

Ultrathin insulator films from first principles:  
a ground and excited-state perspective

von  
Diplom-Chemiker  
Christoph Freysoldt

im Fachbereich Physik der Freien Universität Berlin  
eingereichte  
Dissertation zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM

Berlin 2007



Diese Arbeit wurde in der Zeit von Dezember 2002 bis Dezember 2006 unter der Aufsicht von Herrn Prof. M. Scheffler am Fritz-Haber-Institut der Max-Planck-Gesellschaft durchgeführt.

**Erstgutachter: Prof. M. Scheffler**

**Zweitgutachter: Prof. E.K.U. Gross**

**Disputationstermin: 7. Mai 2007**



# Contents

<b>1</b>	<b>Insulator films: from basic research to technological applications</b>	<b>1</b>
<b>2</b>	<b>Theoretical framework</b>	<b>8</b>
2.1	Density functional theory (DFT)	11
2.1.1	The exchange-correlation functional	14
2.1.2	Comparison of DFT results to experiments	15
2.2	Solving the Kohn-Sham equations: implementation and additional approximations	16
2.2.1	Plane-waves	16
2.2.2	Pseudopotentials	18
2.2.3	<b>k</b> -points	19
2.3	Many-body perturbation theory	20
2.3.1	Single-particle excitations in electron spectroscopy	20
2.3.2	Hedin's equations and the <i>GW</i> approximation	24
2.3.3	The quasiparticle equation	26
2.3.4	<i>GW</i> implementation: the space-time method	28
<b>3</b>	<b><i>GW</i> for surfaces and thin films</b>	<b>32</b>
3.1	Repeated-slab approach	32
3.2	The long-range tail of <i>W</i>	34
3.2.1	<b>k</b> -point sampling and interaction cell	35
3.2.2	The computation of <i>W</i> in reciprocal space	36
3.2.3	Anisotropy in the screened interaction	37
3.3	Screening in slab systems	42
3.3.1	Dielectric model	42
3.3.2	Parallel <b>k</b> -point sampling	43
3.3.3	Periodic repetition in the repeated-slab approach	52
3.4	Summary	56

<b>4 Ultrathin oxides: representative surface models or unique films?</b>	<b>58</b>
4.1 Introduction . . . . .	58
4.2 Properties of the bulk materials . . . . .	61
4.2.1 Sodium chloride . . . . .	61
4.2.2 $\alpha$ -Quartz . . . . .	62
4.2.3 $\alpha$ -Alumina . . . . .	64
4.2.4 Cubic hafnia . . . . .	65
4.2.5 Comparison of the oxides . . . . .	66
4.3 Substrate influence for silica films . . . . .	67
4.3.1 Atomic structure: the siloxane surface . . . . .	67
4.3.2 Electronic structure . . . . .	73
4.3.3 Comparison to UPS and MIES experiments . . . . .	76
4.3.4 How well do supported monolayer films model realistic surfaces? . . . . .	80
4.4 Thickness dependence in freestanding films . . . . .	81
4.4.1 $\alpha$ -Quartz . . . . .	82
4.4.2 $\alpha$ -Alumina . . . . .	86
4.4.3 Cubic hafnia . . . . .	89
4.5 Summary . . . . .	91
<b>5 Excited-state perspectives of ultrathin films</b>	<b>94</b>
5.1 Freestanding slabs . . . . .	96
5.2 NaCl films on Ge(001) . . . . .	101
5.2.1 Experimental situation . . . . .	101
5.2.2 Atomic structure of the interface . . . . .	103
5.2.3 $G_0W_0$ calculations . . . . .	107
5.3 Image effects in oxide films . . . . .	117
5.4 Summary . . . . .	120
<b>6 Conclusions</b>	<b>121</b>
6.1 Summary . . . . .	121
6.2 Outlook . . . . .	124
<b>A Abbreviations</b>	<b>127</b>
<b>B Dielectric models</b>	<b>129</b>
B.1 Image-charge method for dielectric layer models . . . . .	129
B.2 Connection to $G_0W_0$ : the static COHSEX approximation . . . . .	132

