inserted in a former Ga-Ga bond as already observed for the PH_2 fragment in the second example path. The shown example reaction is also leading to a more stable structure with -87 kJ·mol^{-1} and thereby representing an exothermic reaction (ΔE_{React}: -24 kJ·mol^{-1}).

An overview of the obtained adsorption and decomposition reactions for PH_3 on GaP(001) is shown in Figure 4.27. For the reaction paths from a gas phase to a physisorbed structure commonly no barrier was observed. Only for a single reaction path a barrier of $+138$ kJ·mol^{-1} was observed due to the inversion reaction discussed for PH_3. For the adsorption to a chemisorbed structure, reaction paths without a barrier as well as reactions paths with a barrier of up to $+90$ kJ·mol^{-1} are present. In total, 32 adsorption paths were converged. In agreement with the adsorption basins direct adsorption paths are observed as long as the inversion of the PH_3 is avoided.

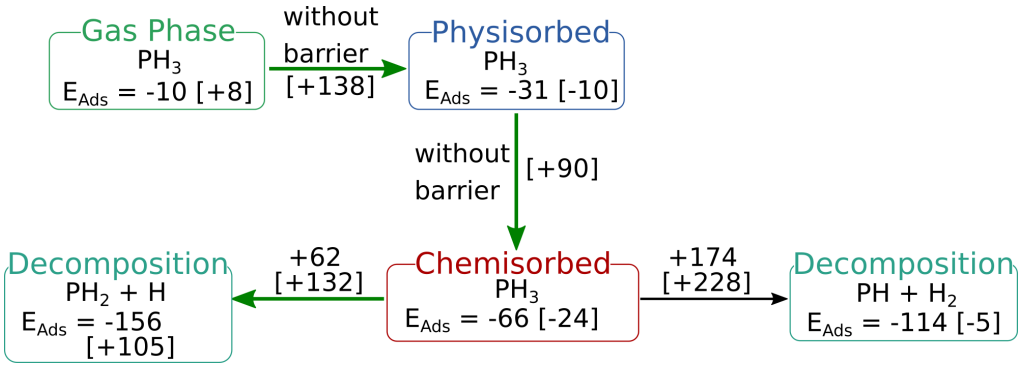


Figure 4.27.: Overview showing the obtained decomposition reactions of PH_3 on GaP(001). Range of adsorption energies E_{Ads} stated at the fragments while range of activation energies stated at the links. The largest value is stated in parentheses. Furthermore, the range of activation energies is stated relative to the most stable fragment structure, respectively. Most likely reaction path marked in green. All values in kJ·mol^{-1}.

For the decomposition of PH_3 only hydrogen transfer reactions R 4.6 and hydrogen elimination reactions R 4.7 were calculated. Also, for PH_3 on GaP(001) only these two types of reactions were then observed by PESE in contrast to GaH_3 for which additional reactions, as for example the transfer of all hydrogen atoms to the surface, were observed. For the hydrogen transfer reaction, barriers in the range of $+62$ to $+132$ kJ·mol^{-1} are obtained based on 28 converged reaction paths, among which only two exothermic reaction steps are observed. Considering all obtained decomposition structures, adsorption energies of -156 to $+105$ kJ·mol^{-1} are present for the decomposition structures showing strongly exothermic as well as endothermic reactions. However, only a minority of six decomposition structures show more attractive adsorption energies relative to the most stable PH_3 adsorption structure of -66 kJ·mol^{-1}. Consequently, the majority of 58 product structures of a hydrogen transfer reactions represent endothermic reactions. For the hydrogen elimination reaction 14 converged paths are obtained with reaction barriers in the range of $+174$ to $+228$ kJ·mol^{-1}. Here, 13 exothermic reaction steps are obtained. In addition, adsorption energies for the decomposition structures of -114 to -5 kJ·mol^{-1} are present whereby a majority of 28 out of 40 adsorption energies is more attractive than the energy of the most stable PH_3 adsorption structure. Overall, as a first decomposition step of PH_3 the hydrogen transfer reaction is due to the smaller barriers most likely. This observation is in agreement to the experimental proposed decomposition mechanism[292,293]. However, as shown by the number of exothermic reaction steps and attractive adsorption energies for product structures, the hydrogen elimination reaction seems to be thermodynamically favored in comparison to the hydrogen transfer reaction even though the most stable decomposition structure is observed for a hydrogen transfer reaction. Furthermore, based on thermodynamic corrections (see section 2.1.7), the formation of H_2 is expected to be even more favorable due to an increasing translational entropy contribution. Still, both decomposition reactions are thermodynamically favored as soon as one of the most stable decomposition structures is reached.

This is a completely different observation in comparison to the gas phase decomposition of PH_3[65]. Here, with $+357$ kJ·mol^{-1} (hydrogen transfer) and $+255$ kJ·mol^{-1} (hydrogen elimination) strongly endothermic reactions are observed. As for GaH_3, it is clearly visible that the surface is able to stabilize the formed decomposition products and thereby leading to thermodynamically favorable decomposition reactions. This effect is observed for different semiconductor surfaces as well. For InP reaction barriers in the range of $+60$ to $+96$ kJ·mol^{-1}[288,289] were observed for the hydrogen transfer reaction and on silicon even smaller barriers of $+44$ to $+73$ kJ·mol^{-1}[285–287] were observed.

Conclusion

The adsorption and decomposition of PH_3 on GaP(001) was studied by PESE. Nine chemisorbed adsorption minima were obtained in the optimization of the adsorption basins representing an adsorption of PH_3 on top of the mixed-dimer, insertion of PH_3 in a Ga-Ga dimer bond and an adsorption at the Ga-atoms in the trench between the topmost Ga-layers. The later structure, shows with $-66\,\mathrm{kJ{\cdot}mol^{-1}}$ the largest adsorption energy for PH_3 on GaP(001). Based on $708\,314$ database structures the basins were generated showing with $28.6\,\%$ the largest basin for another adsorption structure of PH_3 at a Ga-atom in the trench between the topmost layers. The calculation of reaction paths revealed that adsorption paths without as well as adsorption paths with a barrier can be observed based. However, since all adsorption paths with a barrier are showing the inversion of PH_3, it is reasonable to expect additional paths showing a rotation of the PH_3 and consequently a negligible reaction barrier. As the most likely decomposition reaction, the hydrogen transfer reaction was identified. Still, also the hydrogen elimination reaction is accessible at elevated temperatures.

As for GaH_3, it can be expected that the obtained adsorption structures of PH_3 and simple decomposition paths describing the hydrogen transfer R 4.6 or the hydrogen elimination reaction R 4.7 are also found by a thorough manual exploration. Still, for PH_3 on GaP(001) the largest and most complex database for the calculation of the adsorption basins was derived by using PESE. Due to the huge number of $708\,314$ structures, it is evident that such a sampling can not be performed in a manual way. Consequently, new insights are gained by PESE even if further evaluation routines have to be implemented in the future to access all information contained in this dataset.

4.5. Computational Performance of PESE

Besides the capabilities of PESE discussed in the previous sections, its performance is a central. Here, due to the high resource consumption of DFT codes at supercomputing centers extensive benchmarks are performed to fine tune the code performance in consideration of the present hardware[294]. To roughly estimate the additional computational cost of PESE, the time of the workers spent with different tasks was measured.

In Figure 4.28 the total time consumption by the workers is shown for the calculation of the adsorption basins for all studied systems Bi, GaH_3 and PH_3 on GaP(001). The calculation of the adsorption basins is the most complex calculation regarding the communication between the workers and the main core of the evaluator since all data derived by the workers has to be evaluated and added to the database as fast as possible. The computation of decomposition products and also reaction paths is less communication intensive since only the final structures have to be evaluated (decomposition structures) or stored (reaction paths) by the evaluator. Consequently, the calculation of the adsorption basins is a decisive step to evaluate the additional communication overhead caused by PESE.

As shown in Figure 4.28, only two parts contribute significantly to the time consumption of the workers: The time spent in DFT calculations (”DFT”) and the time spent in communication with the main core of the evaluator (”MPI + Idle”). Here, the communication time has two contributions namely the time spent in MPI routines (”MPI”) and the time a worker is waiting

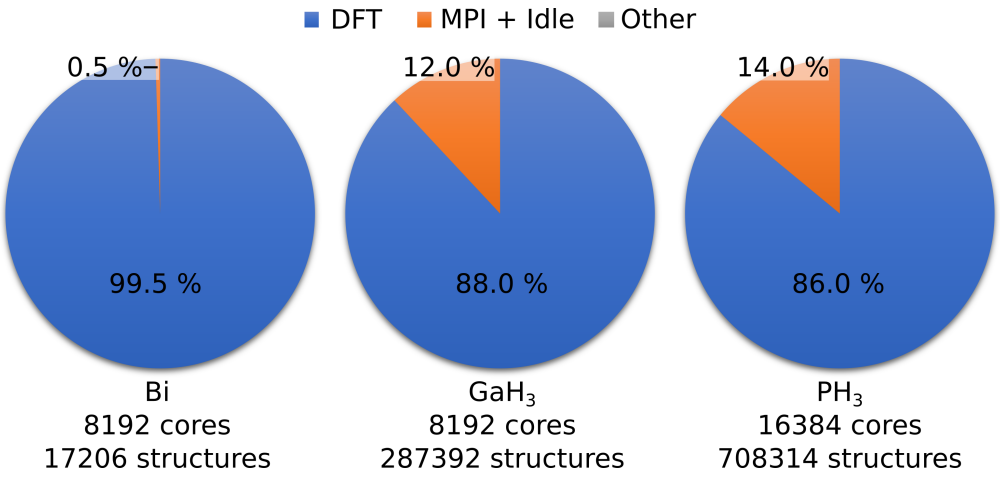


Figure 4.28.: Computational performance of PESE in the basin calculations of Bi, GaH_3 and PH_3 on GaP(001). The time a worker spend with DFT calculations (DFT, blue), MPI routines and idling (MPI + Idle, orange) and with other routines (Other, grey) shown in percentage of the total runtime. In these calculations, each worker was running on 3 nodes and 384 cores of the AMD Epyc 7742 processor.

for a response of the main core of the evaluator (”Idle”). The idle time is a direct consequence of the necessary synchronization between the worker and the evaluator. Unfortunately, the MPI and idle time contributions can not be separated since the time a worker is waiting in a MPI routine to send his results to the evaluator is recognized as MPI time and can not be separated from the actual time for the process of sending all information. However, since a negligible amount of data is exchanged between the workers and the evaluator, it is reasonable to neglect the time consumption of the sending process due to the high bandwidth and low latency of the Infiniband based interconnect[295]. Consequently, the idle time is considered to be the main contribution to the time a worker spent in communication with the main core of the evaluator. All other tasks (”Other”) of the worker, as for example copying all needed files for the DFT calculation to his folder or reading his next task structure, are on their time consumption negligible.

Overall, the time consumption for communication is responsible for 0.5 % (Bi) to 14.0 % (PH_3) of the total time consumption as shown in Figure 4.28. This part is increasing with the system's complexity as indicated by the number of database structures (Bi to GaH_3 and PH_3). An influence regarding the number of used cores can not be drawn from the few data points. Here, an extensive scaling benchmark would be necessary. However, it is visible that PESE can handle $16\,384$ cores with only a minor additional communication overhead of 14 %. With a time consumption of 86 % or more, the DFT calculations remain the decisive performance factor.

In general, a communication overhead can not be avoided for multicore computations[296]. However, a proportion of only 14 % for the communication overhead at this number of cores is a small value in comparison to the parallelization overhead of common quantum chemistry codes[296]. Here, a similar overhead is already reached for less than 100 cores in case a single core is used as reference otherwise a few hundred cores could be added to a calculation to observe a similar increase of the parallelization overhead[296].

The overall performed computations are summarized in Table 4.2. Here, an overview of the created and obtained structures and reaction paths is shown. In this dissertation over one million database structures for the adsorption basins of Bi, GaH_3 and PH_3 on GaP(001) were calculated. Based on 27 minimum structures, 810 possible decomposition structures, 260 decomposition paths and 138 diffusion paths were proposed. Finally, 221 reaction paths were optimized successfully. All these calculations were performed by using up to sixteen thousand cores with only a minor communication overhead of 14 %.

Nevertheless, to enable the calculation of complete reaction networks for precursors as TEGa

and TBP in the future, it is decisive to reduce the computational costs of the DFT calculations. Here, several possibilities emerge: On the one hand, increasing performance could be achieved hardware-wise by using accelerators such as graphic processing units (GPUs)[297–300]. However, this possibility depends on the availability of these components in sufficient numbers at computing centers. On the other hand, PESE can easily be extended to support the combination or replacement of DFT calculations by less demanding models as for example ML approaches[301–304] or extended tight-binding (xTB) methods[305]. Here, a fine balance between accuracy, reliability and computational cost has to be established for the description of group 13 and 15 precursors on semiconductor surfaces.

Table 4.2.: Overview of the created structures and performed calculations by PESE.

System	Bi	GaH_3	PH_3
Basin and Minima			
Initial grid point	810	810	810
Unique orientations	1	12	24
Initial structures	810	9720	19 440
Database structures	17 206	287 392	708 314
Minima structures	241	5376	11 143
Unique minima	204	1549	2199
Chemisorbed structures	14	18	9
Decomposition structures			
Initial structures	—	540	270
Selected structures	—	169	91
Reaction paths			
Diffusion paths	20	48	70
Decomposition paths	—	169	91
Converged paths	19	114	88
Unconverged paths	1	103	73
Ratio conv. / uncon.	19.0	1.1	1.2
Obtained path minima	7	87	41
Obtained transition states	17	142	78
False minima	4	37	34

5

Small Molecule Inhibitors for the Area-Selective Atomic Layer Deposition of Aluminium Oxide and Hafnium Oxide

In the previous part of this thesis PESE was introduced and used to study the reactivity of Bi, GaH_3 and PH_3 on GaP within the CVD process. The second part of this thesis is focusing on the ALD process. As already introduced, the investigation of the ALD process is, similar to the CVD process, motivated by the usage in the manufacturing of electronic devices. Here, due to the miniaturization of electronic devices and circuits the precise fabrication of nanopatterns and their placement have become vital[306–308]. However, the common top-down approaches like lithography are facing problems regarding feature patterning and alignment by dealing with feature sizes below 7 nm[306,309,310]. Therefore, fabrication of such nanopatterns by bottom-up approaches is holding great promises[307,311]. For these area selective deposition (ASD) approaches[308] are heavily investigated. Especially, AS-ALD[312] is of interest due to its self-saturated surface reactions enabling the precise control of material deposition. Consequently, AS-ALD is applied aiming at advanced nanopatterning[39] and feature placement[311]. Here, the elimination of the edge placement error, which describes the difference between the intended and actual alignment of features grown on top of each other, receives special attention[311].

The idea behind AS-ALD is to exploit differences in the reactivity of the used substrate surfaces[39] to obtain selective material growth. By optimizing process conditions and utilized reactants, growth should only occur on the growth surface (GS) while the non-growth surface (NGS) remains untouched. Different approaches are possible to achieve the desired selectivity of the GS over the NGS. In an optimal case the ALD precursors are inherently selective[313] towards a substrate as shown for example for an oxide over nitride surface[314],

metal over oxide surface[315] or even defects over ideal surfaces[313]. However, this inherent selectivity is restricted to very specific combinations of substrates and precursors. Therefore, the possible material compositions are limited. A more flexible approach is to use inhibitor molecules[316] to block the NGS. This approach enables the application of a wide range of material systems since in principle only the inhibitor molecule has to be tuned to the desired substrate. Furthermore, selectivity can be improved by periodically renewing the passivation of the NGS by supplying the inhibitor or by combining the growth cycles with (selective) etch steps[308,317]. Another approach to obtain selectivity is to activate the growth of the material by supplying energy via an electron beam or light or by using catalytically active surfaces[308]. However, this approach is limited by finding catalytically active surfaces or the additional technical challenge of confining the irradiation to the GS by masks[318].

In this thesis the approach to block the NGS by inhibitor molecules was pursued. A common type of inhibitors are molecules forming self assembled monolayers (SAMs)[308,319] (Figure 5.1) by vapor phase[320] or liquid phase[316] processing. The blocking of the NGS is ensured by the large, inert tail groups of the molecules, which support a dense and stable passivation layer[321]. However, due to the costly processing, SAMs are hard to integrate with most vapor-phase processes[322]. Furthermore, the size of SAMs can lead to nonideal growth at interface between the GS and NGS[320,323,324].

In search of easily vaporizable molecules SMIs are proposed as an alternative to SAMs[322]. The great advantage of these SMIs is that they can be integrated in vapor-phase ALD processes. Within this project three different classes of SMIs were investigated and the results are presented in the next sections. The used surface models and the computational methods are introduced in section 5.1 and 5.2, respectively. Subsequently, the application of alkoxysilanes as SMIs is presented in section 5.3 while the usage of methanesulfonic acid (MSA) and diethyl ether (DES) is presented in section 5.4 and 5.5, respectively.

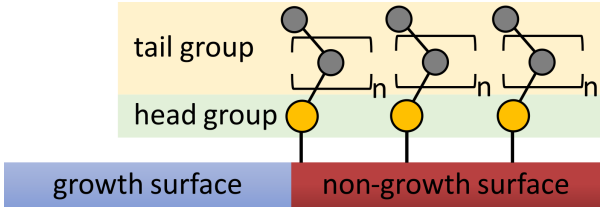


Figure 5.1.: Schematic concept of AS-ALD to block a non-growth surface (NGS) with a self assembled monolayer (SAM) while leaving the growth surface (GS) untouched. As an example for a SAM an alkylthiol is shown. Color code: S (yellow), CH_2 and CH_3 (grey).

5.1. Surface Models for Silicon Oxide and Copper

The reactivity of the studied SMIs was investigated on the SiO_2 and Cu(111) surface. Both materials are highly relevant for the microelectronic industry either as dielectric material (SiO_2)[31] or as conducting material (Cu) for interconnects[325]. In addition, the SiO_2 and Cu surface are heavily studied as exemplary systems for the dielectric on dielectric and dielectric on metal AS-ALD[308].

In Figure 5.2 the computational models for the Cu(111) and SiO_2 surface are shown. The 3×5 Cu(111) slab (Figure 5.2a,b) was derived from an optimized Cu bulk structure with a lattice constant of $a = 3.562$ Å (experimental constant: 3.615 Å[326]). This slab consists of 4 Cu layers, whereby the bottommost Cu layer was frozen in all calculations. In addition, a vacuum of 28.8 Å was added in z-direction to prevent spurious interactions due to PBC.

For SiO_2 a 2×2 (default model) and a 3×3 (used when mentioned) slab model were created from an optimized SiO_2 bulk in the α-quartz phase with parameters of $a = 4.917$ Å and $c = 5.445$ Å (experimental lattice parameters: 4.913 Å and 5.405 Å[327]) by cutting along the (001) plane. As shown in Figure 5.2d the slab contains two O-Si-O layers and a frozen Si-H layer at the bottom. Terminating oxygen atoms at the surface were saturated by hydrogen atoms to reflect a fully hydroxylated α-quartz(001) surface[328–331]. Furthermore, 14.7 (silane SMIs,

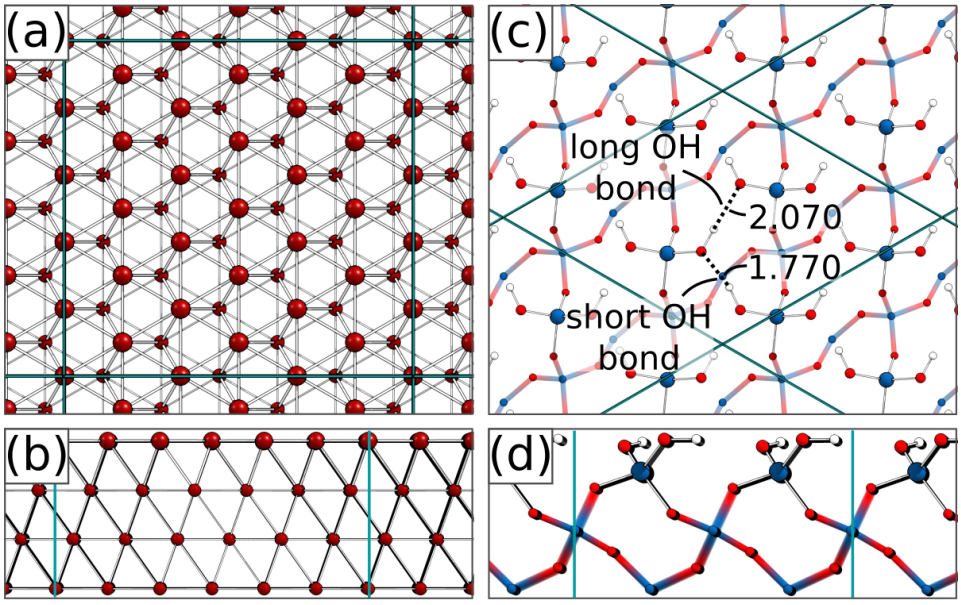


Figure 5.2.: Used surface models of Cu(111) (a,b) and SiO_2(001) (c,d) in top (a,c) and side (b,d) view. OH bond lengths in Å. Unit cell borders shown in teal. For SiO_2 the bottommost H-layer is not shown. Color code: Cu (copper red), Si (blue), O (red), H (white).

MSA) or 19.3 Å (DES) of vacuum was added to the surface model in the z-direction to prevent interactions due to PBC. In Figure 5.2c, it is visible that this surface model contains two distinct adsorption sites namely a long and short OH bond identified based on their bond length. Consequently, all studies where performed with both adsorption sites.

The SiO_2 model presented here is a simplified and computationally less demanding surface model with an OH density of $9.55\,\mathrm{OH/nm^2}$. However, in experiment usually an amorphous SiO_2 surface is present. Here, hydroxyl groups are defined by a distribution of isolated, vicinal, and geminal groups[332] leading to an even smaller density of hydroxyl groups of 4 to $6\,\mathrm{OH/nm^2}$[333–335]. Therefore, our computations serve as an upper bound for the number of hydroxyl groups and thereby the possibility of SMIs to form bonds to the surface.

5.2. Computational Methods

Periodc DFT Calculations

Periodic DFT calculations were performed with the Vienna Ab initio Simulation Package (VASP, version 5.4.4)[233–236]. Here, the functional by Perdew, Burke and Ernzerhof (PBE)[237,238] and the DFT-D3 dispersion correction[90,91] were used. The plane wave cutoff within the PAW approach[86,87] and the k-point grid[239] were converged based on the studied SMI and surface. For the silane based SMIs on SiO_2 a plane wave cutoff of 480 eV and a Γ-centered Γ(331) grid were used, ensuring a remaining inaccuracy of less than $1.5\,\mathrm{kJ{\cdot}mol^{-1}}$ in the total energy. For MSA and DES the adsorption energy on Cu(111) and SiO_2 was considered to individually converge the plane wave cutoff and the k-point grid to a residual inaccuracy of $1\,\mathrm{kJ{\cdot}mol^{-1}}$. Consequently, for MSA on the SiO_2 [Cu(111)] surface a plane wave cutoff of 450 eV [480 eV] and a Γ centered k-point grid of Γ(331) [Γ(661)] were used. For DES a cutoff of 480 eV [480 eV] and a Γ centered Γ(331) [Γ(661)] k-point grid were chosen. All SCF calculations and structure optimizations were converged up to changes in the electronic energy of less than 10^{-6} eV and forces of less than $10^{-2}\,\mathrm{eV{\cdot}Å^{-1}}$, respectively.

Reaction paths were calculated with the NEB[99,100] method introduced in section 2.2.4. Prior to the activation of the climbing image, reaction paths were preoptimized with a force convergence of $10^{-1}\,\mathrm{eV{\cdot}Å^{-1}}$. For the silane based SMIs the Dimer method[336] was applied to refine the transition state in case more than one imaginary mode was observed. For this, the wave function convergence was tightened to 10^{-8} eV. For the calculation of the Gibbs energy (see reference [93]) at $T = 175\,^{\circ}\mathrm{C}$ and $p = 1$ atm, numeric frequencies were calculated

by finite differences with a displacement of 0.01 Å. Here, all atoms except the atoms of the frozen layers were considered in calculation of numeric frequencies. Also, the SCF convergence was tightened to 10^{-7} eV. In case an imaginary mode was obtained, the structure was shifted along the imaginary mode and reoptimized. This approach was repeated until only real modes were obtained.

Gas Phase Calculations

Gas-phase calculations were performed with Gaussian 09 (Revision E.01)[337] and Ahlrichs def2-TZVPP basis set[338]. SCF calculations were converged to a change in electronic energy of less than $10^{-6}\,E_\mathrm{h}$ and structure optimization to forces of less than $4.5 \times 10^{-4}\,E_\mathrm{h}{\cdot}{a_0}^{-1}$ (tight settings). Transition states were obtained by the STQN method[339,340]. Minima and transition states were confirmed by analytic frequency calculations. Gibbs energies were again calculated at $T = 175\,^\circ$C and $p = 1$ atm.

5.3. Alkoxysilane Based Small Molecule Inhibitors for the ALD of Aluminum Oxide

Tuning Molecular Inhibitors and Aluminum Precursors for the Area-Selective Atomic Layer Deposition of Al_2O_3

J. Yarbrough, F. Pieck, D. Grigjanis, I.-K. Oh, P. Maue, R. Tonner-Zech and S. F. Bent

This section is based on a published manuscript[341] with experimental collaborators. The theoretical work on trifunctional SMIs was solely provided by myself, while gas phase calculations regarding the SMI sensibility towards ALD precursor were performed under my supervision by Patrick Maue. Preliminary work on mono- and bifunctional SMIs was conducted by Daniel Grigjanis and completed by myself.

Introduction

For SAMs containing an alkoxysilane headgroup a broad knowledge regarding their blocking mechanism by forming polysiloxanes[342] was already developed. Also, the effect of different blocking groups[343–345] is well known for SAMs in general showing the importance of using methyl groups as blocking group. Furthermore, the effect of tail size regarding blocking performance was already investigated[346]. For SMIs the effects of the reactive and blocking group (Figure 5.3) have not yet been investigated.

Therefore, a thorough study on the effect of blocking group size and number of reactive groups was performed by investigating the alkoxysilane based SMIs shown in Figure 5.3. In a first step the number of methoxy groups was changed from one (methoxytrimethylsilane (MTMS)) to three (trimethoxymethylsilane (TMMS)) to identify the influence of the reactive groups. Here, especially the trifunctional SMIs are promising since similar head groups have shown good performance for SAMs[324,347,348]. In a second step, the effect of the blocking group size was studied by varying the blocking group from a methyl (TMMS) to a propyl group (trimethoxypropylsilane (TMPS)). As the target system the growth of Al_2O_3 with copper as the GS and SiO_2 as the NGS was investigated. The use of a native copper oxide was discarded since no selectivity could be achieved in comparison to the SiO_2. Furthermore, by using trimethylaluminium (TMA) or triethylaluminium (TEA) as Al-precursors the ALD chemistry could be adjusted to the SMI.

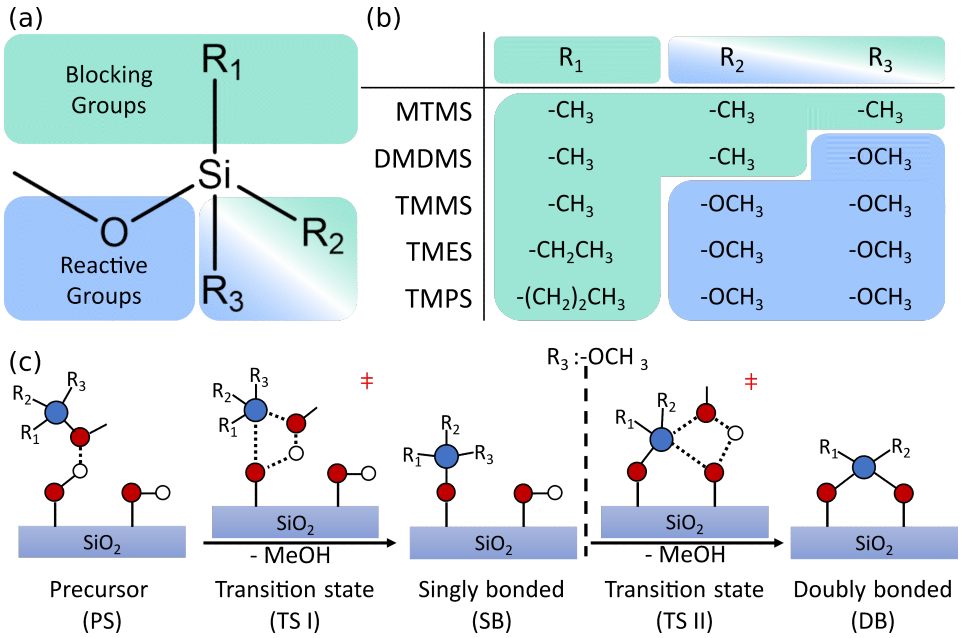


Figure 5.3.: General structure (a) of the studied SMIs. Depending upon the reactive and blocking groups (b), methoxytrimethylsilane (MTMS), dimethoxydimethylsilane (DMDMS), trimethoxymethylsilane (TMMS), trimethoxyethylsilane (TMES) or trimethoxypropylsilane (TMPS) is obtained. Schematic structures of the calculated condensation reactions of the SMIs with OH groups of the SiO_2 surface (c). Color code: Si (blue), O (red), H (white). Reproduced with permission from reference [341].

SMI Reactivity on SiO_2

In this project experimental data[341] obtained by ellipsometry and X-ray photoelectron spectroscopy showed a superior blocking performance of the SMIs in the growth of Al_2O_3 when increasing the number of reactive groups from MTMS to TMMS. However, no significant difference in blocking performance was observed for TMMS, trimethoxyethylsilane (TMES) and TMPS. This observation is surprising since for SAMs larger blocking groups showed superior blocking of HfO_2[346].

To further understand the influences of the reactive and blocking groups on the SMI-surface interactions and the formation of the blocking layer, DFT calculations were performed. In the first part, the adsorption and reactivity of the studied SMIs on the ideal SiO_2 surface is introduced. Here, up to two subsequent condensation reactions were calculated. For this model surface, no third hydroxyl group is sufficiently near a doubly bonded SMI to enable a third condensation reaction. As already discussed, our computations serve as an upper bound for the possibility of SMIs to form several bonds to the surface. In the experiment, it

is unlikely that all methoxy groups of a trifunctional SMI will be involved in reactions with the surface across the entire substrate. However, unreacted methoxy groups may still participate in reactions with nearby SMIs. Hence, the possibility of cross-linking between trifunctional SMIs is also considered in a second part.

Following the general reaction shown in Figure 5.3, we calculated the precursor structure (PS) and the first condensation reaction to a singly bonded structure (SB) with the corresponding transition state (TS I) for each SMI, as well as the subsequent condensation reaction to a doubly bonded structure (DB) with the corresponding transition states (TS II) where relevant. The

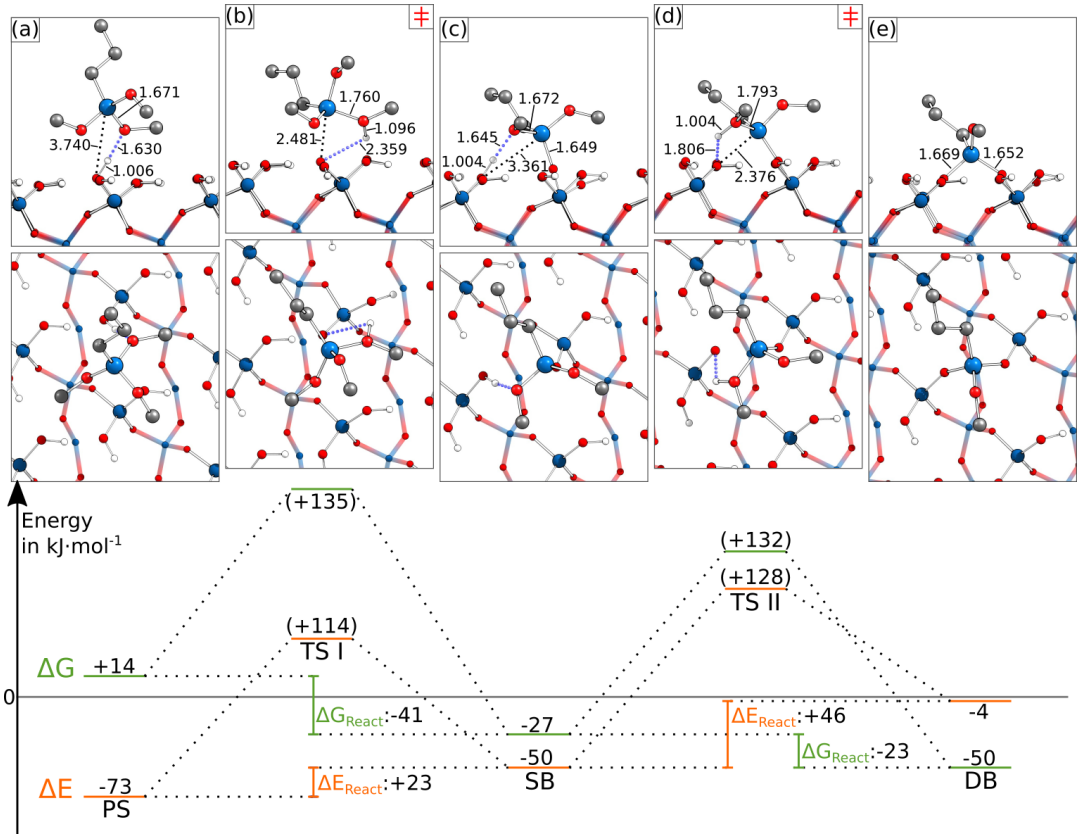


Figure 5.4.: Adsorption and reaction structures of TMPS on SiO_2 with the corresponding energy profile. Top and side view of the precursor structure (PS) in (a), the singly bonded and doubly bonded condensation reaction products (SB, DB) in (c) and (e), and the corresponding transition states (TS I, TS II) in (b) and (d). Selected bond lengths in Å. The electronic (ΔE) adsorption (activation) and reaction (ΔE_{React}) energies are shown in orange and the respective Gibbs energies (ΔG, ΔG_{React}, 175 °C, 1 atm) in green. Activation energies are stated relative to the previous minimum in brackets, while reaction energies include an already desorbed methanol. Color code: Si (blue), O (red), H (white), C (grey). Adapted reprint with permission from reference [341].

surface contains two O_{surf}-H bonds (surface oxygen atoms are indicated by the index "surf") – a short and a long one (Figure 5.2). The reactivity of the short and long OH bond was probed with dimethoxydimethylsilane (DMDMS)[341] and the most likely reaction steps were then investigated with all SMIs. Hence, in all calculations for an SB structure, reaction with the long OH bond was investigated, while for the subsequent second condensation reaction a neighboring short O_{surf}-H bond was used.