---

<b>C Ultrathin oxide films</b>	<b>135</b>
C.1 Alternative terminations for the $\alpha$ -quartz (0001) . . . . .	135
C.2 Surface-projected density of states . . . . .	137
<b>D NaCl films</b>	<b>139</b>
D.1 Atomic orbital projections . . . . .	139
D.2 NaCl: the character of the bulk conduction band . . . . .	140
D.3 Structure of the NaCl/Ge(001) interface . . . . .	144
D.4 STM simulations . . . . .	146
<b>E Computer code developments</b>	<b>149</b>
E.1 Green's function . . . . .	150
E.2 Block inversion of Hermitean packed matrices . . . . .	152
E.3 Disk storage . . . . .	155
E.4 Efficient computation of the matrix elements of $\mathbf{r}$ . . . . .	157
<b>F Computational parameters</b>	<b>160</b>
F.1 Pseudopotentials . . . . .	160
F.2 DFT-LDA calculations . . . . .	160
F.3 $G_0W_0$ calculations . . . . .	161
F.3.1 Time-frequency grids . . . . .	161
F.3.2 System-dependent convergence parameters . . . . .	161
F.3.3 Band cutoff . . . . .	162
<b>Bibliography</b>	<b>165</b>





# Chapter 1

## Insulator films: from basic research to technological applications

Thin films of ionic insulators have gained increasing interest over the last years due to their use in current and possible future technologies. In addition, they offer unique opportunities for studying fundamental questions in physics and chemistry. In this thesis we will concentrate on two choices out of the huge variety of possible materials and applications: wide-gap oxides, notably in the context of heterogeneous catalysis [1–3], and ultrathin sodium chloride films that have recently attracted attention as substrate for single-atom and single-molecule-manipulation experiments [4–6].

Focussing on oxides first, their surfaces and thin films play an important role in various technological applications such as heterogeneous catalysts, corrosion and scratch protection, or electronic devices. The decisive microscopic processes that control their functionality often take place at the atomic scale. For a systematic understanding and improvement of the relevant properties, a detailed atomistic understanding of these materials is therefore highly desired. Let us take heterogeneous catalysts as an example. Many present-day catalysts<sup>1</sup> consist of small metal particles on a ceramic support, which is often based on silica or alumina. A famous example is the catalytic converter in automobiles, in which platinum and rhodium is dispersed on a  $\gamma$ -Al<sub>2</sub>O<sub>3</sub> support with various additives. The support material, though chemically quite inert, exerts an important influence on the catalytic activity of the noble-metal catalyst. While a particular metal may be highly effective on

---

<sup>1</sup>Catalysts for redox reactions often belong to this category, while e.g. isomerisation reactions are catalysed by doped transition metal oxides or zeolites.

one support, it may be inactive on another. The reasons for this behaviour are presently not well understood. One possibility is that the oxide support directly modifies the properties of the metal clusters, another that it takes part in some of the many steps in the catalytic process.

Since real catalysts are very complex systems and not amenable to the full spectrum of the experimental surface science “tool box”, little is known about the microscopic and atomistic mechanisms that determine the catalytic activity under real-life conditions. In the surface science approach (schematically illustrated in Fig. 1.1) to heterogeneous catalysis the full problem is broken down into different aspects that can be addressed with well-defined model catalysts. For this, the ceramic support is replaced by a defined, clean and ideally crystalline oxide surface that can be studied with the different surface science experiments developed over the past decades, most often under ultrahigh vacuum conditions [3, 7, 8]. Although a number of important factors such as the influence of the gas phase (pressure gap, materials gap) or the temperature are neglected, this approach currently provides the best means to learn more about possible mechanisms for the catalytic processes from experiment.

An immediate problem of this approach lies in the insulating nature of the oxide materials. Certain techniques such as photoelectron spectroscopy (PES) or scanning tunnelling microscopy (STM) require electrically conducting samples, limiting their application to metals and semiconductors. The solution for insulating materials has been to grow thin epitaxial films on conducting substrates. The tunnelling current to the substrate is then sufficient to avoid charging problems. This strategy has been successfully applied to wide-gap oxides [1–3], but also to other insulators like NaCl [9, 10]. However, the use of epitaxial films to model the insulator surfaces in model catalysts introduces a further level of approximation as depicted on the right-hand side of Fig. 1.1.

The growth and the characterisation of well-ordered epitaxial insulator films has become an area of active research for this reason. Considerable progress has been made in the last decade for a large variety of oxide materials on various metal substrates [1, 2, 11, 12]. It was found that epitaxial oxide films can exhibit complex and unique structures not known from the bulk materials. Nevertheless, they often show bulk-like features in their vibrational and electronic spectra which has been regarded as an indication that the films represent the bulk materials reasonably well. Recently, however, the detailed atomic structure of several well-ordered oxide films used in this context has been resolved, e.g. silica/Mo(112) [13, 14] and alumina/NiAl(110) [15]. It was found that the films consist of only one or two monolayers of the oxide material, considerably less than previously thought. This raises

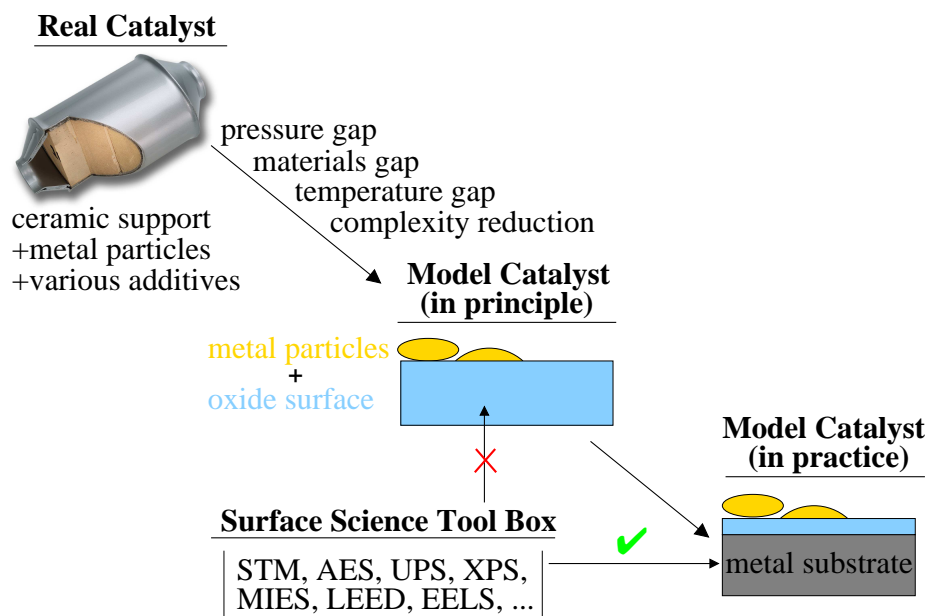


Figure 1.1: Schematic of the surface science approach applied to heterogeneous catalysts.

doubts about whether these films actually resemble the corresponding bulk materials and their surfaces. To study its validity will be a main topic of this thesis.

While earlier studies conducted with different experimental techniques emphasised the similarity of well-ordered thin films with their corresponding bulk materials, recent work has provided evidence for a film-specific behaviour. One example is the valence electron spectrum of thin silica films on Mo(112) which was found to depend on the film thickness [16, 17]. Another example is the adsorption of small linear gold clusters on alumina/NiAl(110) whose orientation was found to be closely matched with the NiAl substrate. Such an orientational preference is unexpected from the film's surface structure, providing strong indication that the substrate modifies the adsorption properties [18].