A prototypical reaction scheme is shown in Figure 5.4 with TMPS on SiO_2. The PS (Figure 5.4a) is shown with the formation of hydrogen bonds between a methoxy group of the SMI and a hydroxyl group at the surface, a common structural motif for all SMIs. Consequently, an exothermic adsorption energy of -73 kJ·mol^{-1} is obtained for TMPS. The transition state for the first condensation reaction (TS I, Figure 5.4b) is then obtained by moving the silicon atom of the SMI above the surface oxygen atom following by significant shortening of the Si-O_{surf} bond between SMI and surface (Δr(Si-O_{surf}): -1.259 Å). Simultaneously, the formation of methanol is already indicated by the proton transfer to the methoxy group (Δr(O-H): -0.534 Å), together with elongation of the Si-O bond within the SMI (Δr(Si-O): $+0.089$ Å) and the former O_{surf}-H bond (Δr(O_{surf}-H): $+1.353$ Å). These structural changes result in an energy barrier of $+114$ kJ·mol^{-1}. The release of methanol and the formation of the Si_{SMI}-O_{surf} bond leads to SB (Figure 5.4c). The desorption of methanol when forming SB ($\Delta E_{desorption,\ methanol}$ = $+65$ kJ·mol^{-1}) leads to an overall endoenergetic reaction (ΔE_{React}(PS→SB): $+23$ kJ·mol^{-1}). However, considering the Gibbs energy of reaction (ΔG_{React}(PS→SB): -41 kJ·mol^{-1}), desorption of a methanol molecule leads to this reaction being exergonic. Furthermore, once the desorbed methanol is purged out of the reaction chamber, any back reaction is prevented.

The second condensation reaction (SB→DB, Figure 5.4c-e) shows the same characteristics as the first. The silicon atom of the SMI is moved closer to an O_{surf} atom of a hydroxyl group at the surface while the formation of methanol is visible. Since the SMI is already attached to the surface and therefore less flexible, the reaction barrier for TS II is increased to $+128$ kJ·mol^{-1} ($+14$ kJ·mol^{-1} higher than that of TS I). After the release of methanol, DB (Figure 5.4e) is obtained. As for the first condensation reaction, an endoenergetic reaction is present for the second condensation reaction (ΔE_{React}(SB→DB): $+46$ kJ·mol^{-1}) while considering Gibbs energies yields an exergonic reaction (ΔG_{React}(SB→DB): -23 kJ·mol^{-1}).

Based on the electronic energies (ΔE), the most stable structure is PS, while subsequent condensation reactions are endothermic. The exact opposite is observed when we include thermodynamic corrections (ΔG). The loss of translational and rotational entropy upon adsorption significantly destabilizes PS. Together with the entropy gain by methanol desorption, this

Table 5.1.: Electronic and Gibbs (shown in parentheses) free adsorption energies for the precursor structure (PS), singly bonded structure (SB) and doubly bonded structure (DB). Corresponding activation energies for subsequent condensation reactions (TS I: PS→SB, TS II: SB→DB) of the studied SMIs with a SiO_2 surface are also shown. All energies in kJ·mol^{-1}. Date from reference [341].

	PS	TS I	SB	TS II	DB
MTMS	-62 (17)	119 (140)	-26 (-17)		
DMDMS	-71 (14)	106 (114)	-32 (-21)	143 (150)	1 (-59)
TMMS	-70 (16)	121 (134)	-40 (-15)	133 (132)	6 (-40)
TMES	-72 (14)	117 (136)	-43 (-17)	128 (128)	0 (-52)
TMPS	-73 (14)	114 (135)	-50 (-27)	128 (132)	-4 (-50)

renders subsequent reactions exergonic. Taking into account the formation and desorption of methanol is thus crucial for favorable thermodynamics and for driving the mechanism towards multiple reactions of the SMI with the SiO_2 surface.

The adsorption and activation energies for all other SMIs are provided in Table 5.1 and the structures are shown in Figure D.1 and D.2. In comparison to TMPS, the reaction structures of the other SMIs show similar features. Larger molecules show stabilization due to larger dispersion interactions which mainly affect the $\triangle$E values. Since larger molecules lose more translational entropy upon adsorption, only minor differences remain in the Gibbs energies of PS. For SB and DB structures, a weak trend towards more exergonic Gibbs energies is present with increasing size of blocking group (TMMS to TMPS) while no trend is present for increasing the number of reactive groups (MTMS to TMMS). In a similar fashion, a stabilization for TS I and TS II is present in the electronic energies for larger blocking groups while this effect vanishes in the Gibbs energy.

Despite these minor differences in the reaction energies, all SMIs show an identical reaction mechanism with the SiO_2 surface. Therefore, the superior blocking properties of trifunctional SMIs cannot be explained by a change in the reaction mechanism. However, following the initial idea of passivating all hydroxyl groups, bi- and trifunctional SMIs can saturate several closely adjacent hydroxyl groups without resulting in a high coverage of SMIs which would lead to intermolecular repulsion. This is especially advantageous for surface areas with a high density of hydroxyl groups. Therefore, we attribute the superior blocking of bi- and trifunctional SMIs in comparison to the monofunctional SMI in part to the ability to more extensively react with surface hydroxyls for the same coverage of SMIs.

Possibility of Siloxane Formation

As shown by these calculations and already discussed, not all reactive groups are reacting with the SiO_2 surface. Therefore, as a second possibility to explain the superior blocking of trifunctional SMIs, the possibility for siloxane formation was investigated with TMPS. The reaction barrier for this reaction was calculated for two cases: In the first case, both SMIs already participated in a condensation reaction with the SiO_2 surface, the product of which is shown in Figure 5.5a. In the second case, one SMI already participated in a condensation reaction with the SiO_2 surface, while the second SMI is bound only by hydrogen bonding to the surface (Figure 5.5b). With values of $+255$ kJ·mol^{-1} (case 1) and $+297$ kJ·mol^{-1} (case 2), very large activation barriers were obtained for the condensation reaction in either configuration, excluding the possibility that direct reaction of two TMPS molecules takes place. Furthermore, the adsorption energies of the reactants for the condensation are $+29$ kJ·mol^{-1} (case 2) to $+41$ kJ·mol^{-1} (case 1) higher in energy than the sum of two isolated TMPS molecules. This shows that an energy penalty resulting in steric repulsion between the TMPS molecules must be overcome to initialize the reaction in addition to the large reaction barriers.

However, under the experimental conditions it is possible that the amorphous SiO_2 retains some water[349,350]. Therefore, a second scenario was investigated for the siloxane formation. Here, the calculations for siloxane formation were repeated after first allowing one methoxy group of a covalently bonded SMI to be hydrolyzed by water. The necessary activation energy for the hydrolysis reaction will be discussed later. For the siloxane formation,

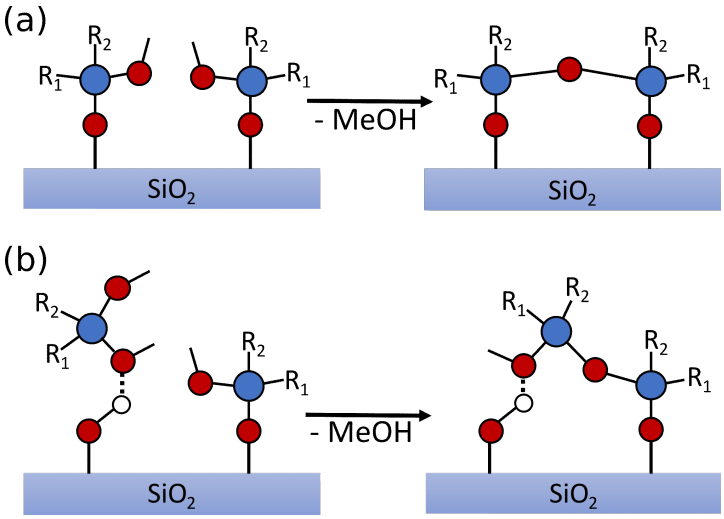


Figure 5.5.: Schematic illustration of the studied siloxane formation reactions. (a) Condensation reaction between two singly bonded SMIs. (b) Condensation reaction in case only one SMI is a singly bonded structure. Color code: Si (blue), O (red), H (white).

the initial hydrolysis drastically reduced the activation barriers to $+133$ kJ·mol^{-1} (case 1) and $+144$ kJ·mol^{-1} (case 2). These barriers are low enough to allow the possibility of siloxane formation under experimental conditions. Furthermore, the exchange of a methoxy group by a hydroxyl group reduces the steric repulsion between the adsorbates. Here, the energy penalties due to steric repulsion of up to $+41$ kJ·mol^{-1} as discussed above can be understood as an upper bound. Hence, these results suggest that surface-level cross-linking between chemisorbed silane molecules may be considered for residual methoxy groups on surface-bound SMIs.

Thus, given the superior blocking ability of trifunctional silanes, our hypothesis is that their performance can be attributed to their ability to block several active sites at the surface at once combined with the potential for surface-level cross-linking compared to monofunctional and bifunctional silanes.

For comparison, all SMIs were also studied on an ideal Cu(111) surface[341]. Here, only dispersion-bound structures without the possibility for a consecutive reaction with the surface were found. Therefore, we do not expect stable SMI films on copper at the experimental temperatures since desorption of SMIs will take place. These differences in the surface reactivity match the experimental observations[341].

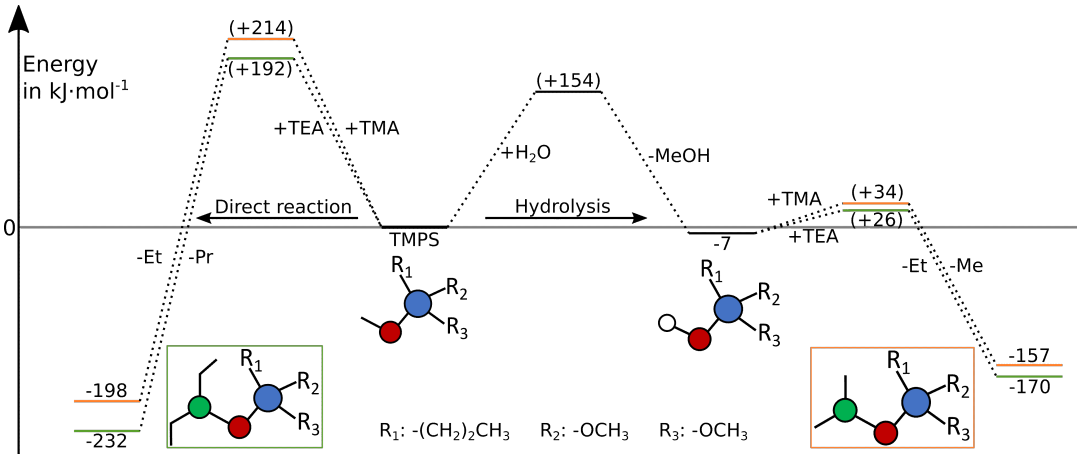


Figure 5.6.: Possible reaction channels of TMPS in gas phase calculations with water, TMA and TEA. Gibbs reaction (activation) energies at $175\,^{\circ}$C and 1 atm in kJ·mol^{-1}. Main products are schematically shown. Color code: Si (blue), O (red), H (white), Al (green). Activation and reaction energies are stated relative to the previous minimum. Reproduced with permission from reference [341].

Sensitivity of SMIs Towards ALD Precursors

While trifunctional SMIs showed the best blocking performance, the influence of the Al-precursor on the weakening of the blocking layer was investigated in an additional part of this project. Here, by experimental data TMA was identified to destroy the blocking layer while better selectivity was observed for TEA[341]. This observation could be explained by the larger size[351] of TEA in comparison to TMA and a thereby slower diffusion into the blocking layer.

To better understand the interactions of TMA, TEA, and water with TMPS - interactions that may lead to degradation of the TMPS blocking layer — DFT calculations were performed. These calculations focused on gas-phase reactivity of TMPS with these three species (Figure 5.6). For the direct reaction of TMA and TEA with TMPS, a high reaction barrier of at least $+192\,\mathrm{kJ{\cdot}mol^{-1}}$ was found. On the contrary, the hydrolysis reaction of TMPS with water shows a smaller barrier of $+154\,\mathrm{kJ{\cdot}mol^{-1}}$. Furthermore, the generated hydroxyl group is highly reactive towards the Al-precursor shown by barriers smaller than $+34\,\mathrm{kJ{\cdot}mol^{-1}}$. Therefore, these calculations indicate that the SMI is more sensitive towards attack by water than attack by the ALD precursor. This is not unexpected since the SMIs are designed to be reactive towards surface hydroxyl groups. However, these gas phase calculations are not sufficient to explain the reduced reactivity of TEA in comparison to TMA. We attribute this to an artifact of the model system since it cannot account for the steric hindrance that a layer of adsorbed TMPS would create. Consequently, more demanding slab calculations are necessary in the future to account for steric surface effects on the reactivity of SMIs towards the Al-precursors.

Conclusion

Adsorption and reactivity of five alkoxysilane based SMIs were studied for the AS-ALD of Al_2O_3 on copper over silicon dioxide. In agreement with experiment, DFT calculations show that not the reaction mechanism is decisive for the different blocking behavior but the number of potential reactions with the surface. Furthermore, the possibility for siloxane formation was investigated. Overall, the superior blocking of trifunctional SMIs is attributed to both the possibility to passivate several reactive groups and the possibility to form an on surface network. In addition, based on gas phase calculations we could identify the coreactant as the crucial molecule to undermine the blocking layer, while the experimental proof of TEA being the better Al-precursor could not be explained due to the limitations of the theoretical model. Also, the defect formation and depletion mechanism of the blocking layer is yet not fully understood. Overall this joint theoretical and experimental study clearly shows the importance of tuning all parameters of the SMI and ALD process to enhance area selective growth.

5.4. Methanesulfonic Acid as the Small Molecule Inhibitor

Area-Selective Al_2O_3 Atomic Layer Deposition with Molecular Deactivation of Copper by Methanesulfonic Acid

J. Yarbrough, F. Pieck, A. Shearer, P. Maue, R. Tonner-Zech and S. Bent

This section is based on a manuscript[352] in preparation with experimental collaborators. Herein, the theoretical work on the reactivity of MSA on Cu(111) was solely performed by myself, while the reactivity investigations of MSA on SiO_2 was acquired by Patrick Maue under my supervision.

Introduction

Besides the alkoxysilane based headgroup introduced in the previous section, common headgroups of molecules forming SAMs are based on phosphonic acid[316,321] or thiols[353,354]. These headgroups are utilized to use metals as the NGS. However, phosphonic acid based SAMs have shown low blocking performance in the growth of Al_2O_3[316]. With MSA a new class of

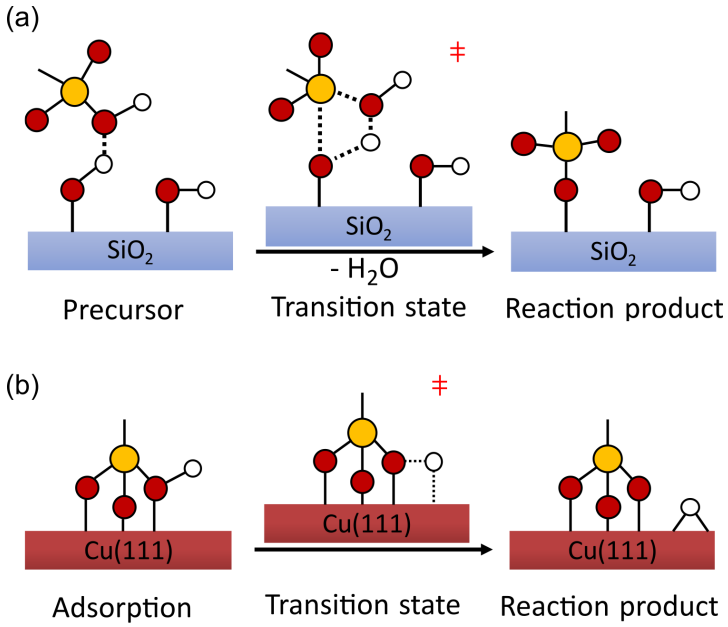


Figure 5.7.: Schematic reactivity of MSA on SiO_2 and Cu(111). Condensation reaction of MSA with SiO_2 in (a) and deprotonation reaction of MSA on copper in (b). Color code: S (yellow), O (red), H (white).

SMI was investigated aiming at the area selective growth of high-κ dielectrics, such as Al_2O_3 on ruthenium, titanium dioxide and silicon dioxide as the GS, while blocking copper and copper oxide as the NGS. For the growth of Al_2O_3, dimethylaluminum isopropoxide (DMAI) and TMA are investigated as possible Al-precursors. Within this project, DFT calculations were performed to understand the different surface reactivity of MSA on copper and silicon dioxide. These calculations followed the schematic reactions of Figure 5.7. For SiO_2, a condensation reaction of MSA with a hydroxyl group of the SiO_2 surface was considered (Figure 5.7a) similar to the alkoxysilane based SMIs, while a deprotonation reaction was investigated for the Cu(111) surface (Figure 5.7b).

MSA Reactivity on Cu(111)

For MSA on Cu(111) the adsorption and reaction structure are shown in Figure 5.8. In addition, less stable product structures are contained in the appendix (Figure D.4). In the initial adsorption structure (Figure 5.8a) MSA is bound to the Cu(111) surface by its oxygen atoms leading to an adsorption energy of -73 kJ·mol^{-1}. With a contribution of -64 kJ·mol^{-1} this interaction is mainly due to dispersion interactions. However, by overcoming a minor barrier of $+29$ kJ·mol^{-1} (Figure 5.8b) MSA can transfer its proton to the Cu(111) surface. This exother-

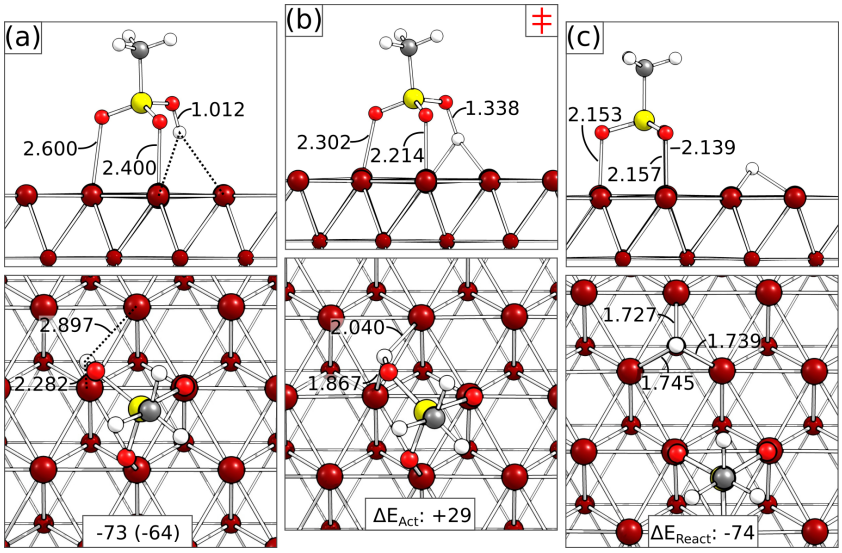


Figure 5.8.: Adsorption and reaction structure of MSA on Cu(111). (a) Initial adsorption structure of MSA. (b) Observed transition state for the deprotonation reaction leading to (c) the product structure. Adsorption energies with their dispersion contribution (shown in parentheses), activation energies (ΔE_{Act}) and reaction energies (ΔE_{React}) in kJ·mol^{-1}. Selected bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Cu (copper red).

mic reaction (ΔE_{React}: -74 kJ·mol^{-1}, Figure 5.8c) leads to a strongly bound product structure with significant electronic interactions between MSA and Cu(111) (ΔE_{Ads}: -147 kJ·mol^{-1}, ΔE_{Elec}: -68 kJ·mol^{-1}). This stronger interaction is in line with the changes in bond lengths along the reaction paths: The O-Cu bonds between MSA and the Cu(111) surface are significantly shortened starting from up to 2.600 Å (Figure 5.8a) to 2.157 Å (Figure 5.8c). Simultaneously, also the H-Cu bonds are shorted from 2.897 Å (Figure 5.8a) to 1.745 Å (Figure 5.8c) while the former O-H bond is elongated from 1.012 Å (Figure 5.8a) to 1.338 Å (Figure 5.8b). Overall, due to the minor reaction barrier the deprotonation of MSA is expected for an adsorption on metallic Cu(111). The thereby formed hydrogen can either diffuse into the bulk[355] or desorb as H_2. Furthermore, to estimate the packing of deprotonated MSA molecules a second deprotonated MSA was placed next to the first one (Figure D.4). Since only a minor repulsion of $+7$ kJ·mol^{-1} is observed, a densely packed MSA layer can be expected.

The observed deprotonation reaction for MSA on copper is in line with basic textbook knowledge for reactions between a strong acid and metal. Consequently, for the interaction of MSA with copper oxide the common type of salt formation reaction along with the release of water as shown in equation R 5.1 could be expected.

$$Cu_2O + 2\underset{\text{MSA}}{(CH_3)SO_3H} \longrightarrow 2\,CuO{-}SO_2(CH_3) + H_2O \qquad \text{(R 5.1)}$$

However, experimental evidence for a reduction of the copper oxide by MSA exists[352]. This can not be explained by the reaction in equation R 5.1. Therefore, we formulate the hypothesis that MSA can decompose and reduce the copper oxide upon adsorption. This could follow the reaction proposed in equation R 5.2.

$$Cu_2{}^{+\mathrm{I}}O + (CH_3)S^{+\mathrm{IV}}O_3H \longrightarrow 2\,Cu^{\pm 0} + S^{+\mathrm{VI}}O_3 + HOCH_3 \qquad \text{(R 5.2)}$$

Following this assumption, MSA is able to reduce the topmost copper oxide layer to metallic copper and block this metallic copper by forming a dense packed MSA layer. This would be in agreement to the experimental observation that MSA is able to block a copper as well as a copper oxide surface[352].

MSA Reactivity on SiO_2

In Figure 5.9 the corresponding adsorption structures for MSA on SiO_2 are shown. As for the alkoxysilane based SMIs different adsorption sites are available on the SiO_2 surface. For MSA on SiO_2, the most stable adsorption structure with an adsorption energy of -75 kJ·mol^{-1}

(Figure 5.9a) is obtained in case MSA adsorbs at a long OH bond while the weakest structure is obtained for the short OH bond (-53 kJ·mol^{-1}, Figure 5.9b). Another type of adsorption structure is observed in case an oxygen atom and the hydroxyl group of MSA are both used for hydrogen bonding, leading to adsorption energies of up to -73 kJ·mol^{-1} (Figure 5.9c). All structures have in common the prominent motif of hydrogen bonding between MSA and OH groups of the SiO_2 surface leading to a significant electronic interaction for the adsorption energies.

Starting from MSA in the long and short OH bond a condensation reaction was calculated. For the structure in Figure 5.9c no reaction path for a condensation reaction was found. However, a change of the conformation to one of the other structures can be expected prior to a condensation reaction. The reaction paths obtained by the NEB method are shown in Figure 5.10 (long OH bond) and Figure D.5 (short OH bond). Since both reactions are nearly identical, only the reaction starting from the long OH bond is discussed here. For this reaction a different conformer for the adsorption of MSA in the long OH bond was identified as the starting point (Figure 5.10a). However, the change in adsorption energy of $+2$ kJ·mol^{-1} in comparison to the most stable conformer is neglectable. The transition state (Figure 5.10b) of the condensation reaction is reached by preforming water. This is accompanied by an elongation of the O_{surf}-H (Δr: $+1.451$ Å) and the O_{MSA}-S bond (Δr: $+0.629$ Å) and a shortening of the O_{MSA}-H bond (Δr: -0.801 Å). Overall, these structural changes lead to a large activation energy

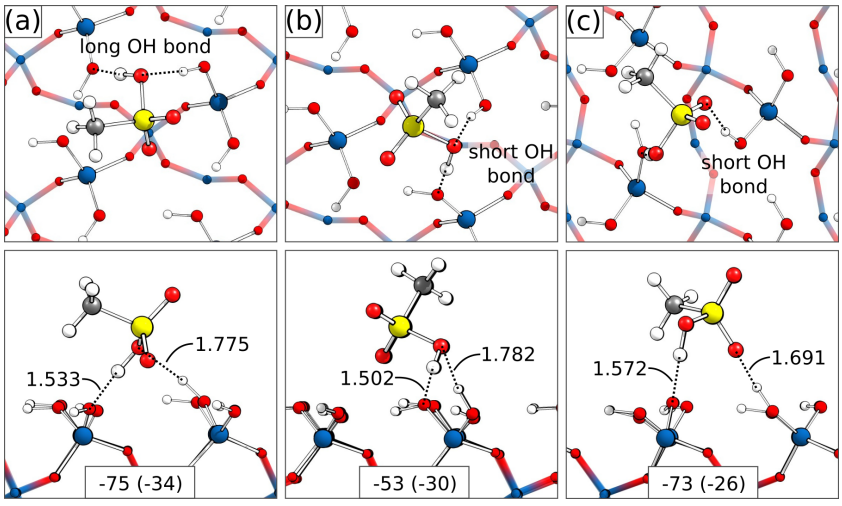


Figure 5.9.: Different adsorption structures of MSA on SiO_2. (a) Adsorption of MSA with its hydroxyl group in a long OH bond or (b) a short OH bond. (c) Adsorption of MSA with an oxygen atom in a short OH bond. Adsorption energies with their dispersion contribution (shown in parentheses) in kJ·mol^{-1}. Selected bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

of $+234\,\mathrm{kJ{\cdot}mol^{-1}}$. Consequently, for MSA on SiO_2 a desorption of MSA is more likely than the condensation reaction. After the release of water the final structure (Figure 5.10c) is obtained. The reaction is slightly endothermic with $+33\,\mathrm{kJ{\cdot}mol^{-1}}$. In comparison to Cu(111), a less favorable reaction with a significant larger reaction barrier is obtained for MSA on SiO_2. This is in line with the experimental observation of a reduced adsorption and worse blocking of MSA on SiO_2 in comparison to Cu(111)[352].

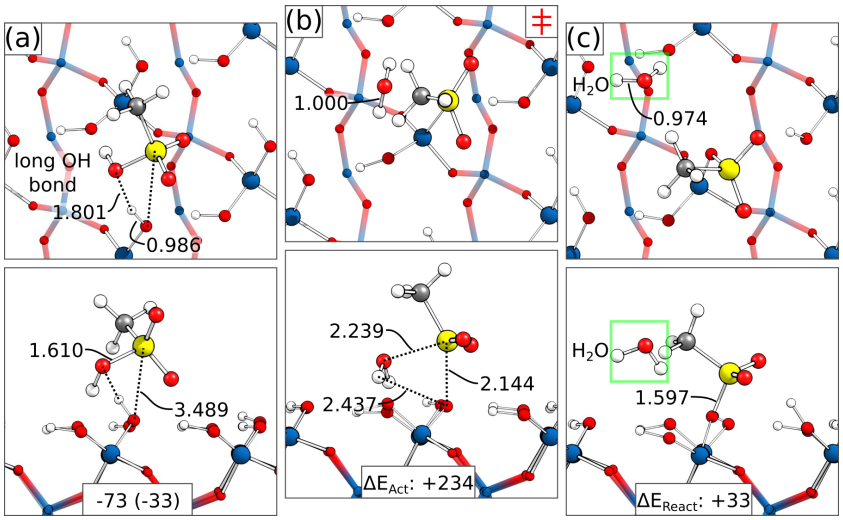


Figure 5.10.: Adsorption and reaction structure of MSA on SiO_2. (a) Starting structure of the reaction path with MSA adsorbed in a long OH bond. (b) Observed transition state for the condensation reaction leading to (c) the product structure. Adsorption energies with their dispersion contribution (shown in parentheses), activation energies (ΔE_{Act}) and reaction energies (ΔE_{React}) in $\mathrm{kJ{\cdot}mol^{-1}}$. Selected bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

Conclusion

Adsorption and reactivity of MSA was studied for the AS-ALD of Al_2O_3 on silicon oxide over copper. In agreement with the experimental data, DFT calculations showed the different reaction mechanisms for MSA on both surfaces. For silicon oxide, condensation reactions with high barriers and unfavorable thermodynamics were obtained, while on copper deprotonation reactions with a minor reaction barrier and favorable thermodynamics were observed. These differences in reactivity can explain the experimental observations that a dense blocking layer of MSA is only formed on copper and not on silicon oxide. Furthermore, a reaction mechanism for the oxidation of MSA was hypothesized to explain the observed reduction and blocking of the native copper oxide.

5.5. Diethyl sulfide for the blocking of copper and silicon oxide

Organothiol Inhibitor Instigated Passivation of Multiple Substrates for Area Selective Atomic Layer Deposition of HfO_2

S. Zoha, B. Gu, F. Pieck, R. Tonner-Zech and H.-B.-R. Lee

This section is based on a manuscript[356] in preparation with experimental collaborators. All theoretical work based on DFT calculations herein was conducted by myself.

Introduction

The previous sections focused on SMIs with either a alkoxysilane or acid based headgroups. As already introduced thiol based headgroups are also prominent[353,354]. Consequently, different thiols like hexadecanethiol[357], dodecanethiol[358–361] or ethanethiol[362] were studied regarding their potential to form SAMs or their usage in the AS-ALD. Especially, the investigation of ethanethiol[362] to grow aluminum oxide on silicon oxide as the GS over copper and cobalt as NGSs is of interest since ethanethiol can also be attributed to the class of SMI. In this project, with DES a structural very similar molecule is studied as SMI. Previous studies on the growth of MoS_2 using DES as a S-precursor already showed that DES is decomposing above $600\,^{\circ}C$ by breaking the S-C bond and releasing ethyl fragments[363]. Furthermore, those ethyl fragments released by DES were attributed to inhibit the growth of MoS_2 on SiO_2[364]. Consequently, DES might have the potential to block metal and silicon oxide surfaces due to the formation of ethyl and thiol fragments. To investigate this idea the blocking of hafnium oxide by DES on copper and silicon oxide as NGS over titanium nitride as GS was investigated.

In Figure 5.11 a schematic overview of the reactions studied is shown. For the SiO_2 surface (Figure 5.11a) three different types of reactions were considered and labeled based on the formed fragment: In the first reaction a proton of a surface OH group is transferred to the DES, while an alkyl fragment is passivating the surface oxygen. In this reaction ethanethiol is formed as a side product. This type of reaction was used to explain the growth inhibition of MoS_2[364]. In the second possible reaction ethane is formed as the side product, while the thiyl fragment is now passivating the surface oxygen. In the third considered reaction, the whole surface OH group is exchanged by the thiol fragment. In this reaction ethanol is formed. For DES on Cu(111) (Figure 5.11b) only one type of reaction was studied. Here, DES is decomposing after adsorption into an ethyl and thiyl radical. To estimate which radical stays bonded to the Cu(111) surface their adsorption energies as well as their recombination to butane and

ethyldisulfide were studied. The reactivity of DES on titanium nitride was not studied by DFT since no experimental evidence for a strong adsorption or reaction was found[356].

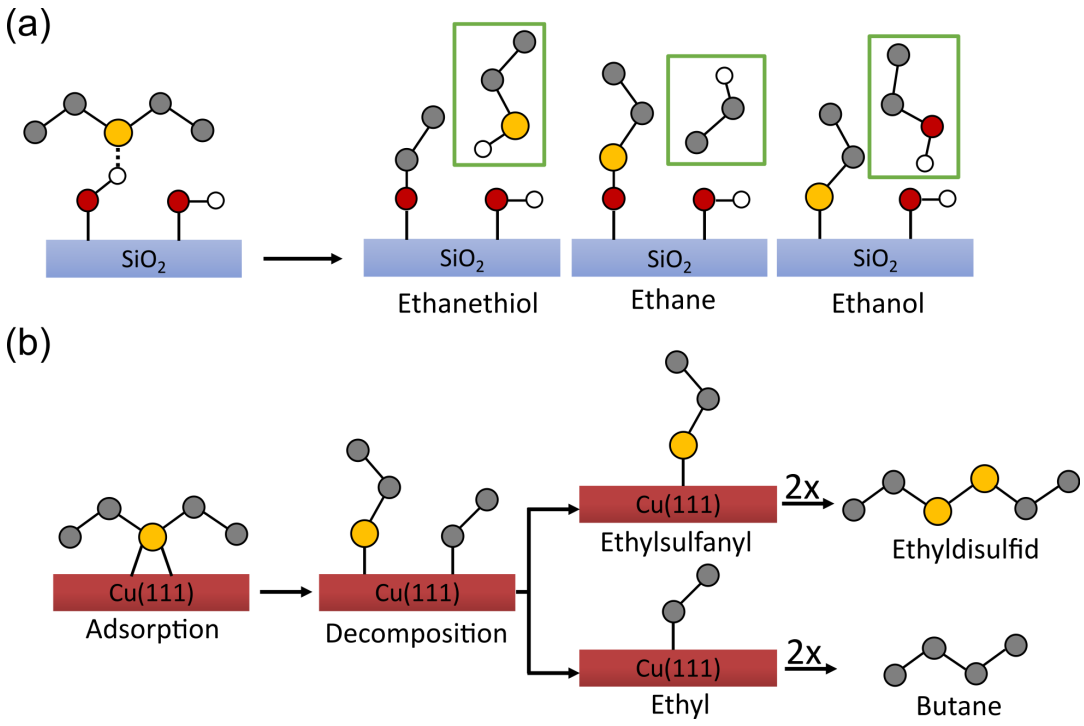


Figure 5.11.: Schematic representation of the studied reactions between DES and the SiO_2 and Cu(111) surfaces. (a) For DES on SiO_2 three different reactions were studied and labeled based on the formed fragment: Formation of ethane, of ethanol and of ethanethiol. (b) For DES on Cu(111) the decomposition of DES and the recombination of the created radicals were studied. Color code: S (yellow), O (red), H (white), CH_2 and CH_3 (grey).