From a conceptual point of view one would thus like to isolate the influence of the substrate from those of the finite thickness or the thickness dependence, to analyse each contribution separately as depicted schematically in Fig. 1.2. However, experimentally this poses several challenges that cannot be met with currently available techniques. Not only is it not possible to produce a freestanding ultrathin oxide film experimentally, even the thickness dependence cannot be investigated for all crystalline films. While

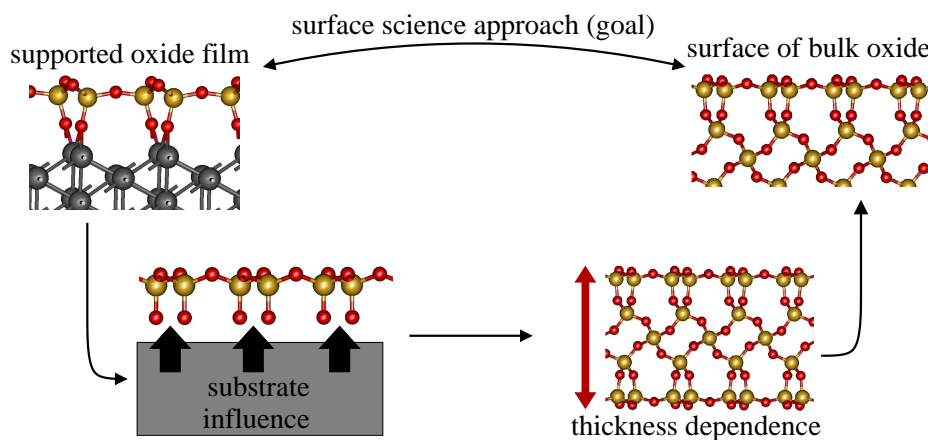


Figure 1.2: Important aspects of the comparison between supported thin films and bulk surfaces.

for some cases such as MgO on Mo(001) [19] or Ag(001) [20, 21] thicker high-quality crystalline films can be grown, most other oxide films become amorphous when the thickness exceeds one or two monolayers.

First principles simulations on the other hand are not bound to these constraints. In our simulations we have atomistic control over the structure and can take the compound system apart in a controlled manner. As shown in Fig. 1.2 for silica, we can then easily transform the  $\text{SiO}_2$  film on Mo(112) into an  $\alpha$ -quartz surface by first stripping away the metal substrate and then increasing the thickness of the free film. Using density-functional theory (DFT) further allows us to compute the equilibrium geometry at every step. In this way, we can not only disentangle the influence of the substrate from the thickness dependence, but moreover answer the question if the surface of an ultrathin silica film behaves like that of  $\alpha$ -quartz or not. We find that the substrate significantly modifies the atomic structure, which then induces noticeable differences in the electronic properties. Likewise, the thickness introduces large variations in the atomic and electronic structure of freestanding films.

The thickness dependence of certain properties is also an important question in microelectronics, where thin oxide films are frequently used. The functionality of integrated circuits for example is based on field-effect transistors (FET), see Fig. 1.3. In a FET, the gate electrode must be electrically insulated from the conducting channel that the gate electrode controls. The properties of this insulating oxide layer (“gate dielectric”) crucially influence the overall performance such as the switching speed or the power consumption [22]. With regard to the on-going miniaturisation in electronic devices,

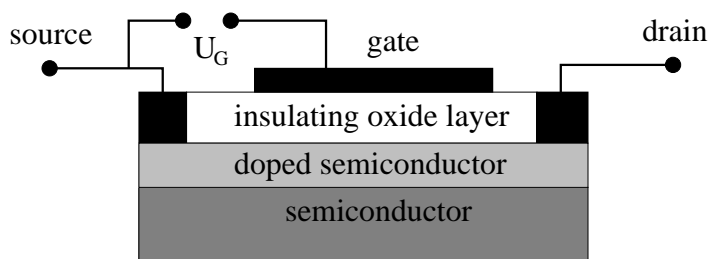


Figure 1.3: Metal-oxide-semiconductor field-effect transistor (MOSFET), schematically. By applying a gate voltage  $U_G$ , the charge carrier density in the doped semiconductor can be varied, which controls the current between source and drain.

the fundamental physical limits for the thickness of these films are of great interest. It is expected that silica, the gate dielectric in current silicon-based devices, will soon have to be replaced by another material with a larger dielectric constant. We have therefore extended our study of the thickness dependence to materials ( $\text{Al}_2\text{O}_3$ ,  $\text{HfO}_2$ ) that have been discussed in this context [22].