DES Reactivity on SiO_2

To calculate the reaction energies of DES on SiO_2 the corresponding adsorption structures were calculated first. In Figure 5.12 the obtained adsorption structures of DES on SiO_2 are shown. The weakest type of adsorption structure is a physisorbed structures with an adsorption energy of $-30\,\mathrm{kJ{\cdot}mol^{-1}}$ (Figure 5.12a). As indicated by the contribution of the dispersion energy of $-29\,\mathrm{kJ{\cdot}mol^{-1}}$ this structure is solely bound by dispersion interactions. More strongly bonded adsorption structures were observed as soon as hydrogen bonds between DES and a long (Figure 5.12b) or short (Figure 5.12c) surface OH bond are formed. Here, the most stable adsorption structure of DES is found for a long OH bond with an adsorption energy of $-61\,\mathrm{kJ{\cdot}mol^{-1}}$.

Starting from these adsorption structures for the long and short OH bond all introduced reactions (Figure 5.11a) between DES and the surface OH group were studied. Since both ad-

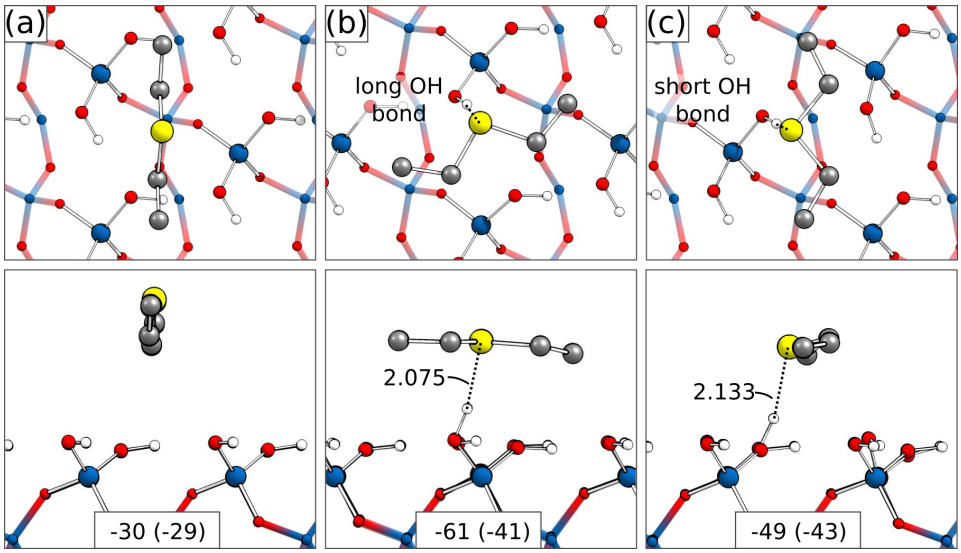


Figure 5.12.: Adsorption of DES on SiO_2 in (a) a physisorbed structure, (b) the long OH bond and (c) the short OH bond. Adsorption energies stated with their dispersion contribution (shown in parentheses) in kJ·mol^{-1}. Selected bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

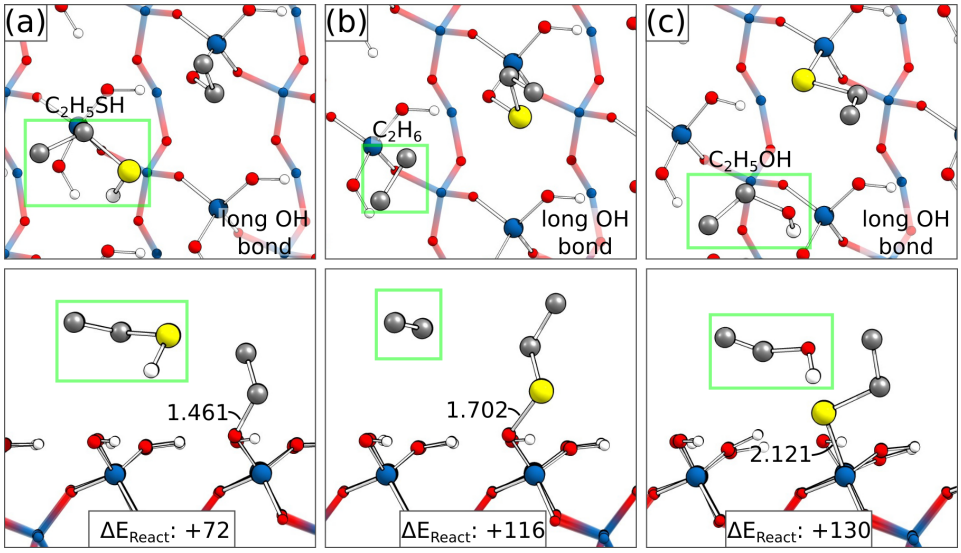


Figure 5.13.: Reaction products of DES on SiO_2 formed at the long OH bond. (a) ethanethiol formation, (b) ethane formation and (c) ethanol formation. The formed molecules are marked in green. Reaction energies are stated relative to the most stable adsorption structure at the long OH bond. Energies in kJ·mol^{-1} and bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

sorption sites agree in their trend, reaction structures for the long OH bond are shown in Figure 5.13, while structures for the short OH bond are shown in the appendix (Figure D.3). For all types of reactions endothermic reaction steps with reaction energies ranging from $+72$ kJ·mol^{-1} for the ethanethiol formation (Figure 5.13a) to $+130$ kJ·mol^{-1} for the ethanol formation (Figure 5.13c) are observed. With $+72$ kJ·mol^{-1} the formation of ethanethiol and the blocking of the surface oxygen by the alkyl chain is the most likely reaction. The reaction barriers for these reactions were not explicitly calculated. However, for an endothermic reaction step the reaction barrier has to be even larger than the change in energy. Consequently, the decomposition of DES is in competition with the desorption of DES since the decomposition barrier will be larger than the adsorption energy.

DES Reactivity on Cu(111)

In Figure 5.14 the corresponding adsorption and decomposition structure of DES on Cu(111) are shown. With -135 kJ·mol^{-1} (Figure 5.14a) a large adsorption energy is observed for DES on Cu(111). This structure is -74 kJ·mol^{-1} more stable than the most stable adsorption structure of DES on SiO_2. Furthermore, with $+10$ kJ·mol^{-1} the decomposition reaction of DES on Cu(111) is only slightly endothermic (Figure 5.14b).

To understand which fragment is responsible for the blocking of Cu(111) also the adsorption energies of the individual fragments were calculated. In these calculations, the individual fragments in the gas phase are used as the reference. With -285 kJ·mol^{-1} for the ethylsulfanyl

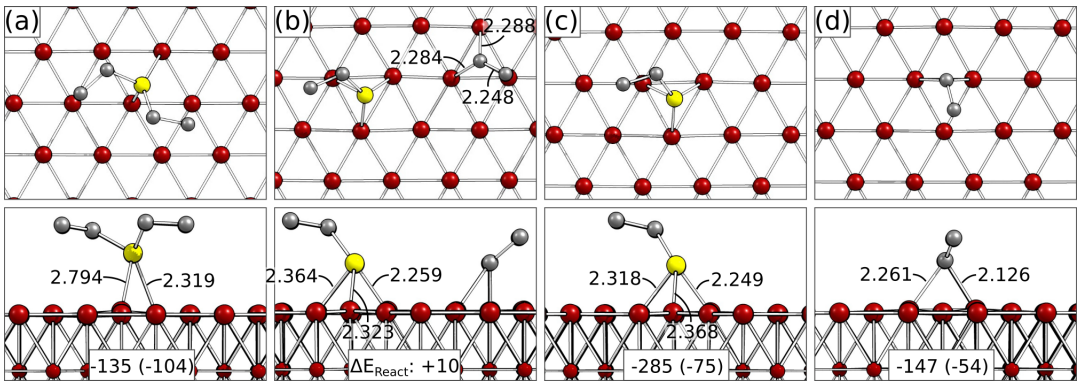


Figure 5.14.: Adsorption and decomposition of DES on Cu(111). (a) Initial adsorption structure of DES on Cu(111). (b) The decomposition structure of DES with the isolated fragments shown in (c) and (d). The reaction energy is stated relative to (a) while the adsorption energy of the fragments is stated relative to the energy of the fragments in the gas phase. Energies in kJ·mol^{-1} and bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Cu (copper red).

(Figure 5.14c) and $-147\,\mathrm{kJ{\cdot}mol^{-1}}$ for the ethyl radical (Figure 5.14d) large adsorption energies are observed for both fragments. Consequently, it is unlikely that any of these radicals will desorb.

However, since a minor diffusion barrier of up to $4\,\mathrm{kJ{\cdot}mol^{-1}}$ was obtained for both radicals, also the possibility for radical recombination prior to a desorption was studied. Following Figure 5.11b the formation of butane and ethyldisulfide were calculated. For the formation and desorption of butane a reaction energy of $-81\,\mathrm{kJ{\cdot}mol^{-1}}$ is observed while the formation and desorption of ethyldisulfide has a reaction energy of $+285\,\mathrm{kJ{\cdot}mol^{-1}}$. Consequently, we conclude that after the decomposition of DES on Cu(111) the ethyl radicals recombine and desorb as butane while the thiyl radicals are responsible for blocking the Cu(111) surface. This is in line with the previously studied reactivity of ethanethiol[362].

Conclusion

Adsorption and reactivity of DES on silicon oxide and copper was studied for the AS-ALD of HfO_2. In agreement to previous studies of ethanethiol and DES, we showed that the SiO_2 oxide is passivated by the formation of ethyl fragments while the copper surface is blocked by ethanethiol fragments. Furthermore, the more favorable adsorption and reaction energies for DES on Cu(111) over SiO_2 match the experimental results of better blocking of Cu(111)[356]. Also, the endothermic nature of the observed reactions explains the demand for high process temperatures of $400\,^{\circ}\mathrm{C}$ during the exposure of DES to the substrates.

6

Summary

In the present dissertation the surface reactivity of small molecules within the thin film growth by CVD and AS-ALD was studied by DFT. In a first part, the need to develop approaches for an automated exploration of reaction networks was formulated since both growth processes are controlled by a vast number of individual reactions. An understanding of these reactions at the atomic level would open the opportunity to fine-tune the growth process of interest and thereby improve the performance of the constructed device.

Consequently, my software PESE was introduced for an automated exploration of reaction networks. Here, two aspects of PESE were presented: The first part was focusing on the implementation concept while the second part was addressing the implemented functionalities. For the implementation, a highly parallel scaling was identified as the determining factor to efficiently use computational resources and explore reaction networks in a reasonable amount of time. Therefore, a highly automated software was developed to minimize human interventions. In addition, a high parallel scaling was ensured by employing a MPI parallelization of PESE with the Evaluator-Worker parallelization model. Also, all time consuming evaluation routines were parallelized to minimize execution time and to maximize the computational resources dedicated to time consuming DFT calculations.

The exploration algorithm itself obtains adsorption minima by a grid-based sampling of the surface of interest. Thereby, the whole sampling is performed for a minimal surface area only, which is defined by the detected symmetry elements of the surface. In addition, by extending the grid layer in the z-direction the adsorption basin of every minima is derived. Here, the volume of the adsorption basins is interpreted as the geometric probability of the adsorbate to adsorb in a certain minimum. To estimate the size of a basin an evaluation routine for the smearing and mapping of all stored structures to an equally spaced grid is used.

Based on the obtained adsorption minima, possible decomposition structures are proposed by PESE. However, only unique minima are considered in this step. The automated identification of duplicate structures is a critical step to reduce human intervention. Here, unique

and duplicate minima are identified based on an RMSD and energy threshold. For the calculation of the RMSD value the Hungarian algorithm is utilized. This algorithm minimizes the obtained RMSD value with respect to the pairing of atoms. This ensures that the ordering of the atoms in the corresponding structures has no longer an effect on the detection of duplicates. Once the unique minima are identified, decomposition structures are obtained by following implemented rules exploiting the connectivity graphs of the studied molecule. For PESE, the decomposition channels of radical cleavage reactions and elimination reactions are considered. Thereby, the usage of PESE is currently restricted to the CVD process. For an extension to the ALD process, the implementation of rules addressing reactions between the adsorbate and the surface will be necessary in the future. Based on every unique minima several decomposition structures are proposed while for the arrangement of the fragments again a grid-based approach is used.

Based on the obtained adsorption and decomposition minima, diffusion and decomposition paths are proposed by PESE. While for every unique decomposition structure a single decomposition paths is proposed, the selection of diffusion paths is more complex. Here, an algorithm introducing a relative length to prefer two short diffusion steps over one long diffusion step is applied. Thereby, PESE can generate a minimal diffusion network connecting all obtained adsorption minima. An initial guess for every reaction path is generated by combining a linear interpolation in cartesian coordinates for the surface model with a linear interpolation in internal coordinates for the adsorbate. In addition to the internal coordinates, also the rotation of the adsorbate and its change in the center of mass is considered. For the optimization of the reaction path a tailored and improved version of the NEB method is used. For performance reasons, a serial optimization of the images is pursued. This enables PESE to change the number of images along a reaction path during the calculation based on a requested energy or spatial resolution. In addition, minima and TSs are detected based on their energy difference to adjacent images and automatically optimized. Furthermore, the approach of Lindgren and coworkers[201] to change the convergence criterium of the images is included to further improve the performance of the method.

The first part of this thesis was completed by studying several systems of interest with PESE. Here, the adsorption and diffusion of a single Bi atom on GaP(001) was investigated as the smallest possible system. The adsorption into a Ga-Ga dimer bond was identified as the most stable adsorption site with an energy of -259 kJ·mol^{-1} while also the largest adsorption basin was found for this structure with a relative size of 31.1 %. In total, 11 chemisorbed minima within a dataset of $17\,206$ structures were obtained. Based on the chemisorbed minima, 20 diffusion paths were generated by PESE and 19 paths were successfully optimized resulting

in the identification of 17 transition state structures. Here, diffusion barriers in the range of $+7$ kJ·mol^{-1} to $+127$ kJ·mol^{-1} were observed underlining the high mobility of Bi on the GaP surface. Also, by the evaluation of the reaction network a preference for Bi to diffuse to one of the most stable structures with the formation of Bi – Ga bonds was identified.

For GaH_3 different adsorption structures with energies in the range of -96 to -148 kJ·mol^{-1} were identified by using PESE. Here, the H-Ga_{Surf} bond formation was identified as a common structural motif. In addition, the most stable structure for GaH_3 was found for an adsorption on top of the mixed-dimer. However, small adsorption basins were observed for all adsorption minima with a relative size of up to 15.5 %. This was discussed as an indication for the absence of a preference for a single adsorption site. Instead, the adsorption of GaH_3 is expected to take place at the mixed-dimer, in the trench between two topmost Ga-layers or on top of a Ga-Ga or Ga-P bond. By presenting several exemplary reactions paths optimized by PESE, the excellent performance of PESE to deal with complex reaction paths including multiple minima and TS was shown. For the reaction network of GaH_3 on GaP(001), 113 reaction paths with 87 minima and 142 TSs along the reaction paths were obtained in total. Within this dataset, the hydrogen transfer reactions for a single hydrogen or all three hydrogens at once were identified as the most likely reactions. By contrast, larger barriers and less favored reaction energies were observed for the hydrogen elimination reaction. Overall, a stabilization by the GaP surface was identified resulting in smaller reaction barriers and more favorable reaction energies in comparison to previous gas phase calculations.

The adsorption and decomposition of PH_3 on GaP(001) was studied by PESE as the third system. Here, the largest dataset covering $708\,314$ structures was obtained in the calculation of the adsorption minima and basins. With -66 kJ·mol^{-1}, the most stable structure was observed for the adsorption of PH_3 at the Ga-atoms in the trench between the topmost Ga-layers. Interestingly, the largest adsorption basin was observed for an adsorption at the adjacent Ga-atom in the trench with a relative size of 28.6 %. Similar to GaH_3, is was shown that the hydrogen transfer reaction is the most likely decomposition reaction while the hydrogen elimination reaction is accessible at elevated temperatures.

The performed calculation to obtain the adsorption basins for Bi, GaH_3 and PH_3 on GaP(001) were used to evaluate the performance of PESE. Here, it was shown that with up to 14.0 % of the computing time a minor communication overhead is caused by the usage of PESE. However, with the implemented parallelization approach of PESE it is now possible to efficiently use $16\,384$ cores for DFT calculations focusing on the reactivity of a single adsorbate. Still, further improvements are needed to be able to study the decomposition of larger precursors and thereby more complex reaction networks.

In a second part of this thesis, the usage of different classes of molecules as SMI for the AS-ALD of high-κ dielectrics such as Al_2O_3 and HfO_2 was investigated. Here, the adsorption and reactivity of five alkoxysilane based SMIs were studied for the AS-ALD of Al_2O_3 on copper over silicon dioxide. By varying the number of reactive (methoxy) and blocking (methyl) groups it was shown that not the reaction mechanism is determining the different blocking behavior but the number of potential reactions with the SiO_2 surface. Furthermore, the siloxane formation was identified as an additional blocking mechanism. However, with up to $+144\,\mathrm{kJ{\cdot}mol^{-1}}$ large reaction barriers were observed for these condensation reactions. In consideration of the results of experimental coworkers, the superior blocking of trifunctional SMIs was attributed to both the possibility to passivate several reactive groups of the SiO_2 surface and the possibility to form an on-surface siloxane network. Based on gas phase calculations, the coreactant was identified as the crucial molecule to weaken the blocking layer, while the experimental proof of TEA being the better Al-precursor could not be explained due to limitations of the theoretical model.

With MSA a second class of SMI was investigated for the AS-ALD of Al_2O_3 on silicon oxide over copper. Here, it was shown by DFT calculations that the different reaction mechanisms for MSA on the SiO_2 and Cu(111) are responsible for the experimentally observed blocking behavior: Condensation reactions with high barriers of up to $+234\,\mathrm{kJ{\cdot}mol^{-1}}$ and unfavorable thermodynamics were observed for SiO_2, while deprotonation reactions with a minor reaction barrier of $+29\,\mathrm{kJ{\cdot}mol^{-1}}$ and favorable thermodynamics were observed for Cu(111). Consequently, a dense blocking layer of MSA is only expected on copper and not on silicon oxide. Furthermore, a reaction mechanism for the oxidation of MSA on Cu(111) was hypothesized to explain the experimentally observed reduction and blocking of the native copper oxide.

The blocking of HfO_2 on silicon oxide and copper was investigated as the third system. Here, a sulfur-based SMI was studied with DES. It was shown that the SiO_2 oxide is passivated by the formation of a ethyl fragments while the copper surface is blocked by the formation of ethanethiol fragments. Furthermore, the more favorable adsorption and reaction energies for DES on Cu(111) over SiO_2 were used to explain the experimental results of better blocking of Cu(111).

Overall, in this thesis a software for the automated exploration of reaction networks of CVD processes was presented. This software was successfully applied to small systems of Bi, GaH_3 and PH_3 on GaP(001). In total, over one million structures were obtained and evaluated to derive adsorption basins for 41 chemisorbed adsorption minima. In addition, 398 reaction paths were automatically proposed, whereby every second path was successfully optimized

by the implemented reaction path optimization algorithm. With this software development the next step towards the computational description of thin film growth by CVD and ALD is made. Also, for the AS-ALD of Al_2O_3 and HfO_2 three new SMI classes were introduced. Here, computational methods were successfully applied to identify the surface reaction mechanisms and understand the experimentally observed blocking behaviors. The hereby gained knowledge will promote the development of novel SMIs and the fine-tuning of the AS-ALD process.

7

Zusammenfassung

In der vorliegende Dissertation wurde die Oberflächenreaktivitat kleiner Moleküle innerhalb des Dünnschichtwachstums durch chemische Gasphasenabscheidung (CVD) und Atomlagenabscheidung (ALD) untersucht. Der verwendete Modellierungsansatz beruht dabei auf der Dichtefunktionaltheorie. In einem ersten Teil der Arbeit wurde die Notwendigkeit formuliert, einen automatisierten Ansatz für die Berechnung von Reaktionsnetzwerken zu entwickeln. Dieser ist von Nöten, da die CVD und ALD Wachstumsprozesse durch eine enorme Vielfalt an Reaktionen bestimmt werden. Ein Verständnis dieser atomaren Prozesse ist wünschenswert, da dadurch eine Optimierung sämtlicher Prozessparameter, wie zum Beispiel die Wahl der Präkursoren oder der optimalen Substrattemperatur, ermöglicht wird, wodurch letztlich die Qualität der konstruierten Bauelemente gesteigert werden kann.

Folglich wurde mit PESE meine Softwareentwicklung für eine automatisierte Berechnung von Reaktionsnetzwerken vorgestellt. Hier wurden zwei Aspekte von PESE hervorgehoben: Der erste Teilaspekt konzentrierte sich auf das Implementierungskonzept während der zweite Teilaspekt die implementierten Funktionalitäten vorstellte. Für die Implementierung wurde eine hohe parallele Skalierung als entscheidender Faktor identifiziert um eine effiziente Nutzung von Rechenressourcen sicherzustellen und das gewünschte Reaktionsnetzwerke innerhalb eines akzeptablen Zeitaufwands zu bestimmen. Als Konsequenz wurde mit PESE eine Software entwickelt, welche ein Reaktionsnetzwerk nahezu ohne Eingriffe des Nutzers berechnet. Darüber hinaus wurde eine hohe parallele Skalierbarkeit durch die Verwendung des MPI Standards und des *Evaluator-Worker* Parallelisierungsmodells sichergestellt. Zudem wurden alle zeitaufwändigen Schritte parallelisiert, um die deren Ausführungszeit zu minimieren und gleichzeitig den Großteil der verwendeten Rechenressourcen den zeitaufwändigen DFT-Berechnungen zuzuweisen.

Die Berechnung des Reaktionsnetzwerkes selbst beginnt mit der Bestimmung von Adsorptionsminima des betrachteten Moleküls. Die dafür nötigen Strukturen werden anhand eines Gitternetz basierten Ansatzes erhalten und mittels DFT optimiert. Das Gitternetz umspannt

jedoch lediglich einen Teil der Oberfläche, welcher anhand der vorliegenden Symmetrieelemente erkannt wird. Zudem kann anhand einer Ausdehnung des Gitternetzes in z-Richtung der Adsorptionstrichter eines jeden Minimums bestimmt werden. Dabei wird das Volumen des Adsorptionstrichters als geometrisches Maß für die Wahrscheinlichkeit einer Adsorption des Adsorbats in ein bestimmtes Minimum interpretiert. Für die Bestimmung des Volumens wurde eigens eine Programmroutine entwickelt, welche alle gespeicherten Strukturen mit Hilfe von Gauß-Funktionen auf ein gleichmäßiges Gitternetz abbildet.

Auf Grundlage der erhaltenen Adsorptionsminima werden mögliche Zerfallsstrukturen von PESE vorgeschlagen. Jedoch werden in diesem Schritt lediglich einzigartige Adsorptionsminima berücksichtigt. Dabei ist die automatische Identifizierung von Duplikaten ein wichtiger Schritt um eine Intervention des Nutzers zu vermeiden. Daher werden einzigartige und doppelte Minima anhand von Grenzwerten für strukturelle und energetische Ähnlichkeit identifiziert. Strukturelle Ähnlichkeit wird dabei anhand der Standardabweichung (engl. *root mean square deviation* (RMSD)) der Atomkoordinaten bestimmt. Für die Berechnung des RMSD Wertes wurde zudem die Ungarische Methode implementiert. Anhand dieser Methode wird ein minimierter RMSD Wert erhalten unter Berücksichtigung der Paarung der Atome. Dadurch wird sichergestellt, dass die Anordnung der Atome in den entsprechenden Strukturen keinen Einfluss auf die Erkennung von Duplikaten hat. Sobald die einzigartigen Minima identifiziert sind, werden die Zerfallsstrukturen anhand von implementierten Regeln erstellt, wobei die Regeln auf Konnektivitätsgraphen des untersuchten Moleküls basieren. Für PESE wurden zwei unterschiedliche Typen von Zerfallsreaktionen implementiert: Die radikalische Spaltung von Bindungen und Eliminierungsreaktionen. Da lediglich Zerfallsreaktionen implementiert sind, ist die Verwendung von PESE momentan auf die Beschreibung des CVD-Prozesses beschränkt. Für eine Erweiterung auf den ALD-Prozess wird zukünftig die Implementierung von Regeln notwendig sein, welche Reaktionen zwischen dem Adsorbat und der Oberfläche beschreiben. Anhand der bisher vorhandenen Regeln werden für alle einzigartigen Minima mehrere Zerfallsstrukturen erstellt. Dabei wird eine sinvolle Anordnung der Fragmente erneut anhand eines Gitternetz basierten Ansatzes sichergestellt.

In einer zweiten Phase verwendet PESE alle erhaltenen Adsorptions- und Zerfallsstrukturen um entsprechende Diffusions- und Zerfallspfade vorzuschlagen. Hierbei wird für jede einzigartige Zerfallsstruktur ein einziger Zerfallspfad vorgeschlagen. Die Auswahl von möglichen Diffusionspfaden ist jedoch deutlich komplexer. Daher wurde ein Algorithmus implementiert, der anhand der relativen Länge von Diffusionspfaden entscheidet, welcher Pfad optimiert wird. Dabei führt die Verwendung der relativen Länge dazu, dass zwei kurze Diffusionspfade gegenüber einem langen Diffusionspfad bevorzugt werden. Auf diese Weise kann PESE ein minima-

les Diffusionsnetzwerk erzeugen, in welchem alle einzigartigen Adsorptionsminima enthalten sind. Die Interpolation eines jeden Reaktionspfades erfolgt durch die Kombination einer linearen Interpolation in kartesischen Koordinaten für die Oberfläche mit einer linearen Interpolation in internen Koordinaten für das Adsorbat. Zusätzlich zu den internen Koordinaten wird auch die Rotation des Adsorbats und die Veränderung des Massenschwerpunkts berücksichtigt. Für die Optimierung jedes Reaktionspfades wurde eine angepasste und verbesserte Version der NEB-Methode implementiert. Dabei wird zu Gunsten der Performance eine serielle Optimierung der Bilder verfolgt. Dies ermöglicht es PESE die Anzahl der Bilder entlang eines Reaktionspfades während der Berechnung auf der Grundlage einer gewünschten Energieauflösung oder strukturellen Auflösung zu ändern. Darüber hinaus werden Minima und Übergangszustände (engl. *transition state* (TS)) anhand ihrers Energieunterschiedes zu benachbarten Bildern erkannt und automatisch optimiert. Zudem wird der Ansatz von Lindgren und Mitarbeitern[201] zur Änderung des Konvergenzkriterium berücksichtigt, um die Performance der implementierten NEB-Methode weiter zu steigern.

Der erste Teil dieser Arbeit wurde vervollständigt indem PESE verwendet wurde verschiedene Systeme zu untersuchen. Dabei wurde anhand der Adsorption und Diffusion eines einzelnen Bi-Atoms auf GaP(001) das kleinste mögliche Testsystem untersucht. Als stabilste Adsorptionsstrukture mit einer Energie von $-259\,\mathrm{kJ{\cdot}mol^{-1}}$ wurde die Insertion in die Bindung des Ga-Ga Dimers identifiziert. Zudem wurde auch der größte Adsorptionstrichters mit einer relativen Größe von $31\,\%$ für diese Struktur gefunden. Insgesamt wurden 11 chemisorbierte Minima innerhalb eines Datensatzes von $17\,206$ Strukturen erhalten. Basierend auf den chemisorbierten Minima wurden 20 Diffusionspfade durch PESE vorgeschlagen und 19 Reaktionspfade erfolgreich optimiert. Dabei wurden 17 Übergangszustände mit entsprechenden Diffusionsbarrieren im Bereich von $+7\,\mathrm{kJ{\cdot}mol^{-1}}$ bis $+127\,\mathrm{kJ{\cdot}mol^{-1}}$ beobachtet, wodurch die hohe Beweglichkeit von Bi auf der GaP-Oberfläche unterstrichen wird. Zudem wurde durch die Auswertung des Reaktionsnetzwerks gezeigt, dass für Bi eine Präferenz besteht zu den stabilsten Strukturen mit Bi – Ga Bindungen zu diffundieren.

Als weiteres System wurden die Adsorption und Zerfallsreaktionen von GaH_3 auf GaP(001) mittels PESE untersucht. Es wurde gezeigt, dass verschiedene Adsorptionsstrukturen mit Energien im Bereich von -96 bis $-148\,\mathrm{kJ{\cdot}mol^{-1}}$ vorliegen. Dabei wurde die Bildung von H-$Ga_{\text{Oberfl.}}$ Bindungen als häufiges Strukturmotiv identifiziert. Zudem wurde gezeigt, dass die stabilste GaH_3 Adsorptionsstruktur für eine Adsorption auf dem Ga-P Dimer vorliegt. Mit einer relativen Größe von bis zu $15.5\,\%$ wurden für die einzelnen Adsorptionsminima jedoch kleinere Adsorptionstrichters beobachtet als für Bi auf GaP. Diese Beobachtung wurde als Hinweis für das Fehlen einer präferierten Adsorptionsstruktur interpretiert. Für die Adsorp-

tion von GaH_3 auf GaP wird stattdessen erwartet, dass sich mehrere GaH_3 Adsorbate auf unterschiedliche Adsorptionsplätze verteilen, darunter den Ga-P Dimer, Ga-Atomen zwischen den obersten Ga-Schichten oder den Ga-Ga Dimer. Anhand beispielhafter Reaktionspfade bestehend aus mehreren Minima und TSs wurde die Verlässlichkeit und Leistungsfähikeit von PESE gezeigt komplexe Reaktionspfade zu optimieren. Für das Reaktionsnetzwerk von GaH_3 auf GaP(001) wurden insgesamt 113 Reaktionspfade mit 87 Minima und 142 TSs entlang der Reaktionspfade erhalten. Innerhalb dieses Datensatzes wurde die Wasserstoffabspaltung als die wahrscheinlichste Reaktion identifiziert. Im Gegensatz dazu wurden größere Barrieren und weniger günstige Reaktionsenergien für die Eliminierung von H_2 beobachtet. Alle Reaktionen haben jedoch gemeinsam, dass eine Stabilisierung durch die GaP-Oberfläche identifiziert wurde, da kleinere Reaktionsbarrieren und günstigere Reaktionsenergien im Vergleich zu Zerfällen in der Gasphase beobachtet wurden.

Die Adsorption und Zerfallsreaktionen von PH_3 auf GaP(001) wurde mittels PESE als drittes System untersucht. Mit $708\,314$ gespeicherten Strukturen wurde der mit Abstand größte Datensatz für die Berechnung der Adsorptionsminima und Adsorptionstrichter erhalten. Mit $-66\,\mathrm{kJ{\cdot}mol^{-1}}$ wurde die stabilste Adsorptionsstruktur für die Adsorption von PH_3 an einem Ga-Atomen im Graben zwischen den obersten Ga-Schichten beobachtet. Der größte Adsorptionstrichter wurde dabei für eine Adsorption an einem benachbarten Ga-Atom beobachtet mit einer relativen Größe von $28.6\,\%$. Analog zu GaH_3 wurde auch für PH_3 gezeigt, dass die Wasserstoffabspaltung die wahrscheinlichste Zerfallsreaktion ist, während die Eliminierung von H_2 erst bei erhöhten Temperaturen zugänglich ist.

Die Berechnungen der Adsorptionstrichter für Bi, GaH_3 und PH_3 auf GaP(001) wurden zudem verwendet, um die Leistung von PESE zu bestimmen. Hier wurde gezeigt, dass mit bis zu $14.0\,\%$ der gesamten Rechenzeit ledigleich ein geringer zusätzlicher Kommunikationsaufwand durch die Verwendung von PESE anfällt. Gleichzeitig ist es mit dem implementierten Ansatz von PESE nun jedoch möglich $16\,384$ Rechenkerne effizient für DFT-Berechnungen zu verwenden. Dennoch werden weitere Verbesserungen angestrebt, um auch Zerfallsreaktionen größerer Moleküle und damit komplexerer Reaktionsnetzwerke untersuchen zu können.

In einem zweiten Teil der Arbeit wurde die Verwendung verschiedener Molekülklassen als Inhibitoren (engl. *small molecule inhibitor* (SMI)) für die Oberflächen selektive Atomlagenabscheidung (engl. *area selective atomic layer deposition* (AS-ALD)) von High-κ-Dielektrika wie Al_2O_3 und HfO_2 untersucht. Als erste Molekülklasse wurden fünf Alkyloxysilan basierte SMIs und deren Verwendung in der AS-ALD von Al_2O_3 auf Kupfer bei gleichzeitiger Blockierung von Siliziumdioxid vorgestellt. Durch Variation der Anzahl der reaktiven (Methoxy-) und

blockierenden (Methyl-) Gruppen konnte gezeigt werden, dass nicht der Reaktionsmechanismus mit der Siliziumdioxid Oberfläche für eine unterschiedlich gute Blockierung entscheidend ist, sondern die Anzahl der potentiellen Reaktionen mit der SiO_2-Oberfläche. Darüber hinaus wurde die Bildung von Siloxanen als zusätzlicher Effekt für die Blockierung der SiO_2 Oberfläche identifiziert. Jedoch wurden mit bis zu $+144\,\mathrm{kJ{\cdot}mol^{-1}}$ große Reaktionsbarrieren für diese Kondensationsreaktionen beobachtet. Unter Einbeziehung der experimentellen Daten der Kooperationspatner wurde gezeigt, dass die untersuchten trifunktionalen SMIs einem mono- oder bifunktionalem SMI vorzuziehen sind, da sie zum einen in der Lage sind, mehrere reaktive Gruppen der SiO_2 Oberfläche zu passivieren und zum anderen ein Siloxan-Netzwerk an der Oberfläche bilden. Durch eine Untersuchtung der SMI Stabilität basierend auf Berechnungen in der Gasphase wurde der O-Präkursor als das kritische Molekül identifiziert, welches die SMI Schicht angreift. Jedoch sind weitere Rechnungen mit dem SiO_2 Oberflächenmodell von Nöten, da die experimentelle Beobachtung, dass Triethylaluminium (engl. *triethylaluminium* (TEA)) der bessere Al-Präkursor ist, innerhalb des limitierten Gasphasenmodels nicht erklärt werden konnte.

Als zweite Molekülklasse wurde Methansulfonsäure (engl. *methanesulfonic acid* (MSA)) als SMI für die AS-ALD von Al_2O_3 auf Siliziumdioxid bei einer Blockierung von Kupfer untersucht. Anhand von DFT-Berechnungen wurde gezeigt, dass die unterschiedlichen Reaktionsmechanismen für MSA auf der SiO_2 und Cu(111) Oberfläche für das experimentell beobachtete Verhalten verantwortlich sind: Für SiO_2 wurden Kondensationsreaktionen mit hohen Barrieren von bis zu $+234\,\mathrm{kJ{\cdot}mol^{-1}}$ und endothermen Reaktionsenergien beobachtet. Im Gegensatz dazu wurden eine Deprotonierung von MSA mit einer geringen Reaktionsbarriere von $+29\,\mathrm{kJ{\cdot}mol^{-1}}$ und exothermer Reaktionsenergie für Cu(111) beobachtet. Folglich wird eine dichtere Packung von MSA Molekülen nur auf der Kupfer und nicht auf der Siliziumdioxid Oberfläche erwartet. Zusätzlich zu der Deprotonierung von MSA auf Cu(111) wurde ein Reaktionsmechanismus für die Oxidation von MSA und eine Reduktion der nativen Kupferoxidschicht entwickelt, um die experimentell beobachtete Reduktion und Blockierung des Kupferoxids zu erklären.

Als drittes System wurde die Blockierung des Wachstums von HfO_2 auf Siliziumdioxid und Kupfer untersucht. Dabei wurde mit Diethylsulfid (engl. *diethyl ether* (DES)) ein schwefelbasierter Inhibitor untersucht. Anhand von Reaktionsenergien wurde gezeigt, dass die SiO_2 Oberfläche durch die Bildung eines Ethylfragments während die Kupferoberfläche durch die Bildung von Ethanthiolfragmenten blockiert wird. Aufgrund von stärker gebundenen Adsorptionsstrukturen und weniger endothermen Reaktionsenergien für DES auf Cu(111) im Vergleich zu SiO_2 konnten die experimentellen Beobachtungen einer besseren Blockierung von Cu(111) erklärt

werden.

Insgesamt wurde in dieser Arbeit eine Software zur automatisierten Berechnung von Reaktionsnetzwerken innerhalb der CVD-Prozesse vorgestellt. Die entwickelte Software wurde erfolgreich auf kleine Systeme von Bi, GaH_3 und PH_3 auf GaP(001) angewendet. Insgesamt wurden über eine Million Strukturen berechnet und ausgewertet, um Adsorptionstrichter für 41 chemisorbierte Adsorptionsminima zu bestimmen. Des Weiteren wurden durch PESE 398 Reaktionspfade vorgeschlagen, wobei jeder zweite Pfad durch den implementierten Optimierungsalgorithmus erfolgreich erhalten wurde. Somit wird mit PESE der nächste Schritt für die theoretische Modellierung aller Prozesse des Dünnschichtwachstums innerhalb des CVD und ALD Prozesses erreicht. Als zweiter Aspekt dieser Arbeit wurden für die AS-ALD von Al_2O_3 und HfO_2 drei neue SMI-Klassen eingeführt. Hier wurden erfolgreich quantenchemische Methoden angewendet um die Reaktionsmechanismen zu identifizieren und die experimentellen Beobachtungen zu verstehen. Das dabei gewonnene Wissen wird die Entwicklung neuartiger SMIs und die Optimierung der AS-ALD Prozesse fördern.

Literature

[1] International Roadmap for Devices and Systems (IRDS™) **2021**, can be found under https://irds.ieee.org, Last accessed: 18.07.2022.

[2] A. C. Jones, M. L. Hitchman in *Chemical Vapour Deposition: Precursors, Processes and Applications*, (Eds.: A. C. Jones, M. L. Hitchman), The Royal Society of Chemistry, London, United Kingdom, **2009**, pp. 1–36, DOI 10.1039/9781847558794-00001.

[3] J. Bardeen, W. H. Brattain, *Phys. Rev.* **1948**, *74*, 230–231, DOI 10.1103/PhysRev.74.230.

[4] H. Ko, K. Takei, R. Kapadia, S. Chuang, H. Fang, P. W. Leu, K. Ganapathi, E. Plis, H. S. Kim, S.-Y. Chen, M. Madsen, A. C. Ford, Y.-L. Chueh, S. Krishna, S. Salahuddin, A. Javey, *Nature* **2010**, *468*, 286–289, DOI 10.1038/nature09541.

[5] H. Riel, L.-E. Wernersson, M. Hong, J. A. del Alamo, *MRS Bull.* **2014**, *39*, 668–677, DOI 10.1557/mrs.2014.137.

[6] K. M. C. Babu, E. Goel, *Silicon* **2022**, DOI 10.1007/s12633-022-01826-0.

[7] A. Pandey, *Silicon* **2022**, DOI 10.1007/s12633-022-01694-8.

[8] F. Dimroth, *Phys. Status Solidi C* **2006**, *3*, 373–379, DOI 10.1002/pssc.200564172.

[9] J. Li, A. Aierken, Y. Liu, Y. Zhuang, X. Yang, J. H. Mo, R. K. Fan, Q. Y. Chen, S. Y. Zhang, Y. M. Huang, Q. Zhang, *Front. Phys.* **2021**, *8*, DOI 10.3389/fphy.2020.631925.

[10] R. M. France, J. F. Geisz, T. Song, W. Olavarria, M. Young, A. Kibbler, M. A. Steiner, *Joule* **2022**, *6*, 1121–1135, DOI 10.1016/j.joule.2022.04.024.

[11] B. Kunert, S. Reinhard, J. Koch, M. Lampalzer, K. Volz, W. Stolz, *Phys. Status Solidi C* **2006**, *3*, 614–618, DOI 10.1002/pssc.200564124.

[12] G. Roelkens, L. Liu, D. Liang, R. Jones, A. Fang, B. Koch, J. Bowers, *Laser Photonics Rev.* **2010**, *4*, 751–779, DOI 10.1002/lpor.200900033.

[13] S. Liebich, M. Zimprich, A. Beyer, C. Lange, D. J. Franzbach, S. Chatterjee, N. Hossain, S. J. Sweeney, K. Volz, B. Kunert, W. Stolz, *Appl. Phys. Lett.* **2011**, *99*, 071109, DOI 10.1063/1.3624927.

[14] Y. Hu, D. Liang, K. Mukherjee, Y. Li, C. Zhang, G. Kurczveil, X. Huang, R. G. Beausoleil, *Light Sci. Appl.* **2019**, *8*, 93, DOI 10.1038/s41377-019-0202-6.

[15] M. Tang, J.-S. Park, Z. Wang, S. Chen, P. Jurczak, A. Seeds, H. Liu, *Prog. Quantum Electron.* **2019**, *66*, 1–18, DOI 10.1016/j.pquantelec.2019.05.002.

[16] Y. Han, K. M. Lau, *J. Appl. Phys.* **2020**, *128*, 200901, DOI 10.1063/5.0029804.

[17] H. Amano, N. Sawaki, I. Akasaki, Y. Toyoda, *Appl. Phys. Lett.* **1986**, *48*, 353–355, DOI 10.1063/1.96549.

[18] S. Nakamura, T. Mukai, M. Senoh, *Appl. Phys. Lett.* **1994**, *64*, 1687–1689, DOI 10.1063/1.111832.

[19] I. Akasaki, H. Amano, *Jpn. J. Appl. Phys.* **1997**, *36*, 5393–5408, DOI 10.1143/jjap.36.5393.

[20] A. Hangleiter, *MRS Bull.* **2003**, *28*, 350–353, DOI 10.1557/mrs2003.99.

[21] J. Cho, J. H. Park, J. K. Kim, E. F. Schubert, *Laser Photonics Rev.* **2017**, *11*, 1600147, DOI 10.1002/lpor.201600147.

[22] M. S. Wong, S. Nakamura, S. P. DenBaars, *ECS J. Solid State Sci. Technol.* **2019**, *9*, 015012, DOI 10.1149/2.0302001jss.

[23] J. Sun, M. Han, Y. Gu, Z. Yang, H. Zeng, *Adv. Opt. Mater.* **2018**, *6*, 1800256, DOI 10.1002/adom.201800256.

[24] A. Ren, L. Yuan, H. Xu, J. Wu, Z. Wang, *J. Mater. Chem. C* **2019**, *7*, 14441–14453, DOI 10.1039/C9TC05738B.

[25] S. Mauthe, Y. Baumgartner, M. Sousa, Q. Ding, M. D. Rossell, A. Schenk, L. Czornomaz, K. E. Moselund, *Nat. Commun.* **2020**, *11*, 4565, DOI 10.1038/s41467-020-18374-z.

[26] C. Liu, J. Guo, L. Yu, J. Li, M. Zhang, H. Li, Y. Shi, D. Dai, *Light Sci. Appl.* **2021**, *10*, 123, DOI 10.1038/s41377-021-00551-4.

[27] R. R. LaPierre, M. Robson, K. M. Azizur-Rahman, P. Kuyanov, *J. Phys. D: Appl. Phys.* **2017**, *50*, 123001, DOI 10.1088/1361-6463/aa5ab3.

[28] X. Chen, C. Liu, S. Mao, *Nano-Micro Lett.* **2020**, *12*, 95, DOI 10.1007/s40820-020-00438-w.

[29] B. Saruhan, R. Lontio Fomekong, S. Nahirniak, *Front. Sens.* **2021**, *2*, DOI 10.3389/fsens.2021.657931.

[30] P. Raju, Q. Li, *J. Electrochem. Soc.* **2022**, *169*, 057518, DOI 10.1149/1945-7111/ac6e0a.

[31] https://irds.ieee.org/topics/semiconductor-materials, Last accessed: 18.07.2022.

[32] B. Kunert, K. Volz, W. Stolz, *ECS Trans.* **2010**, *28*, 219–226, DOI 10.1149/1.3367953.

[33] J. A. del Alamo, *Nature* **2011**, *479*, 317–323, DOI 10.1038/nature10677.

[34] R. W. Johnson, A. Hultqvist, S. F. Bent, *Mater. Today* **2014**, *17*, 236–246, DOI 10.1016/j.mattod.2014.04.026.

[35] Z. Cai, B. Liu, X. Zou, H.-M. Cheng, *Chem. Rev.* **2018**, *118*, 6091–6133, DOI 10.1021/acs.chemrev.7b00536.

[36] M. Ritala, J. Niinistö, *ECS Trans.* **2009**, *25*, 641–652, DOI 10.1149/1.3207651.

[37] R. L. Puurunen, *J. Appl. Phys.* **2005**, *97*, 121301, DOI 10.1063/1.1940727.

[38] S. M. George, *Chem. Rev.* **2010**, *110*, 111–131, DOI 10.1021/cr900056b.

[39] A. J. M. Mackus, A. A. Bol, W. M. M. Kessels, *Nanoscale* **2014**, *6*, 10941–10960, DOI 10.1039/C4NR01954G.

[40] J. Lu, J. W. Elam, P. C. Stair, *Surf. Sci. Rep.* **2016**, *71*, 410–472, DOI 10.1016/j.surfrep.2016.03.003.

[41] A. J. M. Mackus, J. R. Schneider, C. MacIsaac, J. G. Baker, S. F. Bent, *Chem. Mater.* **2019**, *31*, 1142–1183, DOI 10.1021/acs.chemmater.8b02878.

[42] H. M. Manasevit, *Appl. Phys. Lett.* **1968**, *12*, 156–159, DOI 10.1063/1.1651934.

[43] G. B. Stringfellow, *Organometallic vapor-phase epitaxy : theory and practice*, 2nd ed., Academic Press, San Diego, USA, **1999**, DOI 10.1016/B978-0-12-673842-1.X5000-5.

[44] G. Stringfellow, *J. Cryst. Growth* **2004**, *264*, 620–630, DOI 10.1016/j.jcrysgro.2003.12.037.

[45] H. Pedersen, *Chem. Mater.* **2016**, *28*, 691–699, DOI 10.1021/acs.chemmater.5b04553.

[46] L. Sun, G. Yuan, L. Gao, J. Yang, M. Chhowalla, M. H. Gharahcheshmeh, K. K. Gleason, Y. S. Choi, B. H. Hong, Z. Liu, *Nat Rev Methods Primers* **2021**, *1*, 5, DOI 10.1038/s43586-020-00005-y.

[47] D. Muñoz Rojas, J. MacManus-Driscoll, *Mater. Horiz.* **2014**, *1*, 314–320, DOI 10.1039/C3MH00136A.

[48] K. Volz, A. Beyer, W. Witte, J. Ohlmann, I. Németh, B. Kunert, W. Stolz, *J. Cryst. Growth* **2011**, *315*, 37–47, DOI 10.1016/j.jcrysgro.2010.10.036.

[49] R. Straubinger, M. Widemann, J. Belz, L. Nattermann, A. Beyer, K. Volz, *Sci. Rep.* **2018**, *8*, 9048, DOI 10.1038/s41598-018-27286-4.

[50] M. D. Allendorf, T. M. Besmann, R. J. Kee, M. T. Swihart in *Chemical Vapour Deposition: Precursors, Processes and Applications*, (Eds.: A. C. Jones, M. L. Hitchman), The Royal Society of Chemistry, London, United Kingdom, **2009**, pp. 93–157, DOI 10.1039/9781847558794-00093.

[51] S. D. Elliott, *Semicond. Sci. Technol.* **2012**, *27*, 074008, DOI 10.1088/0268-1242/27/7/074008.

[52] J. Neugebauer, T. Hickel, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 438–448, DOI 10.1002/wcms.1125.

[53] H. Pedersen, S. D. Elliott, *Theor. Chem. Acc.* **2014**, *133*, 1476, DOI 10.1007/s00214-014-1476-7.

[54] S. D. Elliott, G. Dey, Y. Maimaiti, H. Ablat, E. A. Filatova, G. N. Fomengia, *Adv. Mater.* **2016**, *28*, 5367–5380, DOI 10.1002/adma.201504043.

[55] N. Cheimarios, G. Kokkoris, A. G. Boudouvis, *Arch. Comput. Methods Eng.* **2021**, *28*, 637–672, DOI 10.1007/s11831-019-09398-w.

[56] L. C. Bannow, O. Rubel, S. C. Badescu, P. Rosenow, J. Hader, J. V. Moloney, R. Tonner, S. W. Koch, *Phys. Rev. B* **2016**, *93*, 205202, DOI 10.1103/PhysRevB.93.205202.

[57] P. Rosenow, L. C. Bannow, E. W. Fischer, W. Stolz, K. Volz, S. W. Koch, R. Tonner, *Phys. Rev. B* **2018**, *97*, 075201, DOI 10.1103/PhysRevB.97.075201.

[58] O. Pulci, W. G. Schmidt, F. Bechstedt, *Phys. status solidi* **2001**, *184*, 105–110, DOI 10.1002/1521-396X(200103)184:1<105::AID-PSSA105>3.0.CO;2-D.

[59] W. G. Schmidt, *Appl. Phys. A Mater. Sci. Process.* **2002**, *75*, 89–99, DOI 10.1007/s003390101058.

[60] P. H. Hahn, W. G. Schmidt, F. Bechstedt, O. Pulci, R. Del Sole, *Phys. Rev. B* **2003**, *68*, 033311, DOI 10.1103/PhysRevB.68.033311.

[61] D. Li, K. Liu, H. Xiao, H. Dong, X. Zu, *J. Alloys Compd.* **2007**, *440*, 229–235, DOI 10.1016/j.jallcom.2006.09.054.

[62] P. Rosenow, R. Tonner, *J. Chem. Phys.* **2016**, *144*, 204706, DOI 10.1063/1.4952603.

[63] A. Beyer, A. Stegmüller, J. O. Oelerich, K. Jandieri, K. Werner, G. Mette, W. Stolz, S. D. Baranovskii, R. Tonner, K. Volz, *Chem. Mater.* **2016**, *28*, 3265–3275, DOI 10.1021/acs.chemmater.5b04896.

[64] I. Kazadojev, D. J. Otway, S. D. Elliott, *Chem. Vap. Depos.* **2013**, *19*, 117–124, DOI 10.1002/cvde.201207025.

[65] A. Stegmüller, P. Rosenow, R. Tonner, *Phys. Chem. Chem. Phys.* **2014**, *16*, 17018–17029, DOI 10.1039/C4CP01584C.

[66] M. Imam, K. Gaul, A. Stegmüller, C. Höglund, J. Jensen, L. Hultman, J. Birch, R. Tonner, H. Pedersen, *J. Mater. Chem. C* **2015**, *3*, 10898–10906, DOI 10.1039/C5TC02293B.

[67] A. Stegmüller, R. Tonner, *Inorg. Chem.* **2015**, *54*, 6363–6372, DOI 10.1021/acs.inorgchem.5b00687.

[68] A. Stegmüller, R. Tonner, *Chem. Vap. Depos.* **2015**, *21*, 161–165, DOI 10.1002/cvde.201504332.

[69] E. Sterzer, A. Beyer, L. Duschek, L. Nattermann, B. Ringler, B. Leube, A. Stegmüller, R. Tonner, C. von Hänisch, W. Stolz, K. Volz, *J. Cryst. Growth* **2016**, *439*, 19–27, DOI 10.1016/j.jcrysgro.2015.12.032.

[70] M. Imam, L. Souqui, J. Herritsch, A. Stegmüller, C. Höglund, S. Schmidt, R. Hall-Wilton, H. Högberg, J. Birch, R. Tonner, H. Pedersen, *J. Phys. Chem. C* **2017**, *121*, 26465–26471, DOI 10.1021/acs.jpcc.7b09538.

[71] C. K. Ande, H. C. M. Knoops, K. de Peuter, M. van Drunen, S. D. Elliott, W. M. M. Kessels, *J. Phys. Chem. Lett.* **2015**, *6*, 3610–3614, DOI 10.1021/acs.jpclett.5b01596.

[72] A. Stegmüller, K. Werner, M. Reutzel, A. Beyer, P. Rosenow, U. Höfer, W. Stolz, K. Volz, M. Dürr, R. Tonner, *Chem. - A Eur. J.* **2016**, *22*, 14920–14928, DOI 10.1002/chem.201602418.

[73] S. D. Elliott, G. Scarel, C. Wiemer, M. Fanciulli, G. Pavia, *Chem. Mater.* **2006**, *18*, 3764–3773, DOI 10.1021/cm0608903.

[74] S. D. Elliott, *Langmuir* **2010**, *26*, 9179–9182, DOI 10.1021/la101207y.

[75] M. Shirazi, S. D. Elliott, *Chem. Mater.* **2013**, *25*, 878–889, DOI 10.1021/cm303630e.

[76] M. Shirazi, S. D. Elliott, *Nanoscale* **2015**, *7*, 6311–6318, DOI 10.1039/C5NR00900F.

[77] T. Weckman, K. Laasonen, *J. Phys. Chem. C* **2016**, *120*, 21460–21471, DOI 10.1021/acs.jpcc.6b06141.

[78] M. Shirazi, S. D. Elliott, *J. Comput. Chem.* **2014**, *35*, 244–259, DOI 10.1002/jcc.23491.

[79] T. Weckman, M. Shirazi, S. D. Elliott, K. Laasonen, *J. Phys. Chem. C* **2018**, *122*, 27044–27058, DOI 10.1021/acs.jpcc.8b06909.

[80] A. Szabo, S. O. Neil, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, rev. 1st ed., Dover Publications, Inc., Mineola, New York, **1996**.

[81] W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, 2nd ed., Wiley-VCH, Weinheim, **2001**.

[82] R. Dronskowski, *Computational Chemistry of Solid State Materials*, 1st ed., Wiley-VCH, Weinheim, **2005**.

[83] G. Czycholl, *Theoretische Festkörperphysik*, 3rd ed., Springer Verlag, Berlin, Heidelberg, **2008**.

[84] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed., Wiley-VCH, Weinheim, **2009**.

[85] R. Evarestov, *Quantum Chemistry of Solids*, 2nd ed., Springer-Verlag, Berlin Heidelberg, **2012**.

[86] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953–17979, DOI 10.1103/PhysRevB.50.17953.

[87] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775, DOI 10.1103/PhysRevB.59.1758.

[88] S. Steiner, S. Khmelevskyi, M. Marsmann, G. Kresse, *Phys. Rev. B* **2016**, *93*, 224425, DOI 10.1103/PhysRevB.93.224425.

[89] S. Kristyán, P. Pulay, *Chem. Phys. Lett.* **1994**, *229*, 175–180, DOI 10.1016/0009-2614(94)01027-7.

[90] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104, DOI 10.1063/1.3382344.

[91] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465, DOI 10.1002/jcc.21759.

[92] E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2006**, *124*, 174104, DOI 10.1063/1.2190220.

[93] L. Pecher, G. Mette, M. Dürr, R. Tonner, *ChemPhysChem* **2017**, *18*, 357–365, DOI 10.1002/cphc.201601134.

[94] D. Sheppard, R. Terrell, G. Henkelman, *J. Chem. Phys.* **2008**, *128*, 134106, DOI 10.1063/1.2841941.

[95] J. R. Shewchuk, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, **1994**.

[96] E. Polak, G. Ribiere, *ESAIM: Math. Model. Numer. Anal.* **1969**, *3*, 35–43.

[97] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, New York, USA, **1992**.

[98] J. Nocedal, *Math. Comput.* **1980**, *35*, 773–782, DOI 10.1090/S0025-5718-1980-0572855-7.

[99] G. Henkelman, B. P. Uberuaga, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9901–9904, DOI 10.1063/1.1329672.

[100] G. Henkelman, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9978–9985, DOI 10.1063/1.1323224.

[101] W. M. C. Sameera, S. Maeda, K. Morokuma, *Acc. Chem. Res.* **2016**, *49*, 763–773, DOI 10.1021/acs.accounts.6b00023.

[102] A. L. Dewyer, A. J. Argüelles, P. M. Zimmerman, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1354, DOI 10.1002/wcms.1354.

[103] C. A. Grambow, A. Jamal, Y.-P. Li, W. H. Green, J. Zádor, Y. V. Suleimanov, *J. Am. Chem. Soc.* **2018**, *140*, 1035–1048, DOI 10.1021/jacs.7b11009.

[104] G. N. Simm, A. C. Vaucher, M. Reiher, *J. Phys. Chem. A* **2019**, *123*, 385–399, DOI 10.1021/acs.jpca.8b10007.

[105] J. P. Unsleber, M. Reiher, *Annu. Rev. Phys. Chem.* **2020**, *71*, 121–142, DOI 10.1146/annurev-physchem-071119-040123.

[106] S. Maeda, Y. Harabuchi, *WIREs Comput. Mol. Sci.* **2021**, *11*, e1538, DOI 10.1002/wcms.1538.

[107] A. Baiardi, S. A. Grimmel, M. Steiner, P. L. Türtscher, J. P. Unsleber, T. Weymuth, M. Reiher, *Acc. Chem. Res.* **2022**, *55*, 35–43, DOI 10.1021/acs.accounts.1c00472.

[108] M. Steiner, M. Reiher, *Top. Catal.* **2022**, *65*, 6–39, DOI 10.1007/s11244-021-01543-9.

[109] C. W. Gao, J. W. Allen, W. H. Green, R. H. West, *Comput. Phys. Commun.* **2016**, *203*, 212–225, DOI 10.1016/j.cpc.2016.02.013.

[110] C. F. Goldsmith, R. H. West, *J. Phys. Chem. C* **2017**, *121*, 9970–9981, DOI 10.1021/acs.jpcc.7b02133.

[111] M. Liu, A. Grinberg Dana, M. S. Johnson, M. J. Goldman, A. Jocher, A. M. Payne, C. A. Grambow, K. Han, N. W. Yee, E. J. Mazeau, K. Blondal, R. H. West, C. F. Goldsmith, W. H. Green, *J. Chem. Inf. Model.* **2021**, *61*, 2686–2696, DOI 10.1021/acs.jcim.0c01480.

[112] P. L. Bhoorasingh, R. H. West, *Phys. Chem. Chem. Phys.* **2015**, *17*, 32173–32182, DOI 10.1039/C5CP04706D.

[113] E. J. Mazeau, P. Satpute, K. Blöndal, C. F. Goldsmith, R. H. West, *ACS Catal.* **2021**, *11*, 7114–7125, DOI 10.1021/acscatal.0c04100.

[114] P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain, R. H. West, *J. Phys. Chem. A* **2017**, *121*, 6896–6904, DOI 10.1021/acs.jpca.7b07361.

[115] K. Blondal, J. Jelic, E. Mazeau, F. Studt, R. H. West, C. F. Goldsmith, *Ind. Eng. Chem. Res.* **2019**, *58*, 17682–17691, DOI 10.1021/acs.iecr.9b01464.

[116] B. Kreitz, K. Sargsyan, K. Blöndal, E. J. Mazeau, R. H. West, G. D. Wehinger, T. Turek, C. F. Goldsmith, *JACS Au* **2021**, *1*, 1656–1673, DOI 10.1021/jacsau.1c00276.

[117] K. Fukui, *Acc. Chem. Res.* **1981**, *14*, 363–368, DOI 10.1021/ar00072a001.

[118] S. Rangarajan, A. Bhan, P. Daoutidis, *Comput. Chem. Eng.* **2012**, *45*, 114–123, DOI 10.1016/j.compchemeng.2012.06.008.

[119] D. Rappoport, C. J. Galvin, D. Y. Zubarev, A. Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, *10*, 897–907, DOI 10.1021/ct401004r.

[120] Y. Guan, V. M. Ingman, B. J. Rooks, S. E. Wheeler, *J. Chem. Theory Comput.* **2018**, *14*, 5249–5261, DOI 10.1021/acs.jctc.8b00578.

[121] D. Rappoport, A. Aspuru-Guzik, *J. Chem. Theory Comput.* **2019**, *15*, 4099–4112, DOI 10.1021/acs.jctc.9b00126.

[122] R. Van de Vijver, J. Zádor, *Comput. Phys. Commun.* **2020**, *248*, 106947, DOI 10.1016/j.cpc.2019.106947.

[123] V. M. Ingman, A. J. Schaefer, L. R. Andreola, S. E. Wheeler, *WIREs Comput. Mol. Sci.* **2021**, *11*, e1510, DOI 10.1002/wcms.1510.

[124] Z. W. Ulissi, A. J. Medford, T. Bligaard, J. K. Nørskov, *Nat. Commun.* **2017**, *8*, 14621, DOI 10.1038/ncomms14621.

[125] E. Martínez-Núñez, *J. Comput. Chem.* **2015**, *36*, 222–234, DOI 10.1002/jcc.23790.

[126] A. Rodríguez, R. Rodríguez-Fernández, S. A. Vázquez, G. L. Barnes, J. J. P. Stewart, E. Martínez-Núñez, *J. Comput. Chem.* **2018**, *39*, 1922–1930, DOI 10.1002/jcc.25370.

[127] S. A. Vázquez, X. L. Otero, E. Martínez-Núñez, *Molecules* **2018**, *23*, 3156, DOI 10.3390/molecules23123156.

[128] E. Martínez-Núñez, G. L. Barnes, D. R. Glowacki, S. Kopec, D. Peláez, A. Rodríguez, R. Rodríguez-Fernández, R. J. Shannon, J. J. P. Stewart, P. G. Tahoces, S. A. Vazquez, *J. Comput. Chem.* **2021**, *42*, 2036–2048, DOI 10.1002/jcc.26734.

[129] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande, T. J. Martínez, *Nat. Chem.* **2014**, *6*, 1044–1048, DOI 10.1038/nchem.2099.

[130] L.-P. Wang, R. T. McGibbon, V. S. Pande, T. J. Martínez, *J. Chem. Theory Comput.* **2016**, *12*, 638–649, DOI 10.1021/acs.jctc.5b00830.

[131] E. Pieri, D. Lahana, A. M. Chang, C. R. Aldaz, K. C. Thompson, T. J. Martínez, *Chem. Sci.* **2021**, *12*, 7294–7307, DOI 10.1039/D1SC00775K.

[132] A. Barducci, M. Bonomi, M. Parrinello, *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843, DOI 10.1002/wcms.31.

[133] G. Piccini, D. Mendels, M. Parrinello, *J. Chem. Theory Comput.* **2018**, *14*, 5040–5044, DOI 10.1021/acs.jctc.8b00634.

[134] I. Gimondi, G. A. Tribello, M. Salvalaglio, *J. Chem. Phys.* **2018**, *149*, 104104, DOI 10.1063/1.5027528.

[135] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862, DOI 10.1021/acs.jctc.9b00143.

[136] D. Mandelli, B. Hirshberg, M. Parrinello, *Phys. Rev. Lett.* **2020**, *125*, 026001, DOI 10.1103/PhysRevLett.125.026001.

[137] U. Raucci, V. Rizzi, M. Parrinello, *J. Phys. Chem. Lett.* **2022**, *13*, 1424–1430, DOI 10.1021/acs.jpclett.1c03993.

[138] X.-J. Zhang, C. Shang, Z.-P. Liu, *J. Chem. Theory Comput.* **2013**, *9*, 3252–3260, DOI 10.1021/ct400238j.

[139] C. Shang, X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2014**, *16*, 17845–17856, DOI 10.1039/C4CP01485E.

[140] X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2015**, *17*, 2757–2769, DOI 10.1039/C4CP04456H.

[141] C. Shang, Z.-P. Liu, *J. Chem. Theory Comput.* **2012**, *8*, 2215–2222, DOI 10.1021/ct300250h.

[142] X.-J. Zhang, C. Shang, Z.-P. Liu, *J. Chem. Theory Comput.* **2013**, *9*, 5745–5753, DOI 10.1021/ct4008475.

[143] X.-J. Zhang, C. Shang, Z.-P. Liu, *J. Chem. Theory Comput.* **2013**, *9*, 5745–5753, DOI 10.1021/ct4008475.

[144] X.-J. Zhang, C. Shang, Z.-P. Liu, *J. Chem. Phys.* **2017**, *147*, 152706, DOI 10.1063/1.4989540.

[145] S.-D. Huang, C. Shang, X.-J. Zhang, Z.-P. Liu, *Chem. Sci.* **2017**, *8*, 6327–6337, DOI 10.1039/c7sc01459g.

[146] S.-D. Huang, C. Shang, P.-L. Kang, X.-J. Zhang, Z.-P. Liu, *WIREs Comput. Mol. Sci.* **2019**, *9*, e1415, DOI 10.1002/wcms.1415.

[147] P. M. Zimmerman, *J. Comput. Chem.* **2013**, *34*, 1385–1392, DOI 10.1002/jcc.23271.

[148] P. M. Zimmerman, *Mol. Simul.* **2015**, *41*, 43–54, DOI 10.1080/08927022.2014.894999.

[149] A. L. Dewyer, P. M. Zimmerman, *Org. Biomol. Chem.* **2017**, *15*, 501–504, DOI 10.1039/C6OB02183B.

[150] B. Peters, A. Heyden, A. T. Bell, A. Chakraborty, *J. Chem. Phys.* **2004**, *120*, 7877–7886, DOI 10.1063/1.1691018.

[151] P. Zimmerman, *J. Chem. Theory Comput.* **2013**, *9*, 3043–3050, DOI 10.1021/ct400319w.

[152] P. M. Zimmerman, *J. Chem. Phys.* **2013**, *138*, 184102, DOI 10.1063/1.4804162.

[153] P. M. Zimmerman, *J. Comput. Chem.* **2015**, *36*, 601–611, DOI 10.1002/jcc.23833.

[154] M. Jafari, P. M. Zimmerman, *J. Comput. Chem.* **2017**, *38*, 645–658, DOI 10.1002/jcc.24720.

[155] G. Shrivastav, E. Vanden-Eijnden, C. F. Abrams, *J. Chem. Phys.* **2019**, *151*, 124112, DOI 10.1063/1.5120372.

[156] M. Jafari, P. M. Zimmerman, *Phys. Chem. Chem. Phys.* **2018**, *20*, 7721–7729, DOI 10.1039/C8CP00044A.

[157] Y. V. Suleimanov, W. H. Green, *J. Chem. Theory Comput.* **2015**, *11*, 4248–4259, DOI 10.1021/acs.jctc.5b00407.

[158] A. Behn, P. M. Zimmerman, A. T. Bell, M. Head-Gordon, *J. Chem. Phys.* **2011**, *135*, 224108, DOI 10.1063/1.3664901.

[159] C. Peng, H. Bernhard Schlegel, *Isr. J. Chem.* **1993**, *33*, 449–454, DOI 10.1002/ijch.199300051.

[160] P. Y. Ayala, H. B. Schlegel, *J. Chem. Phys.* **1997**, *107*, 375–384, DOI 10.1063/1.474398.

[161] S. Habershon, *J. Chem. Phys.* **2015**, *143*, 094106, DOI 10.1063/1.4929992.

[162] S. Habershon, *J. Chem. Theory Comput.* **2016**, *12*, 1786–1798, DOI 10.1021/acs.jctc.6b00005.

[163] Y. Kim, S. Choi, W. Y. Kim, *J. Chem. Theory Comput.* **2014**, *10*, 2419–2426, DOI 10.1021/ct500136x.

[164] M. Bergeler, G. N. Simm, J. Proppe, M. Reiher, *J. Chem. Theory Comput.* **2015**, *11*, 5712–5722, DOI 10.1021/acs.jctc.5b00866.

[165] S. A. Grimmel, M. Reiher, *Chimia* **2021**, *75*, 311, DOI 10.2533/chimia.2021.311.

[166] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683–3701, DOI 10.1039/c3cp44063j.

[167] S. Maeda, Y. Harabuchi, M. Takagi, K. Saita, K. Suzuki, T. Ichino, Y. Sumiya, K. Sugiyama, Y. Ono, *J. Comput. Chem.* **2018**, *39*, 233–250, DOI 10.1002/jcc.25106.

[168] K. Ohno, S. Maeda, *Chem. Phys. Lett.* **2004**, *384*, 277–282, DOI 10.1016/j.cplett.2003.12.030.

[169] S. Maeda, K. Ohno, *J. Phys. Chem. A* **2005**, *109*, 5742–5753, DOI 10.1021/jp0513162.

[170] K. Ohno, S. Maeda, *J. Phys. Chem. A* **2006**, *110*, 8933–8941, DOI 10.1021/jp061149l.

[171] S. Maeda, K. Morokuma, *J. Chem. Phys.* **2010**, *132*, 241102, DOI 10.1063/1.3457903.

[172] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2011**, *7*, 2335–2345, DOI 10.1021/ct200290m.

[173] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, *35*, 166–173, DOI 10.1002/jcc.23481.

[174] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec.* **2016**, *16*, 2232–2248, DOI 10.1002/tcr.201600043.

[175] T. Yoshimura, S. Maeda, T. Taketsugu, M. Sawamura, K. Morokuma, S. Mori, *Chem. Sci.* **2017**, *8*, 4475–4488, DOI 10.1039/C7SC00401J.

[176] C. Choi, R. Elber, *J. Chem. Phys.* **1991**, *94*, 751–760, DOI 10.1063/1.460343.

[177] K. Sugiyama, K. Saita, S. Maeda, *J. Comput. Chem.* **2021**, *42*, 2163–2169, DOI 10.1002/jcc.26746.

[178] S. Maeda, K. Sugiyama, Y. Sumiya, M. Takagi, K. Saita, *Chem. Lett.* **2018**, *47*, 396–399, DOI 10.1246/cl.171194.

[179] S. Ohno, K. Shudo, M. Tanaka, S. Maeda, K. Ohno, *J. Phys. Chem. C* **2010**, *114*, 15671–15677, DOI 10.1021/jp102883c.

[180] K. Sugiyama, Y. Sumiya, M. Takagi, K. Saita, S. Maeda, *Phys. Chem. Chem. Phys.* **2019**, *21*, 14366–14375, DOI 10.1039/C8CP06856A.

[181] F. Pieck, R. Tonner-Zech, PESE - A Potential Energy Surface Explorer for thin film growth, Version 0.1, **2022**, DOI 10.5281/zenodo.6973613.

[182] B. D. Dunnington, J. R. Schmidt, *J. Chem. Theory Comput.* **2012**, *8*, 1902–1911, DOI 10.1021/ct300002t.

[183] H. W. Kuhn, *Nav. Res. Logist. Q.* **1955**, *2*, 83–97, DOI 10.1002/nav.3800020109.

[184] J. Munkres, *J. Soc. Ind. Appl. Math.* **1957**, *5*, 32–38, DOI 10.1137/0105003.

[185] J. M. Vásquez-Pérez, G. U. G. Martínez, A. M. Köster, P. Calaminici, *J. Chem. Phys.* **2009**, *131*, 124126, DOI 10.1063/1.3231134.

[186] J. M. C. Marques, J. L. Llanio-Trujillo, P. E. Abreu, F. B. Pereira, *J. Chem. Inf. Model.* **2010**, *50*, 2129–2140, DOI 10.1021/ci100219f.

[187] M. Griffiths, S. P. Niblett, D. J. Wales, *J. Chem. Theory Comput.* **2017**, *13*, 4914–4931, DOI 10.1021/acs.jctc.7b00543.

[188] M. Andersen, C. Panosetti, K. Reuter, *Front. Chem.* **2019**, *7*, DOI 10.3389/fchem.2019.00202.

[189] K. Reuter, M. Scheffler, *Phys. Rev. B* **2006**, *73*, 045433, DOI 10.1103/PhysRevB.73.045433.

[190] D. A. Patterson, G. Gibson, R. H. Katz, *ACM SIGMOD Rec.* **1988**, *17*, 109–116, DOI 10.1145/971701.50214.

[191] S. R. Kachel, B. P. Klein, J. M. Morbec, M. Schöniger, M. Hutter, M. Schmid, P. Kratzer, B. Meyer, R. Tonner, J. M. Gottfried, *J. Phys. Chem. C* **2020**, *124*, 8257–8268, DOI 10.1021/acs.jpcc.0c00915.

[192] L. Pecher, S. Schmidt, R. Tonner, *J. Phys. Chem. C* **2017**, *121*, 26840–26850, DOI 10.1021/acs.jpcc.7b09148.

[193] G. Mette, C. H. Schwalb, M. Dürr, U. Höfer, *Chem. Phys. Lett.* **2009**, *483*, 209–213, DOI 10.1016/j.cplett.2009.10.073.

[194] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **1996**, *17*, 49–56, DOI 10.1002/(SICI)1096-987X(19960115)17:1<49::AID-JCC5>3.0.CO;2-0.

[195] O. Maßmeyer, J. Haust, T. Hepp, R. Günkel, J. Glowatzki, C. von Hänisch, W. Stolz, K. Volz, *ACS Omega* **2021**, *6*, 28229–28241, DOI 10.1021/acsomega.1c04316.

[196] F. O. Rice, E. Teller, *J. Chem. Phys.* **1938**, *6*, 489–496, DOI 10.1063/1.1750298.

[197] S. I. Miller, *Adv. Phys. Org. Chem.* **1968**, *6*, 185–332, DOI 10.1016/S0065-3160(08)60256-7.

[198] J. Hine, **1977**, *15*, 1–61, DOI 10.1016/S0065-3160(08)60117-3.

[199] R. Hoffmann, R. Gleiter, F. B. Mallory, *J. Am. Chem. Soc.* **1970**, *92*, 1460–1466, DOI 10.1021/ja00709a002.

[200] M. C. Reis, M. Alajarin, M. Marin-Luna, *Phys. Chem. Chem. Phys.* **2022**, *24*, 8064–8075, DOI 10.1039/D2CP00700B.

[201] P. Lindgren, G. Kastlunger, A. A. Peterson, *J. Chem. Theory Comput.* **2019**, *15*, 5787–5793, DOI 10.1021/acs.jctc.9b00633.

[202] E. L. Kolsbjerg, M. N. Groves, B. Hammer, *J. Chem. Phys.* **2016**, *145*, 094107, DOI 10.1063/1.4961868.

[203] B. Kaiser, D. Fertig, J. Ziegler, J. Klett, S. Hoch, W. Jaegermann, *ChemPhysChem* **2012**, *13*, 3053–3060, DOI 10.1002/cphc.201200432.

[204] M. Feifel, T. Rachow, J. Benick, J. Ohlmann, S. Janz, M. Hermle, F. Dimroth, D. Lackner, *IEEE J. Photovoltaics* **2016**, *6*, 384–390, DOI 10.1109/JPHOTOV.2015.2478062.

[205] E. Gillan in *Comprehensive Inorganic Chemistry II (Second Edition)*, (Eds.: J. Reedijk, K. Poeppelmeier), Elsevier, Amsterdam, **2013**, pp. 969–1000, DOI 10.1016/B978-0-08-097774-4.00132-7.

[206] B. Kunert, K. Volz, W. Stolz, *Phys. status solidi* **2007**, *244*, 2730–2739, DOI 10.1002/pssb.200675609.

[207] B. Kunert, K. Volz, J. Koch, W. Stolz, *Appl. Phys. Lett.* **2006**, *88*, 182108, DOI 10.1063/1.2200758.

[208] T. Wegele, A. Beyer, P. Ludewig, P. Rosenow, L. Duschek, K. Jandieri, R. Tonner, W. Stolz, K. Volz, *J. Phys. D. Appl. Phys.* **2016**, *49*, 075108, DOI 10.1088/0022-3727/49/7/075108.

[209] L. Nattermann, A. Beyer, P. Ludewig, T. Hepp, E. Sterzer, K. Volz, *J. Cryst. Growth* **2017**, *463*, 151–155, DOI 10.1016/j.jcrysgro.2017.02.021.

[210] O. Supplie, O. Romanyuk, C. Koppka, M. Steidl, A. Nägelein, A. Paszuk, L. Winterfeld, A. Dobrich, P. Kleinschmidt, E. Runge, T. Hannappel, *Prog. Cryst. Growth Charact. Mater.* **2018**, *64*, 103–132, DOI 10.1016/j.pcrysgrow.2018.07.002.

[211] A. Beyer, K. Volz, *Adv. Mater. Interfaces* **2019**, *6*, 1801951, DOI 10.1002/admi.201801951.

[212] S. F. Fang, K. Adomi, S. Iyer, H. Morkoç, H. Zabel, C. Choi, N. Otsuka, *J. Appl. Phys.* **1990**, *68*, R31–R58, DOI 10.1063/1.346284.

[213] H. Kroemer, *J. Cryst. Growth* **1987**, *81*, 193–204, DOI 10.1016/0022-0248(87)90391-5.

[214] J. T. Boyer, A. N. Blumer, Z. H. Blumer, D. L. Lepkowski, T. J. Grassman, *J. Cryst. Growth* **2021**, *571*, 126251, DOI 10.1016/j.jcrysgro.2021.126251.

[215] P. Farin, M. Marquardt, W. Martyanov, J. Belz, A. Beyer, K. Volz, A. Lenz, *J. Phys.: Condens. Matter* **2019**, *31*, 144001, DOI 10.1088/1361-648X/aafcfb.

[216] B. Kunert, I. Németh, S. Reinhard, K. Volz, W. Stolz, *Thin Solid Films* **2008**, *517*, 140–143, DOI 10.1016/j.tsf.2008.08.077.

[217] I. Németh, B. Kunert, W. Stolz, K. Volz, *J. Cryst. Growth* **2008**, *310*, 1595–1601, DOI 10.1016/j.jcrysgro.2007.11.127.

[218] A. Beyer, J. Ohlmann, S. Liebich, H. Heim, G. Witte, W. Stolz, K. Volz, *J. Appl. Phys.* **2012**, *111*, 083534, DOI 10.1063/1.4706573.

[219] B. Kunert, S. Zinnkann, K. Volz, W. Stolz, *J. Cryst. Growth* **2008**, *310*, 4776–4779, DOI 10.1016/j.jcrysgro.2008.07.097.

[220] O. Supplie, M. M. May, S. Brückner, N. Brezhneva, T. Hannappel, E. V. Skorb, *Adv. Mater. Interfaces* **2017**, *4*, 1601118, DOI 10.1002/admi.201601118.

[221] L. Töben, T. Hannappel, K. Möller, H.-J Crawack, C. Pettenkofer, F. Willig, *Surf. Sci.* **2001**, *494*, L755–L760, DOI 10.1016/S0039-6028(01)01492-3.

[222] H. Döscher, T. Hannappel, *J. Appl. Phys.* **2010**, *107*, 123523, DOI 10.1063/1.3357391.

[223] N. Esser, W. G. Schmidt, J. Bernholc, A. M. Frisch, P. Vogt, M. Zorn, M. Pristovsek, W. Richter, F. Bechstedt, T. Hannappel, S. Visbeck, *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct.* **1999**, *17*, 1691, DOI 10.1116/1.590810.

[224] A. M. Frisch, W. G. Schmidt, J. Bernholc, M. Pristovsek, N. Esser, W. Richter, *Phys. Rev. B* **1999**, *60*, 2488–2494, DOI 10.1103/PhysRevB.60.2488.

[225] W. G. Schmidt, J. L. Fattebert, J. Bernholc, F. Bechstedt, *Surf. Rev. Lett.* **1999**, *06*, 1159–1165, DOI 10.1142/S0218625X99001281.

[226] N. Sanada, S. Mochizuki, S. Ichikawa, N. Utsumi, M. Shimomura, G. Kaneda, A. Takeuchi, Y. Suzuki, Y. Fukuda, S. Tanaka, M. Kamata, *Surf. Sci.* **1999**, *419*, 120–127, DOI 10.1016/S0039-6028(98)00769-9.

[227] W. G. Schmidt, J. Bernholc, F. Bechstedt, *Appl. Surf. Sci.* **2000**, *166*, 179–184, DOI 10.1016/S0169-4332(00)00406-2.

[228] P. Kleinschmidt, H. Döscher, P. Vogt, T. Hannappel, *Phys. Rev. B* **2011**, *83*, 155316, DOI 10.1103/PhysRevB.83.155316.

[229] K. Lüdge, P. Vogt, O. Pulci, N. Esser, F. Bechstedt, W. Richter, *Phys. Rev. B* **2000**, *62*, 11046–11049, DOI 10.1103/PhysRevB.62.11046.

[230] O. Pulci, K. Lüdge, P. Vogt, N. Esser, W. G. Schmidt, W. Richter, F. Bechstedt, *Comput. Mater. Sci.* **2001**, *22*, 32–37, DOI 10.1016/S0927-0256(01)00161-6.

[231] J. Yoshinobu, *Prog. Surf. Sci.* **2004**, *77*, 37–70, DOI 10.1016/j.progsurf.2004.07.001.

[232] L. Pecher, R. Tonner, *Inorganics* **2018**, *6*, 17, DOI 10.3390/inorganics6010017.

[233] G. Kresse, J. Hafner, *Phys. Rev. B* **1993**, *47*, 558–561, DOI 10.1103/PhysRevB.47.558.

[234] G. Kresse, J. Hafner, *Phys. Rev. B* **1994**, *49*, 14251–14269, DOI 10.1103/PhysRevB.49.14251.

[235] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15–50, DOI 10.1016/0927-0256(96)00008-0.

[236] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169–11186, DOI 10.1103/PhysRevB.54.11169.

[237] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868, DOI 10.1103/PhysRevLett.77.3865.

[238] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1997**, *78*, 1396–1396, DOI 10.1103/PhysRevLett.78.1396.

[239] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* **1976**, *13*, 5188–5192, DOI 10.1103/PhysRevB.13.5188.

[240] I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan, *J. Appl. Phys.* **2001**, *89*, 5815–5875, DOI 10.1063/1.1368156.

[241] K. Shiraishi, *J. Phys. Soc. Jpn.* **1990**, *59*, 3455–3458, DOI 10.1143/JPSJ.59.3455.

[242] https://cccbdb.nist.gov, Optimized at PBE/aug-cc-pVTZ level of theory, Last accessed: 10.05.2022.

[243] R. D. Richards, F. Bastiman, J. S. Roberts, R. Beanland, D. Walker, J. P. R. David, *J. Cryst. Growth* **2015**, *425*, 237–240, DOI 10.1016/j.jcrysgro.2015.02.053.

[244] M. Yoshimoto in *Molecular Beam Epitaxy*, (Eds.: H. Asahi, Y. Horikoshi), John Wiley & Sons Ltd, Chichester, UK, **2019**, Chapter 23, pp. 381–394, DOI 10.1002/9781119354987.ch23.

[245] T. Hepp, J. Veletas, R. Günkel, O. Maßmeyer, J. Glowatzki, W. Stolz, S. Chatterjee, K. Volz, *Cryst. Growth Des.* **2021**, *21*, 6307–6313, DOI 10.1021/acs.cgd.1c00754.

[246] T. Hepp, J. Lehr, R. Günkel, O. Maßmeyer, J. Glowatzki, A. Ruiz Perez, S. Reinhard, W. Stolz, K. Volz, *Electron. Lett.* **2022**, *58*, 70–72, DOI 10.1049/ell2.12353.

[247] S. Francoeur, M.-J. Seong, A. Mascarenhas, S. Tixier, M. Adamcyk, T. Tiedje, *Appl. Phys. Lett.* **2003**, *82*, 3874–3876, DOI 10.1063/1.1581983.

[248] S. Tixier, M. Adamcyk, T. Tiedje, S. Francoeur, A. Mascarenhas, P. Wei, F. Schiettekatte, *Appl. Phys. Lett.* **2003**, *82*, 2245–2247, DOI 10.1063/1.1565499.

[249] K. Alberi, O. D. Dubon, W. Walukiewicz, K. M. Yu, K. Bertulis, A. Krotkus, *Appl. Phys. Lett.* **2007**, *91*, 051909, DOI 10.1063/1.2768312.

[250] Y. Zhang, A. Mascarenhas, L.-W. Wang, *Phys. Rev. B* **2005**, *71*, 155201, DOI 10.1103/PhysRevB.71.155201.

[251] B. Fluegel, S. Francoeur, A. Mascarenhas, S. Tixier, E. C. Young, T. Tiedje, *Phys. Rev. Lett.* **2006**, *97*, 067205, DOI 10.1103/PhysRevLett.97.067205.

[252] M. Usman, C. A. Broderick, A. Lindsay, E. P. O'Reilly, *Phys. Rev. B* **2011**, *84*, 245202, DOI 10.1103/PhysRevB.84.245202.

[253] S. J. Sweeney, S. R. Jin, *J. Appl. Phys.* **2013**, *113*, 043110, DOI 10.1063/1.4789624.

[254] T. M. Christian, D. A. Beaton, K. Alberi, B. Fluegel, A. Mascarenhas, *Appl. Phys. Express* **2015**, *8*, 061202, DOI 10.7567/APEX.8.061202.

[255] T. M. Christian, D. A. Beaton, A. Mascarenhas, K. Alberi in International Symposium on Clusters and Nanomaterials, *Vol. 10174*, (Eds.: P. Jena, A. K. Kandalam), Proc. SPIE, Bellingham, Washington USA, **2016**, 101740F, DOI 10.1117/12.2245432.

[256] T. M. Christian, B. Fluegel, D. A. Beaton, K. Alberi, A. Mascarenhas, *Jpn. J. Appl. Phys.* **2016**, *55*, 108002, DOI 10.7567/JJAP.55.108002.

[257] M. P. Polak, P. Scharoch, R. Kudrawiec, *Semicond. Sci. Technol.* **2015**, *30*, 094001, DOI 10.1088/0268-1242/30/9/094001.

[258] D. P. Samajdar, T. D. Das, S. Dhar, *Mater. Sci. Semicond. Process.* **2015**, *40*, 539–542, DOI 10.1016/j.mssp.2015.06.065.

[259] A. Assali, F. Kanouni, Q. Zou, R. Khenata, *Phys. Lett. A* **2020**, *384*, 126147, DOI 10.1016/j.physleta.2019.126147.

[260] B. U. Haq, R. Ahmed, M. Mohamad, A. Shaari, J. Rhee, S. AlFaify, M. B. Kanoun, S. Goumri-Said, *Curr. Appl. Phys.* **2017**, *17*, 162–168, DOI 10.1016/j.cap.2016.10.017.

[261] K. Volz, A. Beyer, W. Witte, J. Ohlmann, I. Németh, B. Kunert, W. Stolz, *J. Cryst. Growth* **2011**, *315*, 37–47, DOI 10.1016/j.jcrysgro.2010.10.036.

[262] W. K. Ford, T. Guo, S. L. Lantz, K. Wan, S. Chang, C. B. Duke, D. L. Lessor, *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct.* **1990**, *8*, 940–947, DOI 10.1116/1.584947.

[263] W. G. Schmidt, F. Bechstedt, G. P. Srivastava, *Surf. Sci. Rep.* **1996**, *25*, 141–223, DOI 10.1016/S0167-5729(96)00006-4.

[264] R. Whittle, A. Murphy, E. Dudzik, I. T. McGovern, A. Hempelmann, C. Nowak, D. R. T. Zahn, A. Cafolla, W. Braun, *J. Synchrotron Radiat.* **1995**, *2*, 256–260, DOI 10.1107/S0909049595009770.

[265] S. C. A. Gay, G. P. Srivastava, *Phys. Rev. B* **2000**, *61*, 2688–2698, DOI 10.1103/PhysRevB.61.2688.

[266] R. Miwa, E. Takahashi, *Surf. Sci.* **2004**, *566-568*, 949–955, DOI 10.1016/j.susc.2004.06.033.

[267] M. Hashassi, A. Ramzi, K. Kourchid, A. Rebey, M. Mbarki, *Mater. Res. Express* **2019**, *6*, 106303, DOI 10.1088/2053-1591/ab3751.

[268] A. Haaland, *Angew. Chem. Int. Ed.* **1989**, *28*, 992–1007, DOI 10.1002/anie.198909921.

[269] G. Frenking, *Angew. Chem. Int. Ed.* **2014**, *53*, 6040–6046, DOI 10.1002/anie.201311022.

[270] C. J. Kirkham, V. Brázdová, D. R. Bowler, *Phys. Rev. B* **2012**, *86*, 035328, DOI 10.1103/PhysRevB.86.035328.

[271] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **2008**, *321*, 792–794, DOI 10.1126/science.1158722.

[272] Y. Ishikawa, K. Yazaki, I. Nakamichi, *Jpn. J. Appl. Phys.* **1989**, *28*, 1272–1273, DOI 10.1143/JJAP.28.1272.

[273] A. J. Downs, C. R. Pulham, *Chem. Soc. Rev.* **1994**, *23*, 175–184, DOI 10.1039/cs9942300175.

[274] S. Aldridge in *The Group 13 Metals Aluminium, Gallium, Indium and Thallium: Chemical Patterns and Peculiarities*, (Eds.: S. Aldridge, A. J. Downs), John Wiley &

Sons, Ltd, Chichester ,UK, **2011**, Chapter 2, pp. 75–147, DOI 10.1002/9780470976548.ch2.

[275] A. J. Downs, T. M. Greene, E. Johnsen, C. R. Pulham, H. E. Robertson, D. A. Wann, *Dalt. Trans.* **2010**, *39*, 5637–5642, DOI 10.1039/c000694g.

[276] M. Imade, M. Kawahara, F. Kawamura, M. Yoshimura, Y. Mori, T. Sasaki, *Mater. Lett.* **2005**, *59*, 4026–4029, DOI 10.1016/j.matlet.2005.07.063.

[277] J. S. Foord, T. J. Whitaker, E. N. Downing, D. O'Hare, A. C. Jones, *Appl. Phys. Lett.* **1993**, *63*, 1270–1272, DOI 10.1063/1.109754.

[278] T. F. Kuech, *Prog. Cryst. Growth Charact. Mater.* **2016**, *62*, 352–370, DOI 10.1016/j.pcrysgrow.2016.04.019.

[279] E. Sterzer, A. Beyer, L. Nattermann, W. Schorn, K. Schlechter, S. Pulz, J. Sundermeyer, W. Stolz, K. Volz, *J. Cryst. Growth* **2016**, *454*, 173–179, DOI 10.1016/j.jcrysgro.2016.08.061.

[280] H. Hardtdegen, M. Mikulics in *Metalorganic Vapor Phase Epitaxy*, (Eds.: S. Irvine, P. Capper), John Wiley & Sons, Ltd, Chichester, UK, **2019**, Chapter 3, pp. 71–108, DOI 10.1002/9781119313021.ch3.

[281] N. A. Besley, J. A. Bryan, *J. Phys. Chem. C* **2008**, *112*, 4308–4314, DOI 10.1021/jp076167x.

[282] J. Glowatzki, O. Maßmeyer, M. Köster, T. Hepp, E. Odofin, C. von Hänisch, W. Stolz, K. Volz, *Organometallics* **2020**, *39*, 1772–1781, DOI 10.1021/acs.organomet.0c00078.

[283] F. Kalantari Fotooh, T. Askari Baghemiyani, *Iran. J. Chem.* **2020**, *2*, 187–196, DOI 10.22036/ijc.2019.168938.1056.

[284] S. Ariaei, *Lab-in-Silico* **2020**, *1*, 44–49, DOI 10.22034/labinsilico20012044.

[285] R. Miotto, A. C. Ferraz, G. Srivastava, *Brazilian J. Phys.* **2002**, *32*, 392–395, DOI 10.1590/S0103-97332002000200041.

[286] S. R. Schofield, N. J. Curson, O. Warschkow, N. A. Marks, H. F. Wilson, M. Y. Simmons, P. V. Smith, M. W. Radny, D. R. McKenzie, R. G. Clark, *J. Phys. Chem. B* **2006**, *110*, 3173–3179, DOI 10.1021/jp054646v.

[287] O. Warschkow, N. J. Curson, S. R. Schofield, N. A. Marks, H. F. Wilson, M. W. Radny, P. V. Smith, T. C. G. Reusch, D. R. McKenzie, M. Y. Simmons, *J. Chem. Phys.* **2016**, *144*, 014705, DOI 10.1063/1.4939124.

[288] R. L. Woo, U. Das, S. F. Cheng, G. Chen, K. Raghavachari, R. F. Hicks, *Surf. Sci.* **2006**, *600*, 4888–4895, DOI 10.1016/j.susc.2006.08.014.

[289] U. Das, K. Raghavachari, R. L. Woo, R. F. Hicks, *Langmuir* **2007**, *23*, 10109–10115, DOI 10.1021/la700790h.

[290] L. Pecher, S. Schmidt, R. Tonner, *J. Phys. Chem. C* **2017**, *121*, 26840–26850, DOI 10.1021/acs.jpcc.7b09148.

[291] L. Pecher, S. Schmidt, R. Tonner, *Beilstein J. Org. Chem.* **2018**, *14*, 2715–2721, DOI 10.3762/bjoc.14.249.

[292] C. A. Larsen, N. I. Buchan, G. B. Stringfellow, *J. Cryst. Growth* **1987**, *85*, 148–153, DOI 10.1016/0022-0248(87)90216-8.

[293] Y. Sun, D. C. Law, R. F. Hicks, *Surf. Sci.* **2003**, *540*, 12–22, DOI 10.1016/S0039-6028(03)00834-3.

[294] V. Stegailov, G. Smirnov, V. Vecher, *Concurr. Comput. Pract. Exp.* **2019**, *31*, e5136, DOI 10.1002/cpe.5136.

[295] https://kb.hlrs.de/platforms/index.php/HPE_Hawk_Hardware_and_Architecture, Last accessed: 30.07.2022.

[296] S. Höfinger, E. Haunschmid, *J. Supercomput.* **2017**, *73*, 4390–4406, DOI 10.1007/s11227-017-2023-9.

[297] M. Hacene, A. Anciaux-Sedrakian, X. Rozanska, D. Klahr, T. Guignon, P. Fleurat-Lessard, *J. Comput. Chem.* **2012**, *33*, 2581–2589, DOI 10.1002/jcc.23096.

[298] W. P. Huhn, B. Lange, V. W. Yu, M. Yoon, V. Blum, *Comput. Phys. Commun.* **2020**, *254*, 107314, DOI 10.1016/j.cpc.2020.107314.

[299] D. B. Williams-Young, W. A. de Jong, H. J. J. van Dam, C. Yang, *Front. Chem.* **2020**, *8*, 581058, DOI 10.3389/fchem.2020.581058.

[300] X. Andrade, C. D. Pemmaraju, A. Kartsev, J. Xiao, A. Lindenberg, S. Rajpurohit, L. Z. Tan, T. Ogitsu, A. A. Correa, *J. Chem. Theory Comput.* **2021**, *17*, 7447–7467, DOI 10.1021/acs.jctc.1c00562.

[301] R. Meyer, K. S. Schmuck, A. W. Hauser, *J. Chem. Theory Comput.* **2019**, *15*, 6513–6523, DOI 10.1021/acs.jctc.9b00708.

[302] O.-P. Koistinen, V. Ásgeirsson, A. Vehtari, H. Jónsson, *J. Chem. Theory Comput.* **2019**, *15*, 6738–6751, DOI 10.1021/acs.jctc.9b00692.

[303] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, *Chem. Rev.* **2021**, *121*, 10073–10141, DOI 10.1021/acs.chemrev.1c00022.

[304] D. Born, J. Kästner, *J. Chem. Theory Comput.* **2021**, *17*, 5955–5967, DOI 10.1021/acs.jctc.1c00517.

[305] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *WIREs Comput. Mol. Sci.* **2021**, *11*, e1493, DOI 10.1002/wcms.1493.

[306] R. Clark, K. Tapily, K.-H. Yu, T. Hakamata, S. Consiglio, D. O'Meara, C. Wajda, J. Smith, G. Leusink, *APL Mater.* **2018**, *6*, 058203, DOI 10.1063/1.5026805.

[307] H.-B.-R. Lee, *Chem. Mater.* **2019**, *31*, 1471–1472, DOI 10.1021/acs.chemmater.9b00654.

[308] G. N. Parsons, R. D. Clark, *Chem. Mater.* **2020**, *32*, 4920–4953, DOI 10.1021/acs.chemmater.0c00722.

[309] H.-D. Yu, M. D. Regulacio, E. Ye, M.-Y. Han, *Chem. Soc. Rev.* **2013**, *42*, 6006–6018, DOI 10.1039/c3cs60113g.

[310] H. J. Levinson, T. A. Brunner in Int. Conf. Extrem. Ultrav. Lithogr. 2018, *Vol. 10809*, (Eds.: K. G. Ronse, P. A. Gargini, E. Hendrickx, P. P. Naulleau, T. Itani), Proc. SPIE, Bellingham, Washington USA, **2018**, pp. 5–11, DOI 10.1117/12.2502791.

[311] A. J. M. Mackus, M. J. M. Merkx, W. M. M. Kessels, *Chem. Mater.* **2019**, *31*, 2–12, DOI 10.1021/acs.chemmater.8b03454.

[312] H.-B.-R. Lee, S. F. Bent, *Chem. Mater.* **2020**, *32*, 3323–3324, DOI 10.1021/acs.chemmater.0c00838.

[313] K. Cao, J. Cai, R. Chen, *Chem. Mater.* **2020**, *32*, 2195–2207, DOI 10.1021/acs.chemmater.9b04647.

[314] J. Lee, J. Lee, H. Oh, C. Kim, J. Kim, D. H. Kim, B. Shong, T. J. Park, W. Kim, *Adv. Funct. Mater.* **2021**, *31*, 2102556, DOI 10.1002/adfm.202102556.

[315] Y.-C. Li, K. Cao, Y.-X. Lan, J.-M. Zhang, M. Gong, Y.-W. Wen, B. Shan, R. Chen, *Molecules* **2021**, *26*, 3056, DOI 10.3390/molecules26103056.

[316] D. Bobb-Semple, K. L. Nardi, N. Draeger, D. M. Hausmann, S. F. Bent, *Chem. Mater.* **2019**, *31*, 1635–1645, DOI 10.1021/acs.chemmater.8b04926.

[317] M. Bonvalot, C. Vallée, C. Mannequin, M. Jaffal, R. Gassilloud, N. Possémé, T. Chevolleau, *Dalton Trans.* **2022**, *51*, 442–450, DOI 10.1039/D1DT03456A.

[318] P. R. Chalker, P. A. Marshall, K. Dawson, I. F. Brunell, C. J. Sutcliffe, R. J. Potter, *AIP Adv.* **2015**, *5*, 017115, DOI 10.1063/1.4905887.

[319] M. Singh, N. Kaur, E. Comini, *J. Mater. Chem. C* **2020**, *8*, 3938–3955, DOI 10.1039/D0TC00388C.

[320] T.-L. Liu, K. L. Nardi, N. Draeger, D. M. Hausmann, S. F. Bent, *ACS Appl. Mater. Interfaces* **2020**, *12*, 42226–42235, DOI 10.1021/acsami.0c08873.

[321] D. Bobb-Semple, L. Zeng, I. Cordova, D. S. Bergsman, D. Nordlund, S. F. Bent, *Langmuir* **2020**, *36*, 12849–12857, DOI 10.1021/acs.langmuir.0c01974.

[322] J. Yarbrough, A. B. Shearer, S. F. Bent, *J. Vac. Sci. Technol. A* **2021**, *39*, 021002, DOI 10.1116/6.0000840.

[323] R. Chen, H. Kim, P. C. McIntyre, D. W. Porter, S. F. Bent, *Appl. Phys. Lett.* **2005**, *86*, 191910, DOI 10.1063/1.1922076.

[324] J. Hong, D. W. Porter, R. Sreenivasan, P. C. McIntyre, S. F. Bent, *Langmuir* **2007**, *23*, 1160–1165, DOI 10.1021/la0606401.

[325] A. Simon, O. van der Straten, N. A. Lanzillo, C.-C. Yang, T. Nogami, D. C. Edelstein, *J. Vac. Sci. Technol. A* **2020**, *38*, 053402, DOI 10.1116/6.0000170.

[326] M. E. Straumanis, L. S. Yu, *Acta Crystallogr. Sect. A* **1969**, *25*, 676–682, DOI 10.1107/S0567739469001549.

[327] Y. Le Page, G. Donnay, *Acta Crystallogr. Sect. B* **1976**, *32*, 2456–2459, DOI 10.1107/S0567740876007966.

[328] J. Yang, E. G. Wang, *Phys. Rev. B* **2006**, *73*, 035406, DOI 10.1103/PhysRevB.73.035406.

[329] T. P. M. Goumans, A. Wander, W. A. Brown, C. R. A. Catlow, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2146–2152, DOI 10.1039/B701176H.

[330] F. Musso, M. Sodupe, M. Corno, P. Ugliengo, *J. Phys. Chem. C* **2009**, *113*, 17876–17884, DOI 10.1021/jp905325m.

[331] H. Wang, Z. Chai, D. Wang, *Dalt. Trans.* **2015**, *44*, 1646–1654, DOI 10.1039/C4DT02872D.

[332] L. T. Zhuravlev, *Colloids Surf. A: Physicochem. Eng. Asp.* **2000**, *173*, 1–38, DOI 10.1016/S0927-7757(00)00556-2.

[333] A. Delabie, S. Sioncke, J. Rip, S. Van Elshocht, G. Pourtois, M. Mueller, B. Beckhoff, K. Pierloot, *J. Vac. Sci. Technol. A* **2012**, *30*, 01A127, DOI 10.1116/1.3664090.

[334] C. S. Ewing, S. Bhavsar, G. Veser, J. J. McCarthy, J. K. Johnson, *Langmuir* **2014**, *30*, 5133–5141, DOI 10.1021/la500422p.

[335] A. S. Sandupatla, K. Alexopoulos, M.-F. Reyniers, G. B. Marin, *J. Phys. Chem. C* **2015**, *119*, 18380–18388, DOI 10.1021/acs.jpcc.5b05261.

[336] G. Henkelman, H. Jónsson, *J. Chem. Phys.* **1999**, *111*, 7010–7022, DOI 10.1063/1.480097.

[337] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao,

N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 09 Revision E.01, Gaussian Inc. Wallingford CT, 2016.

[338] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305, DOI 10.1039/b508541a.

[339] C. Peng, B. H. Schlegel, *Isr. J. Chem.* **1993**, *33*, 449–454, DOI 10.1002/ijch.199300051.

[340] C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, *J. Comput. Chem.* **1996**, *17*, 49–56, DOI 10.1002/(SICI)1096-987X(19960115)17:1<49::AID-JCC5>3.0.CO;2-0.

[341] J. Yarbrough, F. Pieck, D. Grigjanis, I.-K. Oh, P. Maue, R. Tonner-Zech, S. F. Bent, *Chem. Mater.* **2022**, *34*, 4646–4659, DOI 10.1021/acs.chemmater.2c00513.

[342] A. Ulman, *Chem. Rev.* **1996**, *96*, 1533–1554, DOI 10.1021/cr9502357.

[343] Y. Xu, C. B. Musgrave, *Chem. Mater.* **2004**, *16*, 646–653, DOI 10.1021/cm035009p.

[344] M. Haran, P. Clancy, *J. Mater. Chem.* **2007**, *17*, 3927–3937, DOI 10.1039/b702664c.

[345] S. Patwardhan, D. H. Cao, G. C. Schatz, A. B. F. Martinson, *ACS Appl. Energy Mater.* **2019**, *2*, 4618–4628, DOI 10.1021/acsaem.8b02202.

[346] R. Chen, H. Kim, P. C. McIntyre, S. F. Bent, *Chem. Mater.* **2005**, *17*, 536–544, DOI 10.1021/cm0486666.

[347] A. Chandekar, S. K. Sengupta, J. E. Whitten, *Appl. Surf. Sci.* **2010**, *256*, 2742–2749, DOI 10.1016/j.apsusc.2009.11.020.

[348] I. Zyulkov, V. Madhiwala, E. Voronina, M. Snelgrove, J. Bogan, R. O'Connor, S. De Gendt, S. Armini, *ACS Appl. Mater. Interfaces* **2020**, *12*, 4678–4688, DOI 10.1021/acsami.9b14596.

[349] T. Takahagi, H. Sakaue, S. Shingubara, *Jpn. J. Appl. Phys.* **2001**, *40*, 6198–6201, DOI 10.1143/JJAP.40.6198.

[350] D. B. Asay, S. H. Kim, *J. Phys. Chem. B* **2005**, *109*, 16760–16763, DOI 10.1021/jp053042o.

[351] I.-K. Oh, T. E. Sandoval, T.-L. Liu, N. E. Richey, S. F. Bent, *Chem. Mater.* **2021**, *33*, 3926–3935, DOI 10.1021/acs.chemmater.0c04718.

[352] J. Yarbrough, F. Pieck, A. B. Shearer, P. Maue, R. Tonner-Zech, S. F. Bent, manuscript in preparation.

[353] J. R. Avila, E. J. DeMarco, J. D. Emery, O. K. Farha, M. J. Pellin, J. T. Hupp, A. B. F. Martinson, *ACS Appl. Mater. Interfaces* **2014**, *6*, 11891–11898, DOI 10.1021/am503008j.

[354] F. S. Minaye Hashemi, B. R. Birchansky, S. F. Bent, *ACS Appl. Mater. Interfaces* **2016**, *8*, 33264–33272, DOI 10.1021/acsami.6b09960.

[355] H. Magnusson, K. Frisk, *J. Phase Equilibria Diffus.* **2017**, *38*, 65–69, DOI 10.1007/s11669-017-0518-y.

[356] S. Zoha, B. Gu, F. Pieck, R. Tonner-Zech, H.-B.-R. Lee, manuscript in preparation.

[357] J. M. McLellan, M. Geissler, Y. Xia, *Chem. Phys. Lett.* **2005**, *408*, 80–83, DOI 10.1016/j.cplett.2005.04.010.

[358] E. Färm, M. Vehkamäki, M. Ritala, M. Leskelä, *Semicond. Sci. Technol.* **2012**, *27*, 074004, DOI 10.1088/0268-1242/27/7/074004.

[359] M. D. Sampson, J. D. Emery, M. J. Pellin, A. B. F. Martinson, *ACS Appl. Mater. Interfaces* **2017**, *9*, 33429–33436, DOI 10.1021/acsami.7b01410.

[360] D. S. Bergsman, T.-L. Liu, R. G. Closser, K. L. Nardi, N. Draeger, D. M. Hausmann, S. F. Bent, *Chem. Mater.* **2018**, *30*, 5694–5703, DOI 10.1021/acs.chemmater.8b02150.

[361] T.-L. Liu, S. F. Bent in Adv. Patterning Mater. Process. XXXVI, (Eds.: R. Gronheid, D. P. Sanders), Proc. SPIE, Bellingham, Washington USA, **2019**, pp. 56 –62, DOI 10.1117/12.2519845.

[362] H. G. Kim, M. Kim, B. Gu, M. R. Khan, B. G. Ko, S. Yasmeen, C. S. Kim, S.-H. Kwon, J. Kim, J. Kwon, K. Jin, B. Cho, J.-S. Chun, B. Shong, H.-B.-R. Lee, *Chem. Mater.* **2020**, *32*, 8921–8929, DOI 10.1021/acs.chemmater.0c02798.

[363] C. M. Schaefer, J. M. Caicedo Roque, G. Sauthier, J. Bousquet, C. Hébert, J. R. Sperling, A. Pérez-Tomás, J. Santiso, E. del Corro, J. A. Garrido, *Chem. Mater.* **2021**, *33*, 4474–4487, DOI 10.1021/acs.chemmater.1c00646.

[364] W. Jeon, Y. Cho, S. Jo, J.-H. Ahn, S.-J. Jeong, *Adv. Mater.* **2017**, *29*, 1703031, DOI 10.1002/adma.201703031.

[365] G. G. Slabaugh, Computing Euler angles from a rotation matrix, http://eecs.qmul.ac.uk/~gslabaugh/publications/euler.pdf, Last accessed 03.03.2022.

[366] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, New York, **1955**.

[367] R. A. Pilgrim, http://csclab.murraystate.edu/~bob.pilgrim/445/munkres.html, https://brc2.com/the-algorithm-workshop/, Last accessed 28.02.2022, **2000**.

[368] W. Kabsch, *Acta Crystallogr. Sect. A* **1976**, *32*, 922–923, DOI 10.1107/S0567739476001873.

[369] W. Kabsch, *Acta Crystallogr. Sect. A* **1978**, *34*, 827–828, DOI 10.1107/S0567739478001680.

[370] K. G. Murty, *Oper. Res.* **1968**, *16*, 682–687, DOI 10.1287/opre.16.3.682.

A

Mathematics

A.1. Basic Math

Root Mean Square Deviation

The definition for the RMSD between two structures $\mathbb{R}_1$ and $\mathbb{R}_2$ is shown in equation A.1. Here, $\mathbf{R}_{1,A}$ and $\mathbf{R}_{2,A}$ are the atomic coordinates of the M atoms in structure $\mathbb{R}_1$ and $\mathbb{R}_2$, respectively.

$$\begin{aligned} \text{RMSD}(\mathbb{R}_1, \mathbb{R}_2) &= \sqrt{\frac{1}{M}\sum_{A=1}^{M}(R_{1,A}^x - R_{2,A}^x)^2 + (R_{1,A}^y - R_{2,A}^y)^2 + (R_{1,A}^z - R_{2,A}^z)^2} \\ &= \sqrt{\frac{1}{M}\sum_{A=1}^{M}\|\mathbf{R}_{1,A} - \mathbf{R}_{2,A}\|^2} \end{aligned} \tag{A.1}$$

Centroid and Center of Mass

The centroid $\mathbf{R}_{cen}$ and the center of mass $\mathbf{R}_{com}$ of a structure are calculated based on the coordinates $\mathbf{R}_A$ of the M atoms. As shown in equation A.2 the coordinates are weighted based on the atomic mass m_A relative to the total mass of the system m_{total} for the center of mass.

$$\begin{aligned} \mathbf{R}_{cen} &= \frac{1}{M}\sum_{A=1}^{M}\mathbf{R}_A \\ \mathbf{R}_{com} &= \frac{1}{m_{total}}\sum_{A=1}^{M} m_A\mathbf{R}_A \end{aligned} \tag{A.2}$$

Matrix Inversion

For squared, positive defined matrices $\mathbb{A}$ the Cholesky decomposition (equation A.3) is an efficient approach do calculate the inverse matrix.

$$\mathbb{A} = \mathbb{L}\mathbb{L}^T \tag{A.3}$$

Here, the common definition of an inverse matrix in equation A.4 is formulated in a column wise manner (equation A.5) where **i** and **e** are the columns of the inverse matrix $\mathbb{A}^{-1}$ and the unit matrix $\mathbb{E}$, respectively.

$$\mathbb{A}\mathbb{A}^{-1} = \mathbb{E} \tag{A.4}$$

$$\mathbb{A}\mathbf{i} = \mathbf{e} \tag{A.5}$$

The advantage of this formulation is that the columns **i** can easily be calculated based on the triangular matrices $\mathbb{L}$ and $\mathbb{L}^T$. First the vectors **y** are obtained by forward substitution as shown in equation A.6.

$$\mathbb{L} \cdot \mathbf{y} = \mathbf{e} \tag{A.6}$$

Then the columns **i** of the inverse matrix are obtained by backward substitution as shown in equation A.7.

$$\mathbb{L}^T \cdot \mathbf{i} = \mathbf{y} \tag{A.7}$$

For non-squared matrices like the Wilson B matrix the calculation of an inverse matrix is not possible. Here, a pseudo-inverse can be derived. For this the matrix $\mathbb{B}$ is decomposed by the singular value decomposition (SVD) as in equation A.8.

$$\mathbb{B} = \mathbb{V}\mathbb{S}\mathbb{W}^T \tag{A.8}$$

The pseudo-inverse is then obtained by inverting the singular values $\mathbb{S}$. To promote numerical stability, only diagonal entries larger than 0.1 are inverted in PESE. Smaller values are set to zero. The pseudo-inverse is then obtained by equation A.9.

$$\mathbb{B}^{-1} = \mathbb{W}\mathbb{S}^{-1}\mathbb{V}^T \tag{A.9}$$

A.2. Finite Differences

Finite differences can be used to calculate the derivative of a function at any point x by calculating the function value at small displacements $x \pm \Delta x$. In equation A.10 the central difference is shown (displacement by a positive and negative difference).

$$f'(x) = \frac{f(x + \frac{1}{2}\Delta x) - f(x - \frac{1}{2}\Delta x)}{\Delta x} \tag{A.10}$$

The second derivative of a function at point x can also be obtained by finite differences (equation A.11). Here, the first derivative of the function can be used in case it is known. Otherwise, the second derivative is derived based on function values.

$$\begin{aligned} f''(x) &= \frac{f'(x + \frac{1}{2}\Delta x) - f'(x - \frac{1}{2}\Delta x)}{\Delta x} \\ &= \frac{f(x + \Delta x) - 2f(x) + f(x - \Delta x)}{\Delta x^2} \end{aligned} \tag{A.11}$$

For the calculation of the second derivative of the energy with respect to the atomic positions the force can be utilized resulting in equation A.12.

$$E''(\mathbb{R}) = \frac{\mathbb{F}(\mathbb{R} - \frac{1}{2}\Delta\mathbb{R}) - \mathbb{F}(\mathbb{R} + \frac{1}{2}\Delta\mathbb{R})}{\Delta\mathbb{R}} \tag{A.12}$$

A.3. Describing Rotations

General Structure of a Rotation Matrix

A rotation matrix $\mathbb{U}$ is built upon rotations around the principal axes of the coordinate system. Within this work I follow the order of the rotations shown in equation A.13.

$$\mathbb{U} = \mathbb{Z}\mathbb{Y}\mathbb{X} \tag{A.13}$$

Here, the individual rotations around each principal axis are defined by the rotation angles ψ, θ and ϕ as shown in equation A.14.

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\psi) & -\sin(\psi) \\ 0 & \sin(\psi) & \cos(\psi) \end{pmatrix}}_{\mathbb{X}} \quad \underbrace{\begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}}_{\mathbb{Y}} \quad \underbrace{\begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbb{Z}} \tag{A.14}$$

With the individual rotation matrices the final rotation matrix $\mathbb{U}$ is obtained as shown in equation A.15. The rotation described by this rotation matrix is an extrinsic rotation. Therefore, for all types of structural comparisons it is necessary to center the structure by for example its centroid or center of mass after the rotation.

$$\underbrace{\begin{pmatrix} \cos(\theta)\cos(\phi) & \sin(\psi)\sin(\theta)\cos(\phi) - \cos(\psi)\sin(\phi) & \cos(\psi)\sin(\theta)\cos(\phi) + \sin(\psi)\sin(\phi) \\ \cos(\theta)\sin(\phi) & \sin(\psi)\sin(\theta)\sin(\phi) + \cos(\psi)\cos(\phi) & \cos(\psi)\sin(\theta)\sin(\phi) - \sin(\psi)\cos(\phi) \\ -\sin(\theta) & \sin(\psi)\cos(\theta) & \cos(\psi)\cos(\theta) \end{pmatrix}}_{\mathbb{U}} \tag{A.15}$$

Vector Rotation

To align two vectors **a** and **b** by a rotation only a single rotation angle α is needed. This angle describes the rotation of one vector in the plane spanned by both vectors. The angle α is obtained by the scalar product of the two vectors as shown in equation A.16.

$$\cos(\alpha) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \tag{A.16}$$

In the implementation of this equations $\cos(\alpha)$ should be checked to be in the range -1 to 1, i.e. α is in the range π to 0, to prevent any abnormal termination due to numeric noise.

For the rotation a normalized normal **n** of the plane spanned by the vectors **a** and **b** is needed. This normal is obtained by the cross product of **a** and **b** (equation A.17).

$$\mathbf{n} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|} \tag{A.17}$$

This definition of the normal only works in case **a** and **b** are neither parallel nor antiparallel, i.e. α must not be close to 0 or π. Otherwise, **n** has to be defined by using the elements of the vector **a** as shown in equation A.18 to get a non-zero vector. An equivalent formulation for **n** is obtained by using the corresponding elements of the vector **b**.

$$\mathbf{n} = \begin{pmatrix} -a_2 \\ a_1 \\ 0 \end{pmatrix} \quad \text{or} \quad \mathbf{n} = \begin{pmatrix} 0 \\ -a_3 \\ a_2 \end{pmatrix} \tag{A.18}$$

A rotation matrix $\mathbb{U}$ with the elements U_{ij} describing the rotation can then be defined by the

elements of the normal **n** and the rotation angle α as shown in equation A.19.

$$\begin{aligned}
U_{11} &= n_1^2 \cdot (1 - \cos(\alpha)) + \cos(\alpha) \\
U_{12} &= n_1 \cdot n_2 \cdot (1 - \cos(\alpha)) - n_3 \cdot \sin(\alpha) \\
U_{13} &= n_1 \cdot n_3 \cdot (1 - \cos(\alpha)) + n_2 \cdot \sin(\alpha) \\
U_{21} &= n_2 \cdot n_1 \cdot (1 - \cos(\alpha)) + n_3 \cdot \sin(\alpha) \\
U_{22} &= n_2^2 \cdot (1 - \cos(\alpha)) + \cos(\alpha) \\
U_{23} &= n_2 \cdot n_3 \cdot (1 - \cos(\alpha)) - n_1 \cdot \sin(\alpha) \\
U_{31} &= n_3 \cdot n_1 \cdot (1 - \cos(\alpha)) - n_2 \cdot \sin(\alpha) \\
U_{32} &= n_3 \cdot n_2 \cdot (1 - \cos(\alpha)) + n_1 \cdot \sin(\alpha) \\
U_{33} &= n_3^2 \cdot (1 - \cos(\alpha)) + \cos(\alpha)
\end{aligned} \tag{A.19}$$

Decomposition of a Rotation Matrix in Individual Rotations

Following the equations described by Gregory G. Slabaugh[365] a rotation matrix $\mathbb{U}$ with elements U_{ij} can be decomposed into the rotations around the individual coordinate axes. Here, several possible solutions can be obtained. In case the element U_{31} is not equal to ± 1 the angles θ, ψ and ϕ are obtained following equations A.20 to A.22.

$$\begin{aligned}
\theta_1 &= -\operatorname{asin}(U_{31}) \\
\theta_2 &= \pi - \theta_1
\end{aligned} \tag{A.20}$$

$$\begin{aligned}
\psi_1 &= \operatorname{atan2}\left(\frac{U_{32}}{\cos(\theta_1)}, \frac{U_{33}}{\cos(\theta_1)}\right) \\
\psi_2 &= \operatorname{atan2}\left(\frac{U_{32}}{\cos(\theta_2)}, \frac{U_{33}}{\cos(\theta_2)}\right)
\end{aligned} \tag{A.21}$$

$$\begin{aligned}
\phi_1 &= \operatorname{atan2}\left(\frac{U_{21}}{\cos(\theta_1)}, \frac{U_{11}}{\cos(\theta_1)}\right) \\
\phi_2 &= \operatorname{atan2}\left(\frac{U_{21}}{\cos(\theta_2)}, \frac{U_{11}}{\cos(\theta_2)}\right)
\end{aligned} \tag{A.22}$$

In case the element U_{31} is equal to ± 1 the angle ϕ can be set to zero. To obtain the angles θ and ϕ equations A.23 are used in case the element U_{31} is equal to -1. Otherwise, the

equations A.24 are used.

$$\theta = \pi/2$$
$$\psi = \phi + \text{atan2}\,(U_{12}, U_{13}) \tag{A.23}$$

$$\theta = -\pi/2$$
$$\psi = -\phi + \text{atan2}\,(-U_{12}, -U_{13}) \tag{A.24}$$

A.4. Internal Coordinates

Calculation of Internal Coordinates

Internal coordinates were calculated following the standard reference of Wilson, Decius and Cross[366]. PESE is considering four types of internal coordinates: Bond lengths, bond angles, dihedral angles and out-of-plane angles as explained within the next paragraphs. In case two atoms are closer than 1.3 times the sum of their covalent radii they are considered as bonded. The bond vector $\mathbf{r}_{12}$ (Figure A.1) is then calculated based on the atomic coordinates. Here, the bond length r_{12} is the magnitude of the bond vector. If several fragments were obtained following this definition, the fragments are bonded to each other based on their shortest distance to each other in case a single molecule is requested. For the calculation of the angles it is advantageous to normalize the bond vector to obtain a unit vector $\mathbf{e}_{12}$.

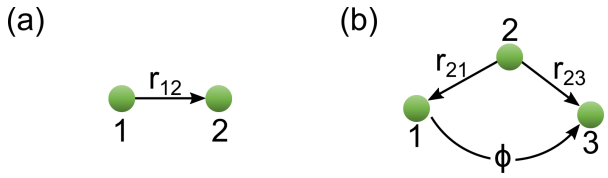


Figure A.1.: First set of internal coordinates. (a) Bond length r derived from the bond vector **r** and (b) bond angle ϕ.

Once the bond lengths are known, the bond angles ϕ are calculated between every pair of atoms bonded to the same third atom. In case unit vectors are used, the calculation of the bond angle (Figure A.1) simplifies to equation A.25.

$$\cos(\phi) = \mathbf{e}_{21} \cdot \mathbf{e}_{23} \tag{A.25}$$

In a similar fashion dihedral angles are calculated for every group of four atoms sequentially bonded as shown in Figure A.2. The dihedral angle φ is then calculated following equation A.26. However, the value obtained here is always in the range 0 to π.

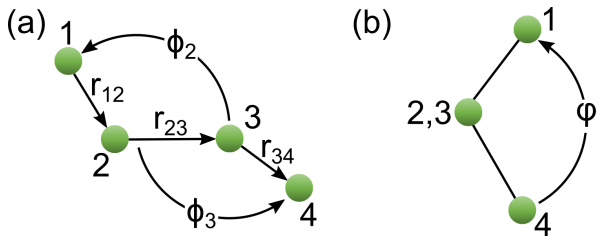


Figure A.2.: Second set of internal coordinates. Dihedral angle φ in (a) top view and (b) side view.

$$\cos(\varphi) = \frac{(\mathbf{e}_{12} \times \mathbf{e}_{23}) \cdot (\mathbf{e}_{23} \times \mathbf{e}_{34})}{\sin(\phi_2)\sin(\phi_3)} \tag{A.26}$$

Therefore, equation A.27 is considered. In case this equation is fulfilled the sign of the dihedral angel has to be negative. Consequently, the dihedral angles are obtained by definition in the range $-\pi$ to π.

$$\frac{\mathbf{e}_{12} \times \mathbf{e}_{23}}{\sin(\phi_2)} \cdot \mathbf{e}_{34} < 0 \tag{A.27}$$

The fourth internal coordinate used are the out-of-plane angles. These are in particular useful in case molecules with four atoms but without a dihedral angle like GaH_3, PH_3 or NH_3 are studied. A sketch of the out-of-plane angle θ is shown in Figure A.3.

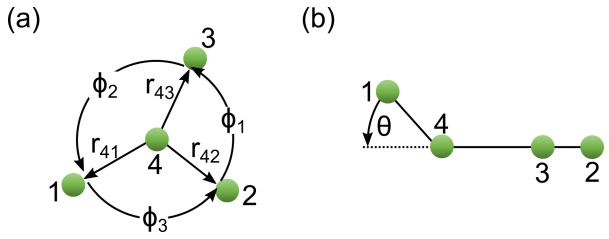


Figure A.3.: Third set of internal coordinates. Out of plane angle θ in (a) top view and (b) side view.

In PESE out-of-plane angles are calculated for every group of three atoms connected to the same fourth atom. The angles are derived following equation A.28.

$$\sin(\theta) = \frac{\mathbf{e}_{42} \times \mathbf{e}_{43}}{\sin(\phi_1)} \cdot \mathbf{e}_{41} \tag{A.28}$$

The Wilson B Matrix

For the transformation between internal and cartesian coordinates the Wilson B matrix $\mathbb{B}$ is needed as shown in equation A.29. Here, $\delta\mathbf{x}$ and $\delta\mathbf{q}$ represent a slight change of the cartesian

and internal coordinates, respectively.

$$\begin{aligned} \delta\mathbf{q} &= \mathbb{B} \cdot \delta\mathbf{x} \\ \delta\mathbf{x} &= \mathbb{B}^{-1} \cdot \delta\mathbf{q} \end{aligned} \tag{A.29}$$

Every element B_{ij} of the Wilson B matrix is a partial derivative of the internal coordinate q_i with respect to the cartesian coordinate x_j (equation A.30).

$$B_{ij} = \frac{\partial q_i}{\partial x_j} \tag{A.30}$$

For the calculation of these elements I again follow the definition of Wilson, Decius and Cross[366]. Here, the elements B_{ij} are not derived component wise but atom wise. This means that instead of calculating the elements B_{i,x_1}, B_{i,y_1} and B_{i,z_1} step by step the vector $\mathbf{s}_{i1} = (B_{i,x_1} B_{i,y_1} B_{i,z_1})$ giving the relation of atom 1 on the coordinate q_i is derived. Following the sketches of Figure A.1 to A.3 the elements of the Wilson B matrix are derived by the equations A.31 (bond length), A.32 (bond angle), A.33 (dihedral angle) and A.34 (out-of-plane angle).

$$\begin{aligned} \mathbf{s}_{i1} &= \mathbf{e}_{21} \\ \mathbf{s}_{i2} &= \mathbf{e}_{12} \end{aligned} \tag{A.31}$$

$$\begin{aligned} \mathbf{s}_{i1} &= \frac{\cos(\phi)\mathbf{e}_{21} - \mathbf{e}_{23}}{r_{21}\sin(\phi)} \\ \mathbf{s}_{i2} &= \frac{[(r_{21} - r_{23}\cos(\phi))\mathbf{e}_{21} + (r_{23} - r_{21}\cos(\phi))\mathbf{e}_{23}]}{r_{21} r_{23}\sin(\phi)} \\ \mathbf{s}_{i3} &= \frac{\cos(\phi)\mathbf{e}_{23} - \mathbf{e}_{21}}{r_{23}\sin(\phi)} \end{aligned} \tag{A.32}$$

$$
\begin{aligned}
\mathbf{s}_{i1} &= -\frac{\mathbf{e}_{12} \times \mathbf{e}_{23}}{r_{12} \sin^2(\phi_2)} \\
\mathbf{s}_{i2} &= \frac{r_{23} - r_{12}\cos(\phi_2)}{r_{23} r_{12} \sin(\phi_2)} \frac{\mathbf{e}_{12} \times \mathbf{e}_{23}}{\sin(\phi_2)} - \frac{\cos(\phi_3)}{r_{23} \sin(\phi_3)} \frac{\mathbf{e}_{43} \times \mathbf{e}_{32}}{\sin(\phi_3)} \\
\mathbf{s}_{i3} &= \frac{r_{32} - r_{43}\cos(\phi_3)}{r_{32} r_{43} \sin(\phi_3)} \frac{\mathbf{e}_{43} \times \mathbf{e}_{32}}{\sin(\phi_3)} - \frac{\cos(\phi_2)}{r_{32} \sin(\phi_2)} \frac{\mathbf{e}_{12} \times \mathbf{e}_{23}}{\sin(\phi_2)} \\
\mathbf{s}_{i4} &= -\frac{\mathbf{e}_{43} \times \mathbf{e}_{32}}{r_{43} \sin^2(\phi_3)}
\end{aligned}
\tag{A.33}
$$

$$
\begin{aligned}
\mathbf{s}_{i1} &= \frac{1}{r_{41}} \left(\frac{\mathbf{e}_{42} \times \mathbf{e}_{43}}{\cos(\theta) \sin(\phi_1)} - \tan(\theta) \mathbf{e}_{41} \right) \\
\mathbf{s}_{i2} &= \frac{1}{r_{42}} \left[\frac{\mathbf{e}_{43} \times \mathbf{e}_{41}}{\cos(\theta) \sin(\phi_1)} - \frac{\tan(\theta)}{\sin^2(\phi_1)} \left(\mathbf{e}_{42} - \cos(\phi_1) \mathbf{e}_{43} \right) \right] \\
\mathbf{s}_{i3} &= \frac{1}{r_{43}} \left[\frac{\mathbf{e}_{41} \times \mathbf{e}_{42}}{\cos(\theta) \sin(\phi_1)} - \frac{\tan(\theta)}{\sin^2(\phi_1)} \left(\mathbf{e}_{43} - \cos(\phi_1) \mathbf{e}_{42} \right) \right] \\
\mathbf{s}_{i4} &= -\mathbf{s}_{i1} - \mathbf{s}_{i2} - \mathbf{s}_{i3}
\end{aligned}
\tag{A.34}
$$

B

Algorythm

B.1. Hungarian Algorithm

The Hungarian algorithm also known as Kuhn-Munkres algorithm[183,184] can be used to solve the assignment problem. Here, I am using the Hungarian algorithm to search the assignment of atoms minimizing the RMSD between two structures. For this the coordinates $\mathbf{R}_A$ of atom A in the first structure and the coordinates $\mathbf{R}_B$ of atom B in the second structure are used to define the elements of the cost matrix as shown in equation B.1.

$$C_{AB} = \|\mathbf{R}_A - \mathbf{R}_B\|^2 \tag{B.1}$$

By manipulating the cost matrix $\mathbb{C}$ the Hungarian algorithm finds the assignment between atoms A and B, which minimizes the sum over the selected entries of $\mathbb{C}$ and thereby also the RMSD. In the following, the algorithm is explained step-by-step using the example of Figure 3.6. On the right side the step of the algorithm is explained while the outcome is shown on the left side.

6	3	1
1	4	7
1	2	5

Step 0
Arbitrary cost matrix obtained by equation B.1.

5	2	0
0	3	6
0	1	4

Step 1
In this step the smallest element of every row is subtracted from all elements in the respective row. Then the algorithm proceeds with step 2.

5	2	0*
0*	3	6
0	1	4

Step 2

In the second step the algorithm searches for every element equal to zero. In case no starred zero is in the same row or column as the selected element, this element is starred. Here, different outcomes are possible based on the search behavior (column-wise vs. row-wise) of the algorithm. However, the overall outcome of the Hungarian algorithm is not influenced by this. The algorithm proceeds with step 3.

5	2	0*
0*	3	6
0	1	4

Step 3

In step 3 every column with a starred zero is covered. In case every column is covered an optimal assignment is found. Otherwise, the algorithm continues with step 4.

5	2	0*
0*	3	6
0	1	4

Step 4

This step depends on the uncovered elements. In case an uncovered zero is present this element is primed. Furthermore, in case no starred zero is in the same row as the primed zero the algorithm proceeds with step 5. In case a starred zero is in the same row as the primed zero, this row is covered and the column containing the starred zero is uncovered. These steps are repeated until no uncovered zero is left. Then the smallest uncovered value is stored and the algorithm proceeds with step 6.

5	1	0*
0*	2	6
0	0	4

Step 6

The value obtained in step 4 is added to every element of a covered row and subtracted from every element of an uncovered column. In this way doubly covered elements are increased while uncovered elements are decreased. After this the algorithm returns to step 4.

5	1	0*
0*	2	6
0	0'	4

Step 4

For a description see previous steps.

5	1	0*
0*	2	6
0	0*	4

Step 5

In this step the set of primed and starred zeros is turned into a set of only starred zeros. For this a sequence of primed and starred zeros is generated. The first element of this sequence is the uncovered primed zero of step 4. The next element is a starred zero in the same column as the primed zero. In case no starred zero is present in the same column the sequence is completed. Otherwise, the next element is a primed zero in the same row as the starred zero of the previous step. These steps are repeated until the sequence is completed. Then all starred zeros of the sequence are unstarred and all primed zeros of the sequence are starred. Also, all other primes are removed and all lines are uncovered. After this the algorithm jumps back to step 3.

5	1	0*
0*	2	6
0	0*	4

Step 3

For a description see previous steps.

6	3	1
1	4	7
1	2	5

Algorithm converged

Since all columns are covered a final assignment was found. The individual pairs are identified by the starred zeros. The cost of the assignment can be obtained with the initial cost matrix.

My Fortran implementations follows the C# code[367] written by Robert A. Pilgrim.

B.2. Kabsch Algorithm

The Kabsch algorithm[368,369] is used to find the best solution for a rotation matrix mapping two sets of vectors. In our case these sets of vectors are the atomic coordinates of two structures $\mathbb{X}$ and $\mathbb{Y}$ stored in $3 \times M$ matrices with M the number of atoms. Before the algorithm is used both structures $\mathbb{X}$ and $\mathbb{Y}$ are moved to their centroid. Then the algorithm proceeds by calculating the matrix $\mathbb{C}$ as shown in equation B.2.

$$\mathbb{C} = \mathbb{X}\mathbb{Y}^T \tag{B.2}$$

To calculate the best rotation matrix $\mathbb{U}$ the inverse of $\mathbb{C}$ could be used. However, this matrix is not defined in all cases. Therefore, $\mathbb{C}$ is decomposed by using a SVD routine (equation B.3).

$$\mathbb{C} = \mathbb{V}\$\mathbb{W}^T \tag{B.3}$$

Also, the sign of the determinant of $\mathbb{C}$ is used to define a matrix $\mathbb{T}$. In equation B.4 the constant c is set to 1 in case the determinant of $\mathbb{C}$ has a positive sign otherwise c is set to -1.

$$\mathbb{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix} \tag{B.4}$$

With the singular vectors of $\mathbb{C}$ and $\mathbb{T}$ the rotation $\mathbb{U}$ can finally be obtained by equation B.5.

$$\mathbb{U} = \mathbb{W}\mathbb{T}\mathbb{V}^T \tag{B.5}$$

B.3. Murty Algorithm

With the Hungarian algorithm the best pairing of atoms can be found. However, in some cases also the second, third or n-th best pairing is of interest. In such a case the Murty algorithm[370] can be used. The basic idea of this algorithm is to start with the best pairing obtained by for example the Hungarian algorithm and gradually worsen it.

$$a(1) = [(1,9),(2,7),(3,3),(4,8),(5,6),(6,4),(7,10),(8,1),(9,5),(10,2)] \tag{B.6}$$

In equation B.6 an example[370] for a best pairing is shown. Here, $(1,9)$ indicates that the element 1 and 9 are paired. The second best pairing has to differ from this pairing by at least two elements, which correspond to one permutation. This permutation is forced by excluding one pairing of $a(1)$. The Murty algorithm formulates all possible ways to exclude one pairing in a smart way as shown in equation B.7: Either the pairing $\overline{(1,9)}$ is not included in the second best pairing or it is included $(1,9)$. In case, the pairing $(1,9)$ is included the pair $\overline{(2,7)}$ is not included and so forth. It is important to recognize that the so called nodes M_1 to M_9 describe all combinatorial possible pairings except the best pairing $a(1)$. Consequently, solving every node by the Hungarian algorithm results in the second best pairing. In practice, this is achieved by manipulating the original cost matrix individually for every node. Every element that is not included is set to infinity or at least to a huge value. For every element

that must be included all other elements of the same row or column are set to infinity.

$$
\begin{aligned}
M_1 &= \left[\overline{(1,9)}\right] \\
M_2 &= \left[(1,9), \overline{(2,7)}\right] \\
M_3 &= \left[(1,9), (2,7), \overline{(3,3)}\right] \\
M_4 &= \left[(1,9), (2,7), (3,3), \overline{(4,8)}\right] \\
M_5 &= \left[(1,9), (2,7), (3,3), (4,8), \overline{(5,6)}\right] \\
M_6 &= \left[(1,9), (2,7), (3,3), (4,8), (5,6), \overline{(6,4)}\right] \\
M_7 &= \left[(1,9), (2,7), (3,3), (4,8), (5,6), (6,4), \overline{(7,10)}\right] \\
M_8 &= \left[(1,9), (2,7), (3,3), (4,8), (5,6), (6,4), (7,10), \overline{(8,1)}\right] \\
M_9 &= \left[(1,9), (2,7), (3,3), (4,8), (5,6), (6,4), (7,10), (8,1), \overline{(9,5)}\right]
\end{aligned}
\tag{B.7}
$$

In the same fashion the n-th best pairing can be obtained by taking the (n-1) best pairing and partitioning the node of this pairing. Here, it is important that all restrictions are conserved. To clarify this we assume that M_4 resulted in the second best pairing $a(2)$ shown in equation B.8.

$$
a(2) = [(1,9), (2,7), (3,3), (4,2), (5,6), (6,4), (7,8), (8,1), (9,5), (10,10)] \tag{B.8}
$$

To obtain the third best pairing the node M_4 is partitioned by $a(2)$ resulting in the additional nodes M_{10} to M_{15} shown in equation B.9. As shown all constraints from the original node M_4 are conserved. Within the complete set of nodes the third best pairing can be found since it

describes all combinatorial possible pairings except the best and second best pairing.

$$
\begin{aligned}
M_{10} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},\overline{(4,2)}\right] \\
M_{11} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},(4,2),\overline{(5,6)}\right] \\
M_{12} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},(4,2),(5,6),\overline{(6,4)}\right] \\
M_{13} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},(4,2),(5,6),(6,4),\overline{(7,8)}\right] \\
M_{14} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},(4,2),(5,6),(6,4),(7,8),\overline{(8,1)}\right] \\
M_{15} &= \left[(1,9),(2,7),(3,3),\overline{(4,8)},(4,2),(5,6),(6,4),(7,8),(8,1),\overline{(9,5)}\right]
\end{aligned}
\tag{B.9}
$$

B.4. Internal Coordinate Transformation

While internal coordinates can be directly derived from the cartesian coordinates as shown in appendix A.4, the transformation from internal to cartesian coordinates is much more complex. Following Peng, Ayala, Schlegel and Frisch[194], the transformation of a small change in internal coordinates $\Delta\mathbf{q}$ to cartesian coordinates is obtained by equation B.10. Here, a change of internal coordinates $\Delta\mathbf{q}$ is transformed by the Wilson B matrix $\mathbb{B}$ and added to the old cartesian coordinates $\mathbf{x}_0$ to obtain the new set of cartesian coordinates $\mathbf{x}_1$. $\mathbb{G}^{-1}$ represents the inverse of $\mathbb{G} = \mathbb{B}\mathbb{B}^T$ and is thereby also derived from the Wilson B matrix.

$$
\mathbf{x}_1 = \mathbf{x}_0 + \mathbb{B}^T\mathbb{G}^{-1}\Delta\mathbf{q} \tag{B.10}
$$

Instead of strictly following equation B.10, PESE is using the pseudo inverse $\mathbb{B}^{-1}$ of the Wilson B matrix. This matrix can easily be obtained by a SVD as shown in appendix A.1. As shown in equation B.11 both approaches are equivalent.

$$
\begin{aligned}
\mathbf{x}_1 &= \mathbf{x}_0 + \mathbb{B}^T\mathbb{G}^{-1}\Delta\mathbf{q} \\
&= \mathbf{x}_0 + \mathbb{B}^T(\mathbb{B}\mathbb{B}^T)^{-1}\Delta\mathbf{q} \\
&= \mathbf{x}_0 + \underbrace{(\mathbb{B}^T\mathbb{B}^{-T})}_{=\mathbb{1}}\mathbb{B}^{-1}\Delta\mathbf{q} \\
&= \mathbf{x}_0 + \mathbb{B}^{-1}\Delta\mathbf{q}
\end{aligned}
\tag{B.11}
$$

The transformation of internal to cartesian coordinates following equation B.11 is an iterative transformation since cartesian and internal coordinates are not linearly related. Therefore, based on $\mathbf{x}_1$ a new set of internal coordinates $\mathbf{q}_1$ is derived and the difference is compared to $\Delta\mathbf{q}$. The remaining difference $\Delta\Delta\mathbf{q} = \Delta\mathbf{q} - (\mathbf{q}_1 - \mathbf{q}_0)$ is again transformed by equation B.11. This step is repeated until the change in the cartesian coordinates is smaller than 0.001 Å. In addition, a maximum number of iterations has to be selected to terminate the iterative transformation in case a convergence is not obtained.

For PESE a special problem arises since reaction paths are interpolated. Here, the change in internal coordinates can be large. However, the introduced approach converges only for minor changes in internal coordinates. I circumvent this problem by adding a substantial change in the internal coordinates in small portions to the cartesian coordinates. Therefore, a second iterative loop is implemented. Here, the currently applied change in internal coordinates $\Delta\mathbf{q}$ is halved every time the transformation of equation B.11 is not converging. In case a portion was successfully added to the cartesian coordinates the next portion is transformed. With this second iterative loop only problematic transformations are split in small portions, while simpler transformations are done in a single step. In PESE the maximum number of portions is set to 1000. This means that a change of an angle by 100° could be split in portions of 0.1° if necessary. However, convergence is not guaranteed for every reaction path and an interpolation in cartesian coordinates must be considered as fallback procedure.

C

Reactivity of Atoms and Small Molecules on Gallium Phosphide

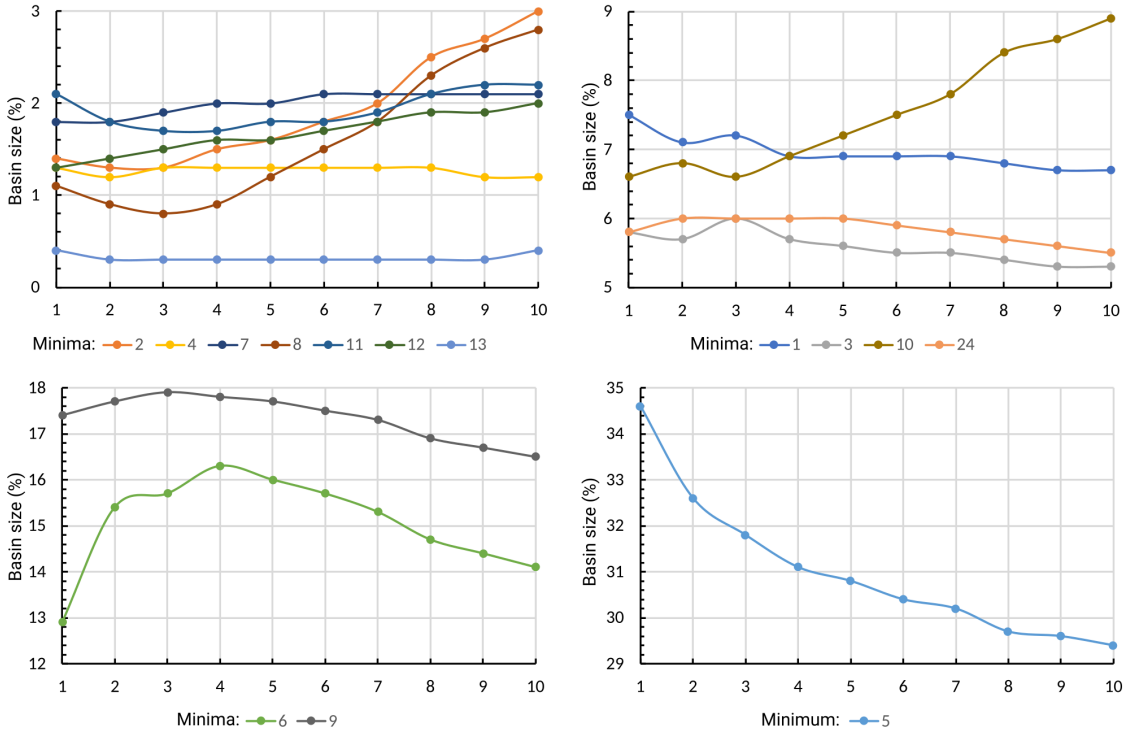


Figure C.1.: Dependence of the relative basin sizes on the smearing keyword ω shown for Bi on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. The extend of the smearing in terms of final grid points n is used as x-axis.

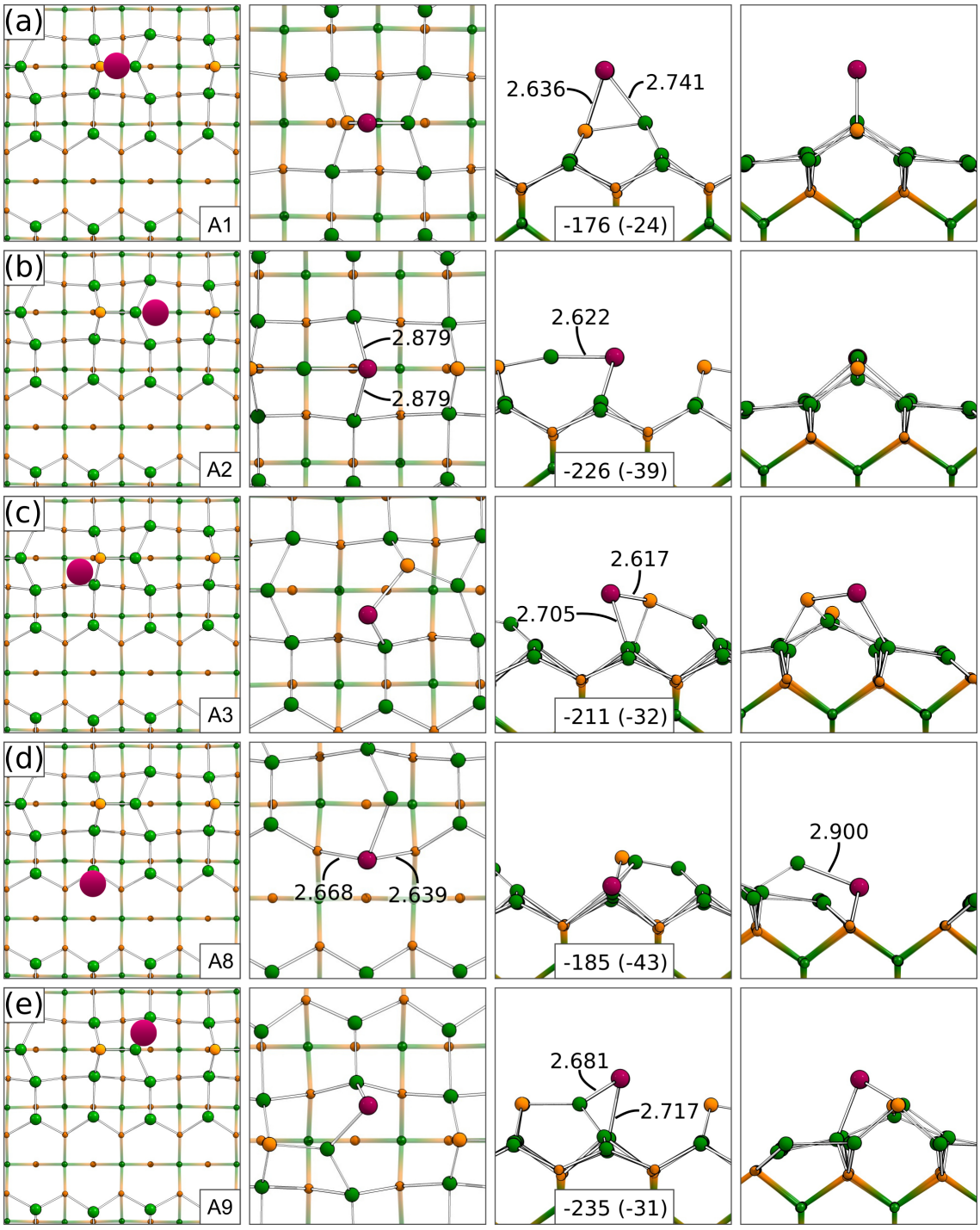


Figure C.2.: Top and side view on adsorption structures of Bi on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima (”A”) follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), Bi (purple).

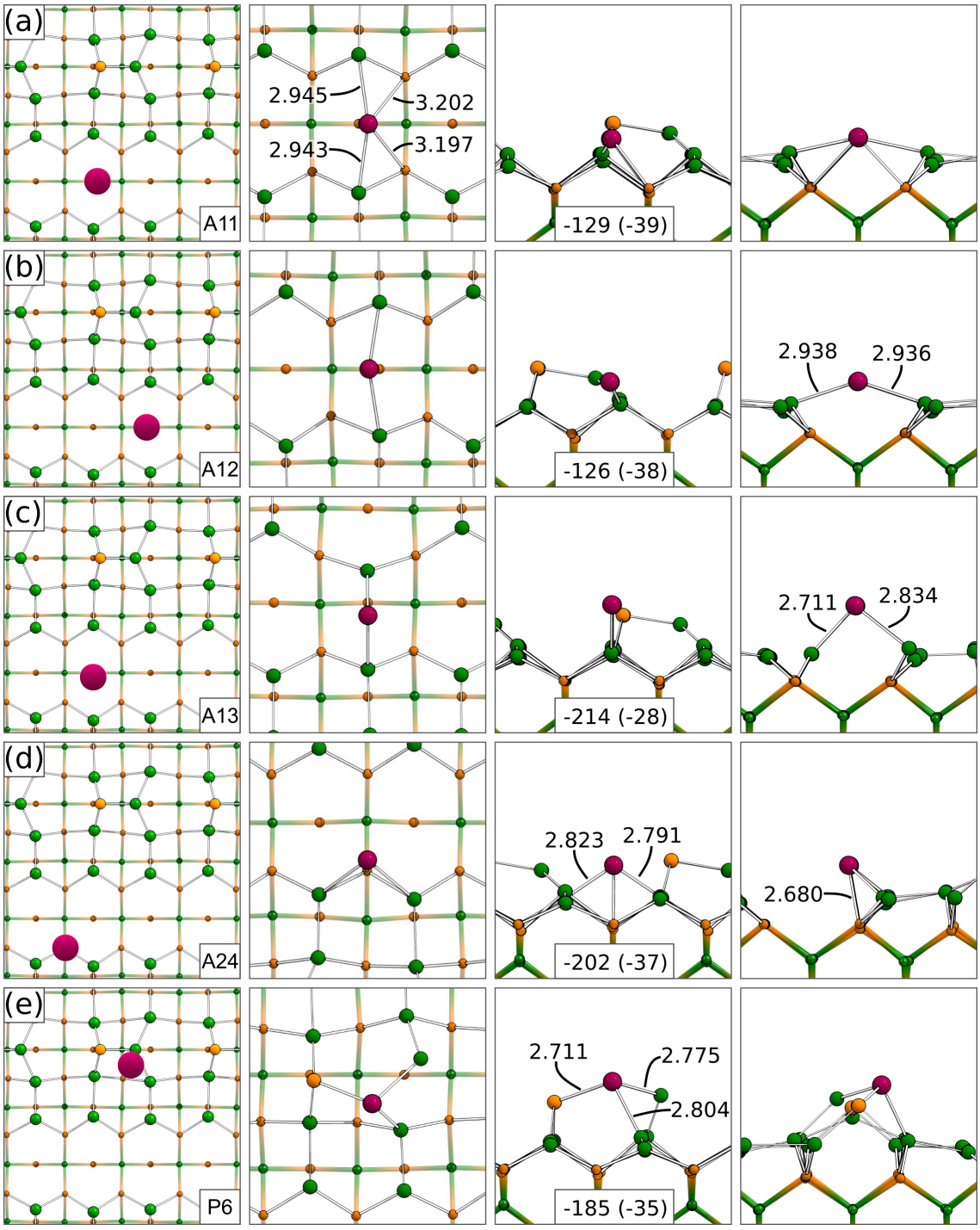


Figure C.3.: Top and side view on adsorption structures of Bi on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima (”A”) follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), Bi (purple).

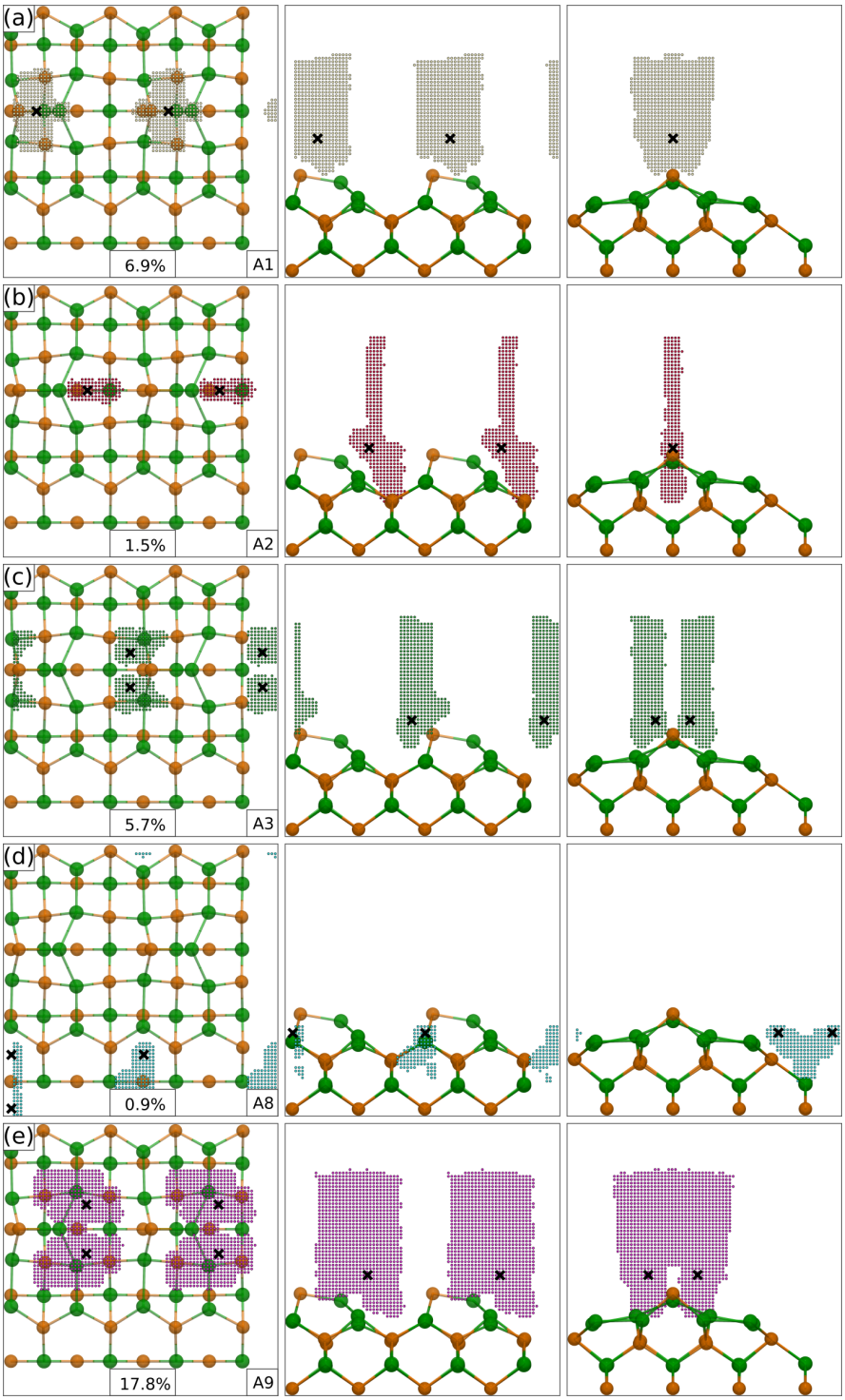


Figure C.4.: Top and side view on adsorption basins of Bi on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange).

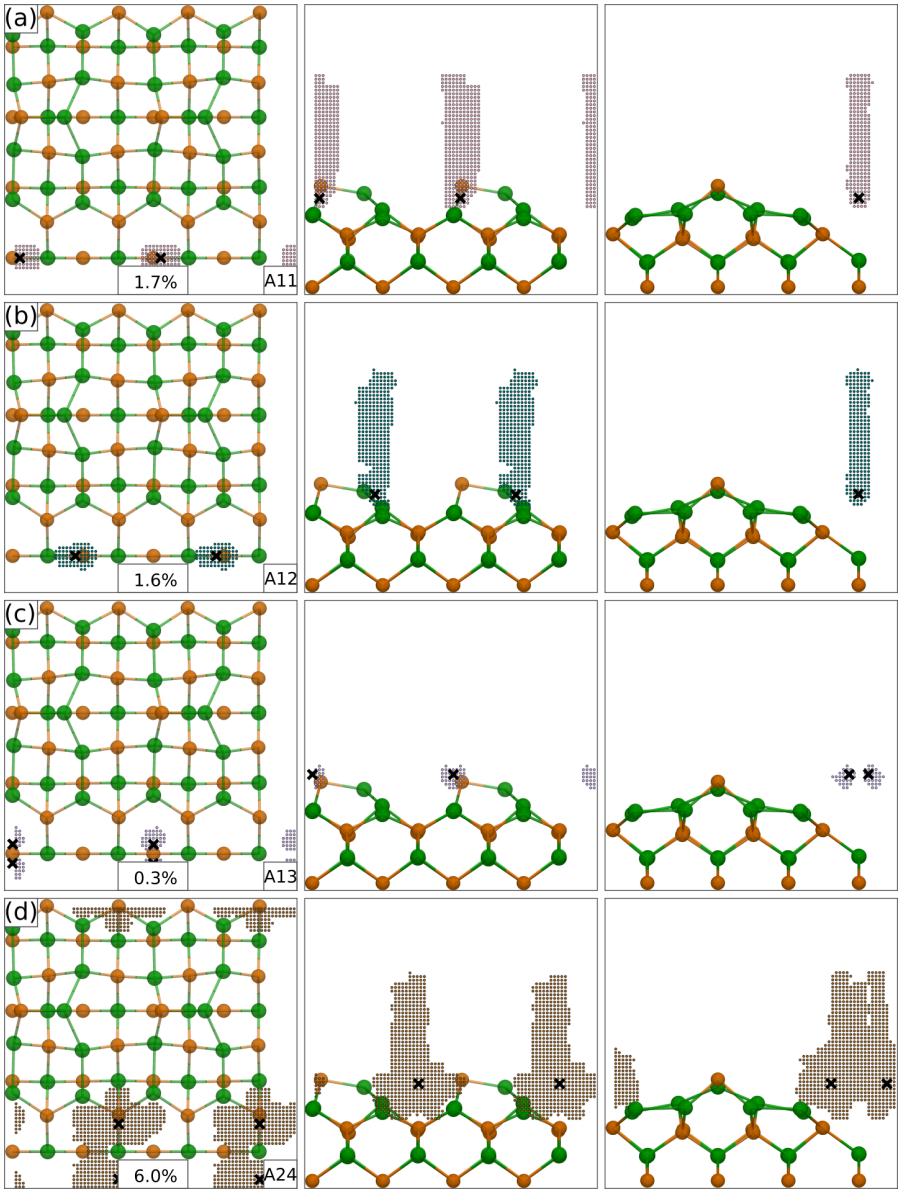


Figure C.5.: Top and side view on adsorption basins of Bi on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange).

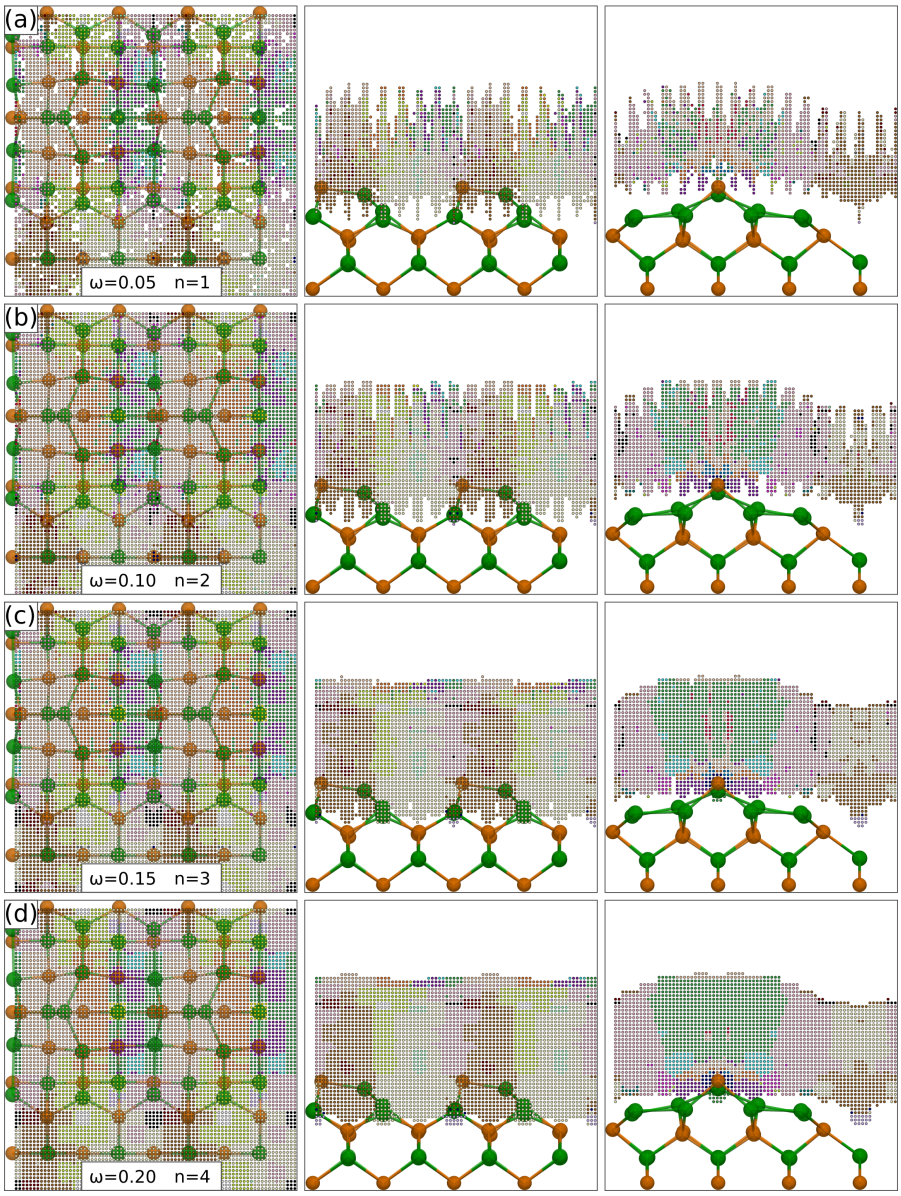


Figure C.6.: Influence of the smearing keyword ω on the final adsorption basin mapping for GaH_3 on GaP(001). Every final grid point, which is assigned by more than $1\,\%$ to a minimum, is shown as a small point in the color of the minimum. For ω settings of (a) $\omega = 0.05$, (b) $\omega = 0.10$, (c) $\omega = 0.15$ and (d) $\omega = 0.20$ are shown. In addition, the range of the smearing is stated in terms of final grid points n. Color code: Ga (green), P (orange).

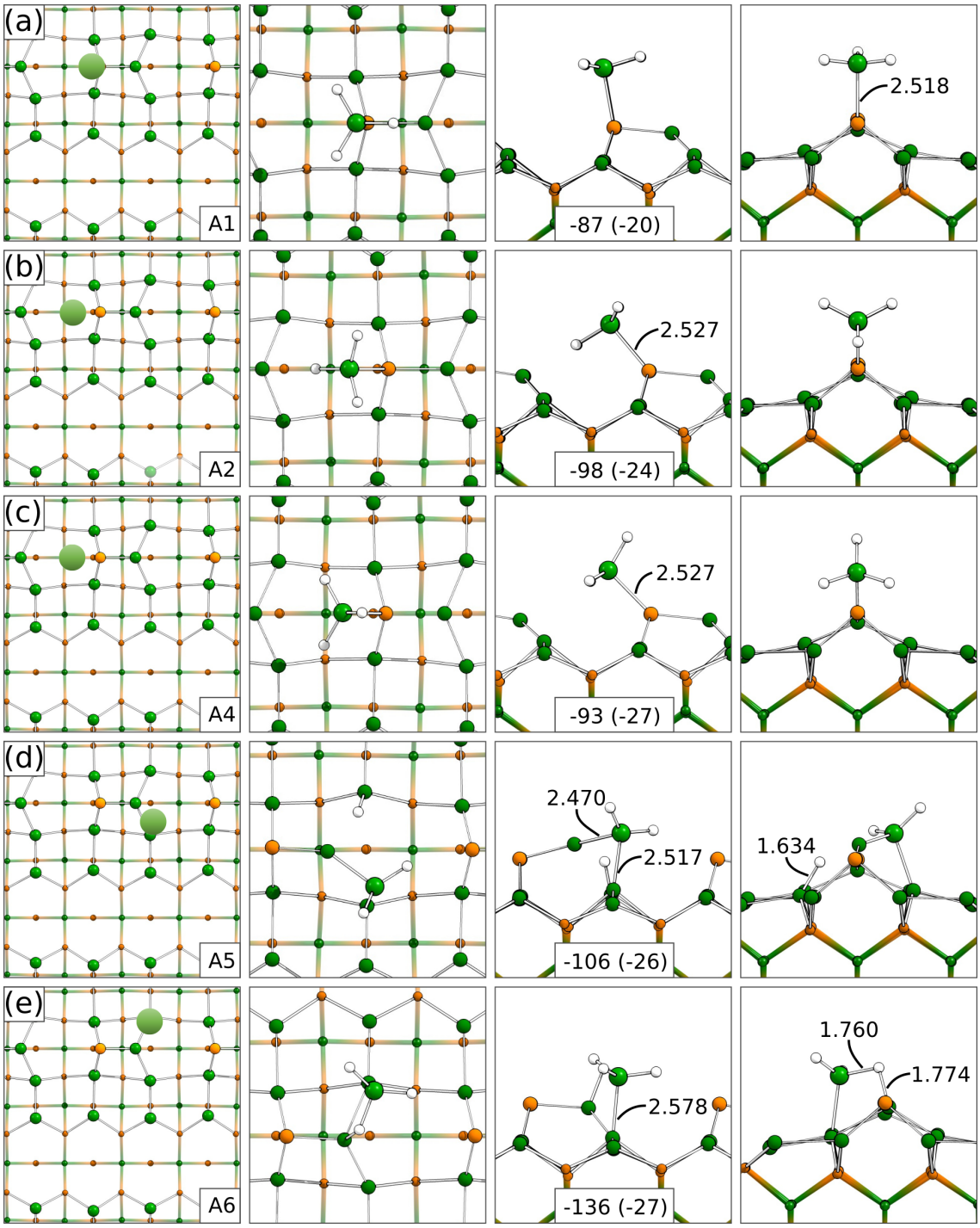


Figure C.7.: Top and side view on adsorption structures of GaH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima (”A”) follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

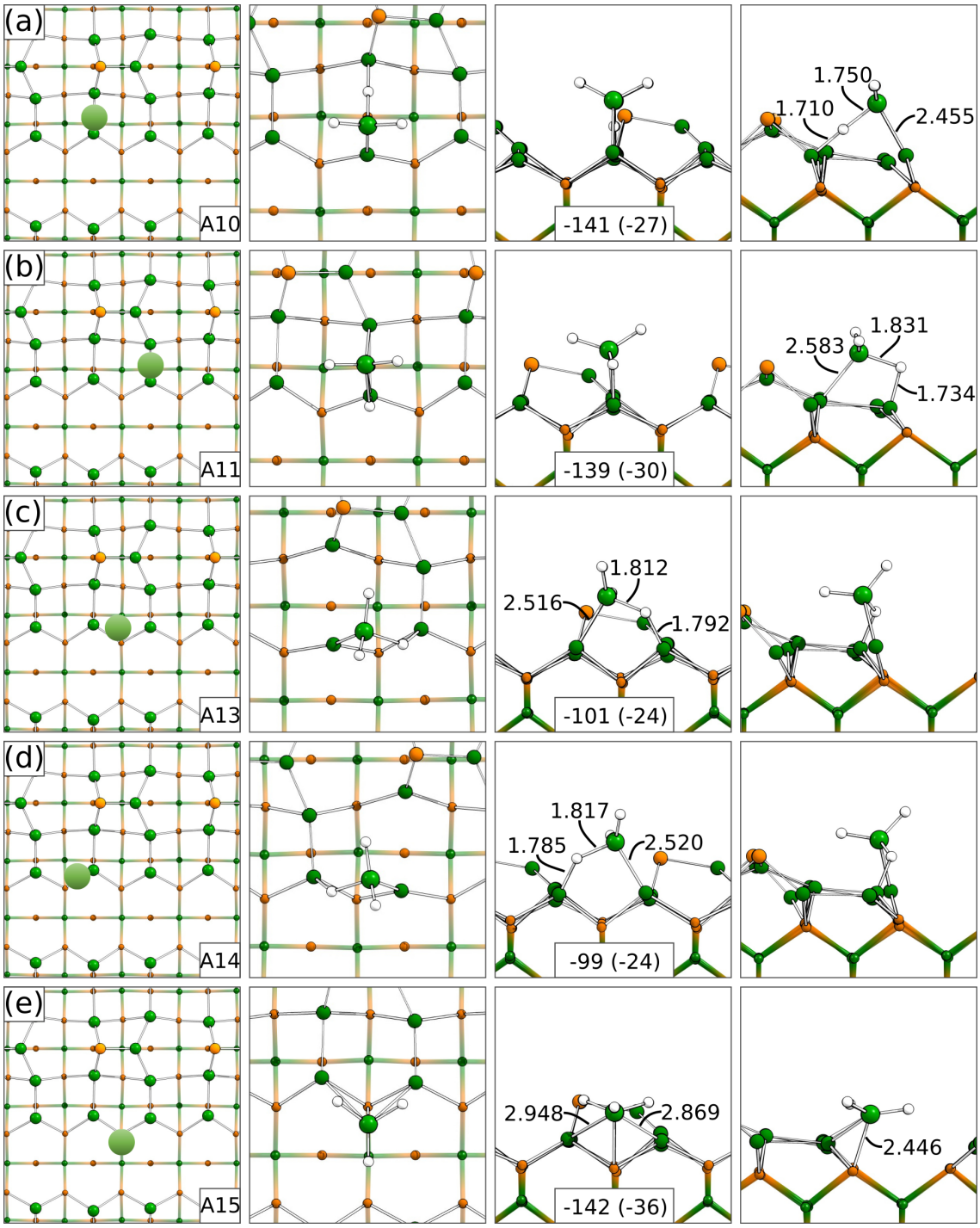


Figure C.8.: Top and side view on adsorption structures of GaH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in $kJ \cdot mol^{-1}$ with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima ("A") follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

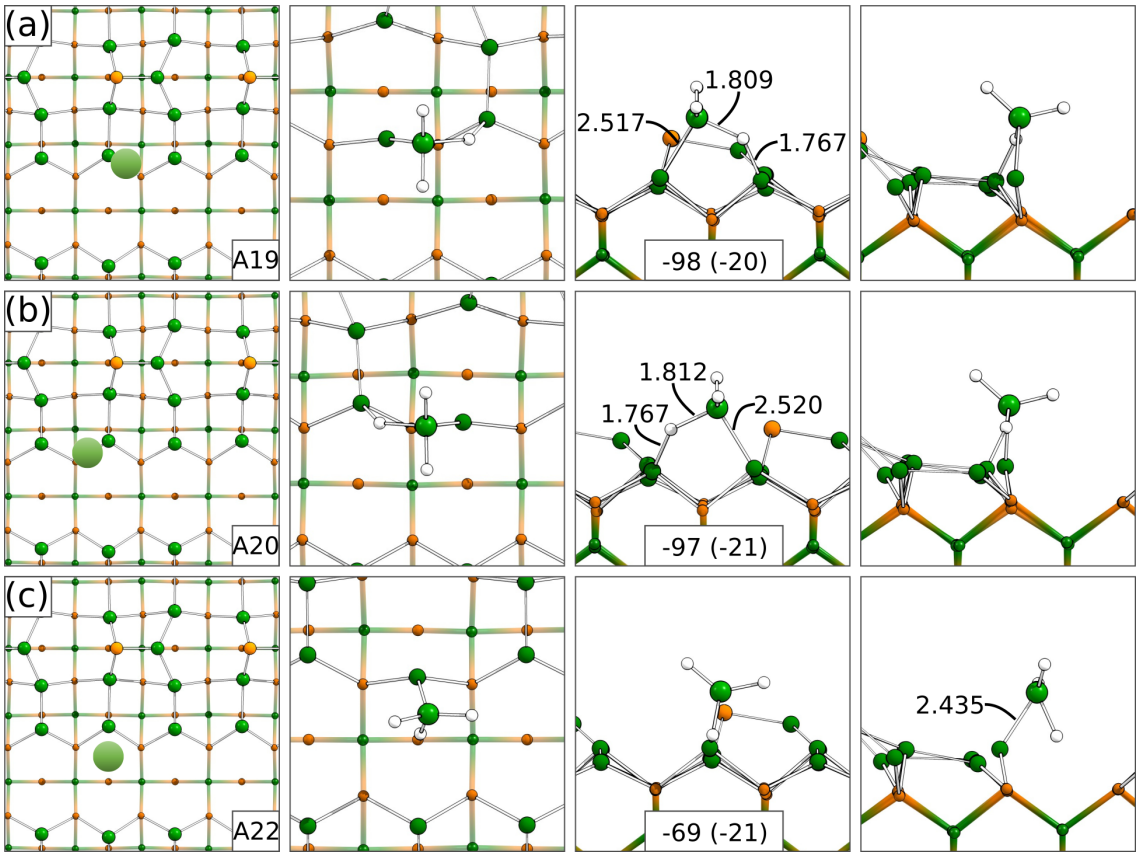


Figure C.9.: Top and side view on adsorption structures of GaH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima ("A") follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

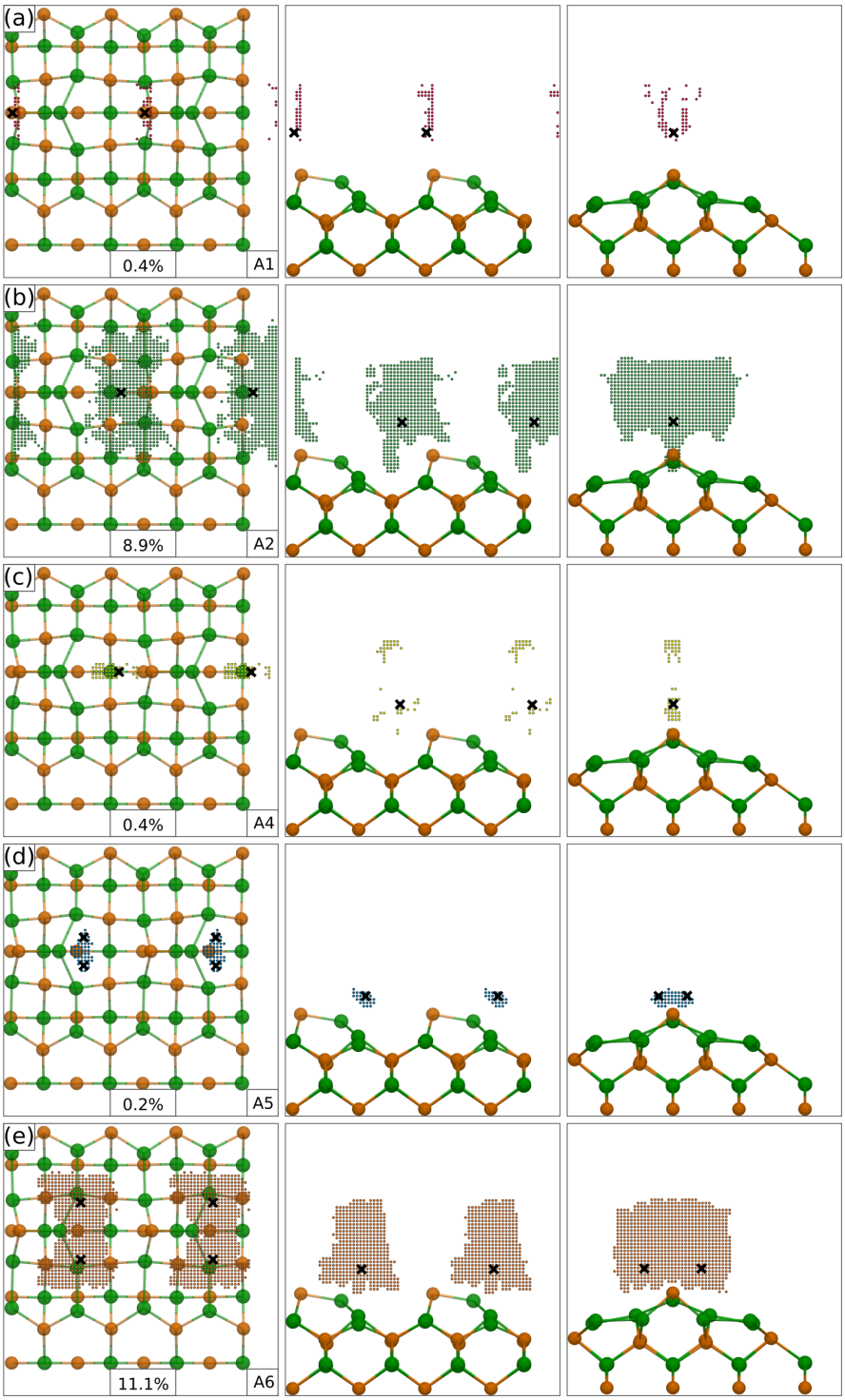


Figure C.10.: Top and side view on adsorption basins of GaH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

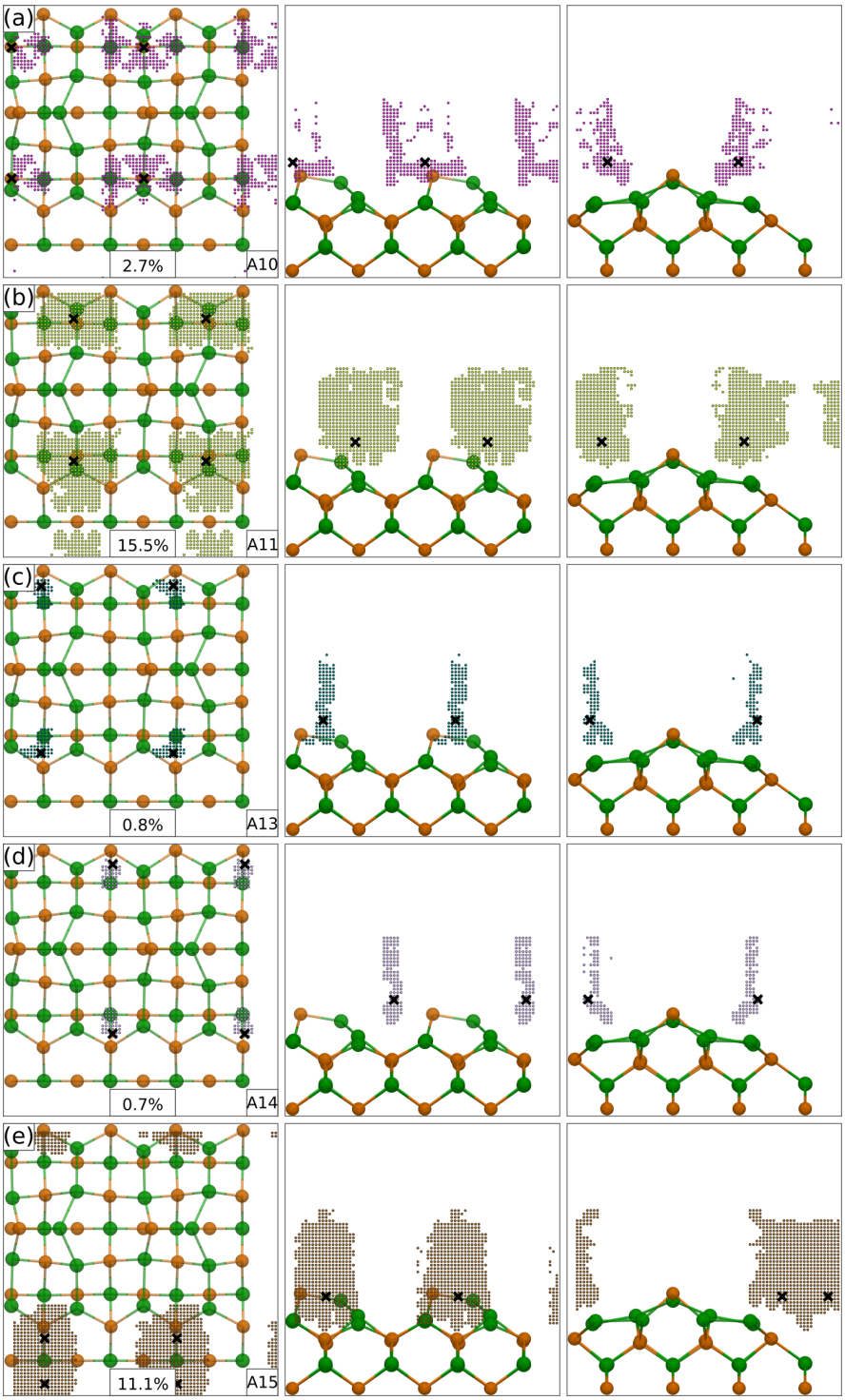


Figure C.11.: Top and side view on adsorption basins of GaH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

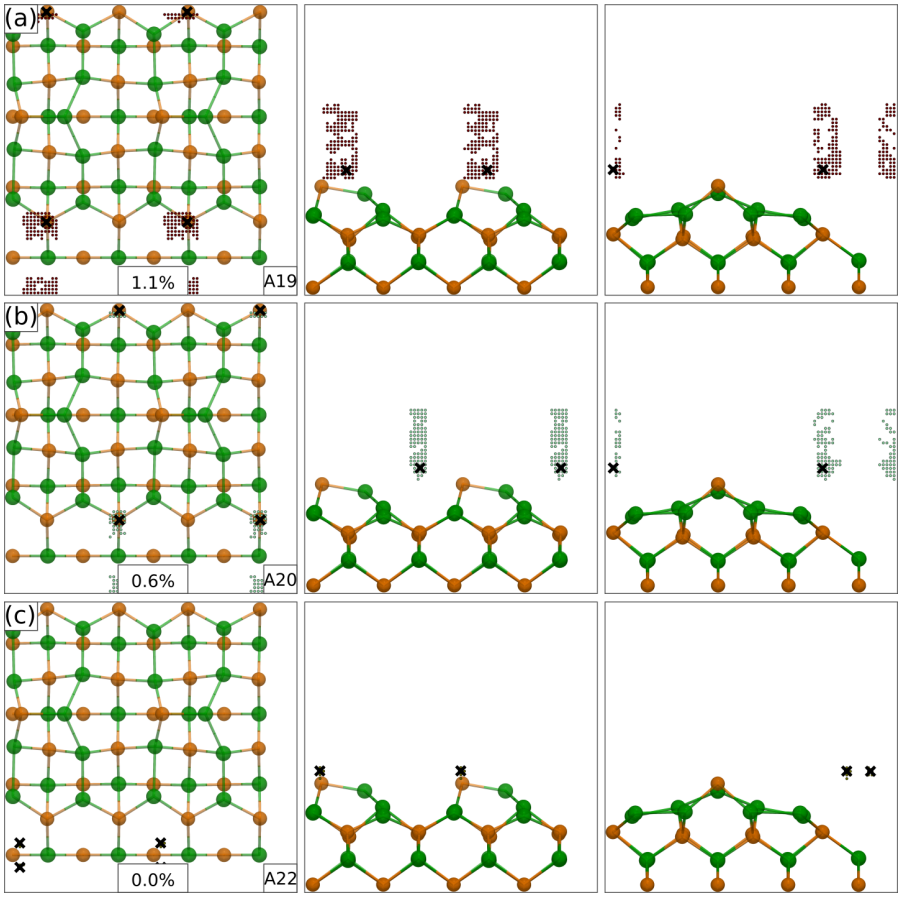


Figure C.12.: Top and side view on adsorption basins of GaH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

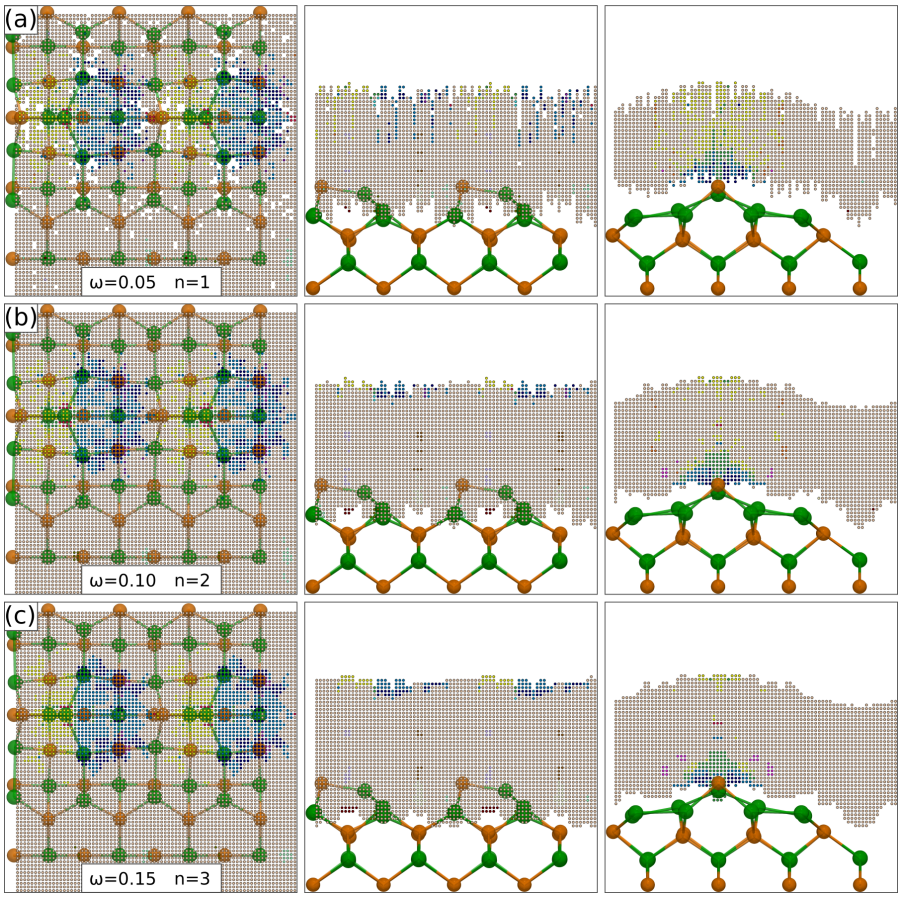


Figure C.13.: Influence of the smearing keyword ω on the final adsorption basin mapping for PH_3 on GaP(001). Every final grid point, which is assigned by more than 1 % to a minimum, is shown as a small point in the color of the minimum. For ω settings of (a) $\omega = 0.05$, (b) $\omega = 0.10$ and (c) $\omega = 0.15$ are shown. In addition, the range of the smearing is stated in terms of final grid points n. Color code: Ga (green), P (orange).

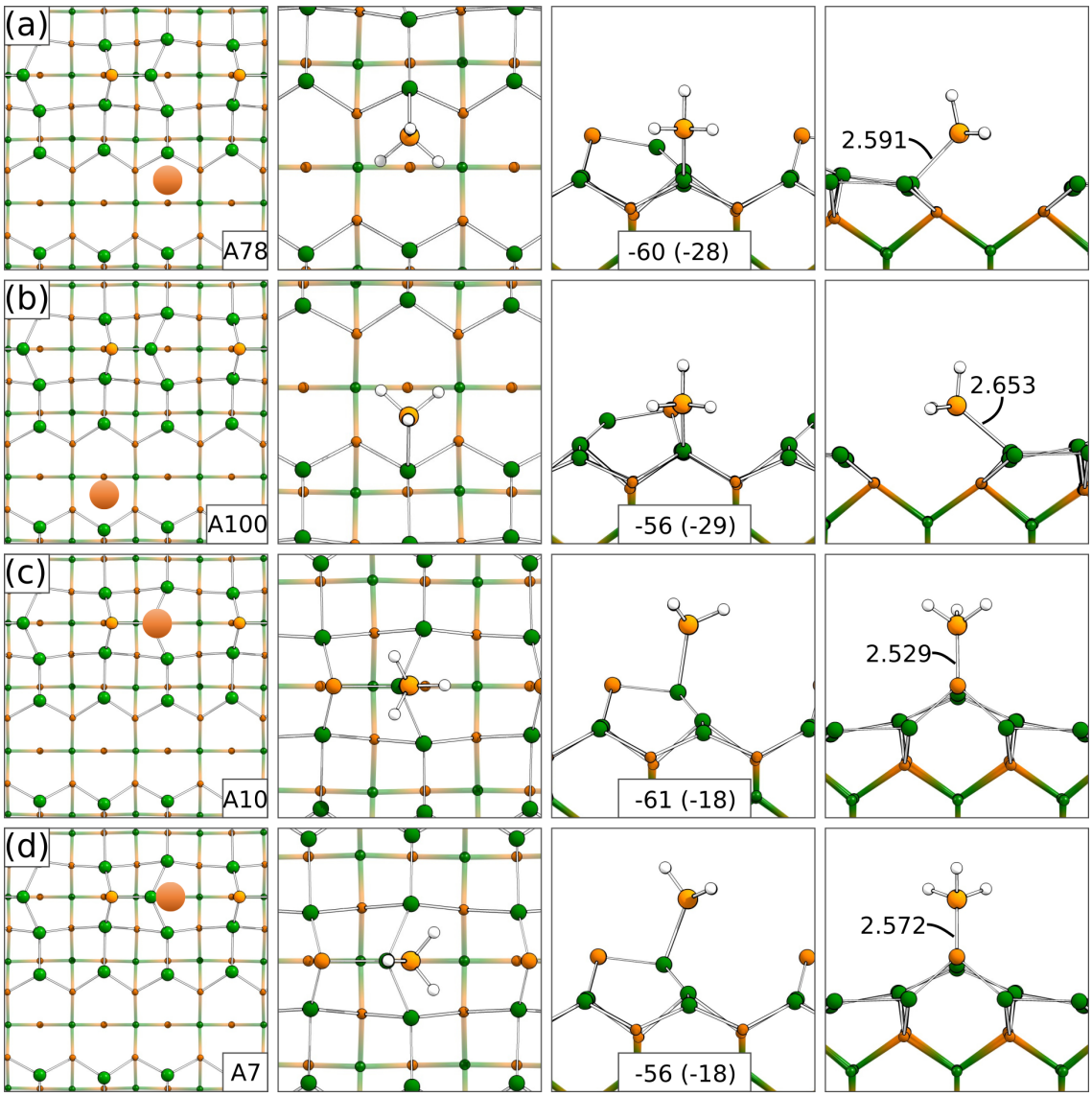


Figure C.14.: Top and side view on adsorption structures of PH_3 on GaP(001). In the first column the relative position of the adsorbate is sketched. Adsorption energies in kJ·mol^{-1} with their dispersion contribution in parentheses. Bond lengths in Å. The numeration of the adsorption minima ("A") follows the internal nomenclature of PESE. Color code: Ga (green), P (orange), H (white).

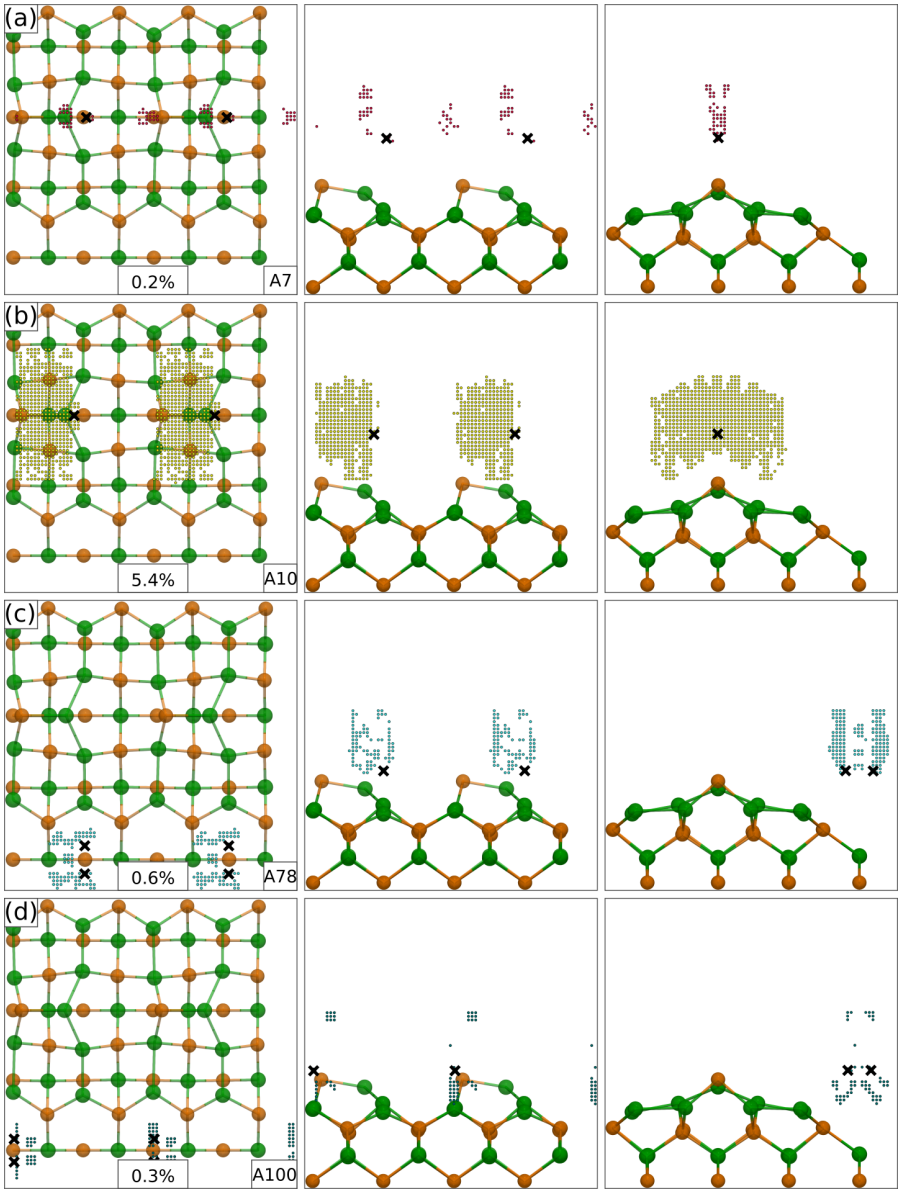


Figure C.15.: Top and side view on adsorption basins of PH_3 on GaP(001). The relative size of every basin to the cumulated size of all basins is stated in %. Final grid points (small points), which are predominantly and by more than $1\,\%$ assigned to a minimum, are colored according to the minimum (large point). Color code: Ga (green), P (orange), H (white).

D

Small Molecule Inhibitors for the Area-Selective Atomic Layer Deposition of Aluminium Oxide and Hafnium Oxide

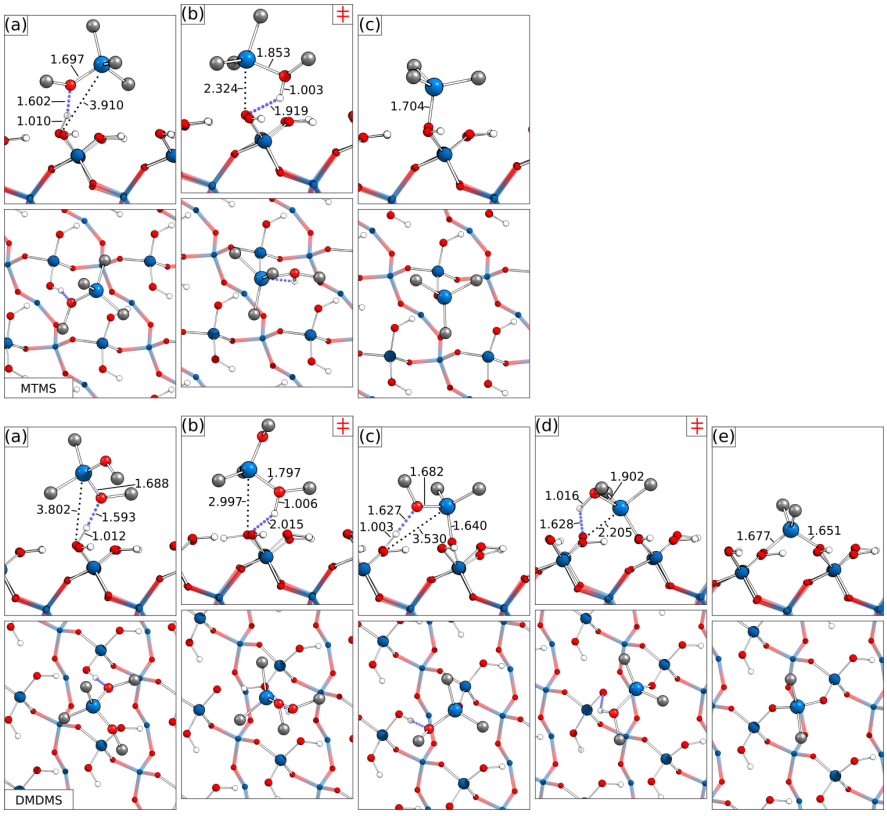


Figure D.1.: Top and side view on MTMS (first row) and DMDMS (second row). Structures (b) and (d) are transition states (‡). Bond lengths in Å. Color code: Si (blue), O (red), H (white), C (grey). Reproduced with permission from reference [341].

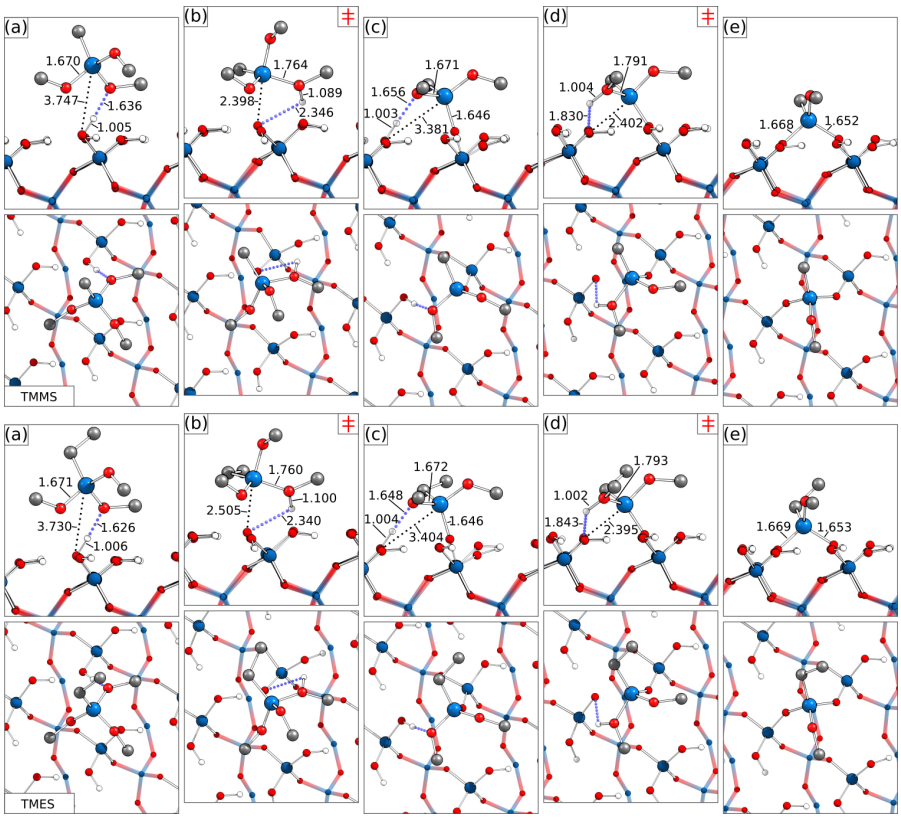


Figure D.2.: Top and side view on TMMS (first row) and TMES (second row). Structures (b) and (d) are transition states (‡). Bond lengths in Å. Color code: Si (blue), O (red), H (white), C (grey). Reproduced with permission from reference [341].

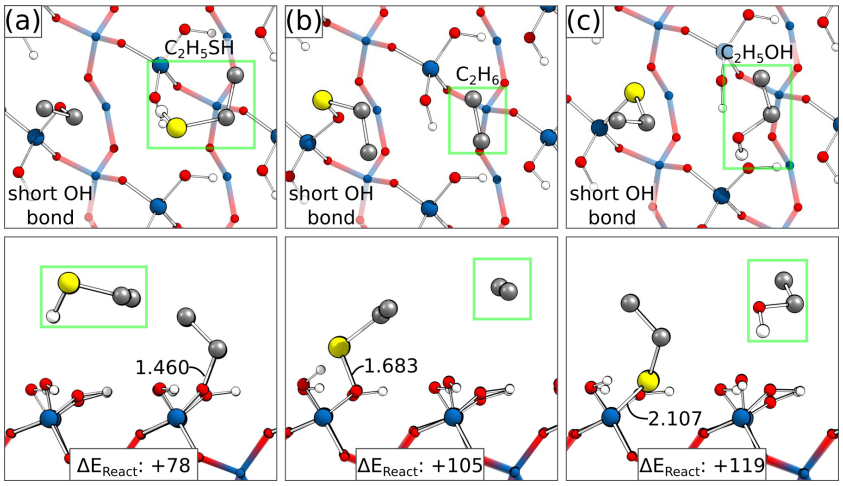


Figure D.3.: Reaction products of DES on SiO_2 formed at the short OH bond. (a) ethanethiol formation, (b) ethane formation and (c) ethanol formation. The formed molecules are marked in green. Reaction energies are stated relative to the most stable adsorption structure at the short OH bond. Energies in $kJ \cdot mol^{-1}$ and bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

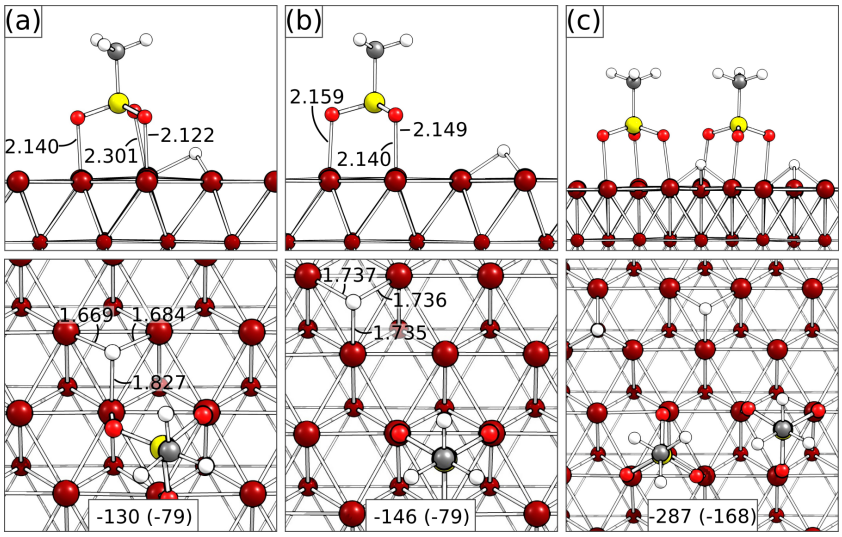


Figure D.4.: Reaction structures of MSA on Cu(111). (a,b) Alternative product structures of MSA after deprotonation. (c) Structure of two aligned MSA molecules after deprotonation. Since the adsorption energy for (c) is nearly twice the adsorption energy of (b) only a minor repulsion between the MSAs is present and a close packing can be expected. Adsorption energies with their dispersion contribution (shown in parentheses) in kJ·mol^{-1}. Bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Cu (copper red).

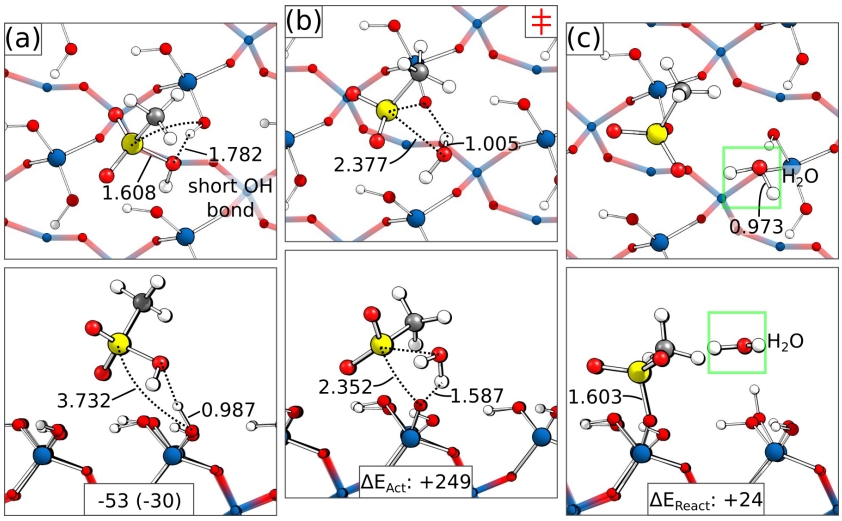


Figure D.5.: Adsorption and reaction structure of MSA on SiO_2. (a) Starting structure of the reaction path with MSA adsorbed in a short OH bond. (b) Observed transition state for the condensation reaction leading to (c) the product structure. Adsorption energies with their dispersion contribution (shown in parentheses), activation energies (ΔE_{Act}) and reaction energies (ΔE_{React}) in kJ·mol^{-1}. Bond lengths in Å. Color code: S (yellow), O (red), H (white), C (grey), Si (blue).

Erklärung

Ich erkläre, dass eine Promotion noch an keiner anderen Hochschule als der Philipps-Universität Marburg, Fachbereich Chemie, versucht wurde.

Hiermit versichere ich, dass ich die vorliegende Dissertation

Ab Initio Modelling of Chemical Vapor and Area-Selective Atomic Layer Deposition
Developing an Automated Exploration of Surface Reaction Networks

selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

_____ _____

Ort/Datum Unterschrift (Vor- und Nachname)