We will show in Sec. 4 that oxide films of one or two monolayer thickness are in general not representative for the bulk materials and their surfaces, with considerable differences in the atomic and electronic structure. For NaCl films, on the other hand, we find only negligible variations when we focus on the electronic ground-state effects. When comparing with spectroscopic measurements, however, ground-state DFT offers a slightly warped view, because self-energy effects are not included at this level of theory. The self-energy connects the DFT electronic structure to the quasiparticle spectrum measured in photoemission spectroscopy or other experiments that involve charged excited states. Since this type of spectroscopic tools is routinely employed to characterise epitaxial films, it is important to understand the origin of the experimental features in detail.

We employ the  $G_0W_0$  approximation (GWA), where  $G_0$  is the Green's function and  $W_0$  the screened Coulomb interaction, as a practical way to compute the self-energy corrections. In practice,  $G_0W_0$  not only introduces quasiparticle effects (e.g. the correct value of the gap or finite lifetimes), but also corrects for deficiencies in the underlying DFT starting point such as the self-interaction contained in many approximate density functionals. The GWA has been shown to describe the valence electron spectra in good agreement with experiment for many weakly correlated bulk systems such as the main group semiconductors and simple metals [23]. In order to use existing computer codes based on periodic boundary conditions for systems with bro-

ken periodicity such as films or surfaces, an artificial periodicity is introduced by repeating a slab of material in the direction perpendicular to its surface (repeated-slab approach). However, recent work on e.g. semiconductor slabs or nanowires has demonstrated that long-range polarisation effects require a special treatment for low-dimensional systems [24–27], questioning the reliability of the repeated-slab approach for  $G_0W_0$  calculations. In Sec. 3, we will present a thorough review of polarisation effects in  $G_0W_0$  calculations for the repeated-slab approach. From this, an efficient scheme will be developed to compute accurate quasiparticle band structures that no longer depend on the computational parameters in the numerical calculation.

Using this improved scheme, we have studied self-energy effects for sodium chloride films. NaCl is *the* prototypical ionic insulator, and it is therefore one of the most important toy systems for studying fundamental questions arising in this context. Recently, NaCl films have attracted renewed interest for the manipulation of single atoms [4] or molecules [5] in STM. Since STM experiments require a conducting substrate, the majority of studies is performed for metals or doped semiconductors. When a molecule is adsorbed on their surfaces, the molecular orbitals usually couple strongly to the substrate states or in other words the metallic substrate required for the experiment perturbs the adsorbate’s electronic structure. To study weakly perturbed adsorbates, ultrathin NaCl films of 1–4 monolayers (ML) thickness have been successfully used as insulating spacer layers to decouple the adsorbate levels from the substrate below, in this case Cu(001), Cu(111), or Cu(113) [6]. Repp *et al.* found that the STM then images the unperturbed molecular orbitals of the adsorbate [28]. Using the STM tip as local electrode, it is also possible to switch the charge state of an adsorbate, e.g. a gold atom, reversibly by applying short voltage pulses [4]. The high quantum yields achievable for this process suggest a two-step process in which the additional electron localises on the adsorbate before it tunnels into the substrate [6]. This implies a quasiparticle picture for the primary process, whose energetics is therefore expected to be well described by  $G_0W_0$  calculations.<sup>2</sup>

We have studied NaCl films on Ge(001), which represents a prototypical insulator/semiconductor interface. On this substrate, flat epitaxial NaCl films with a few ML thickness can be grown [9], and STM experiments can be carried out [29]. Since the structure of the NaCl/Ge interface is not known, we will first present a model derived from DFT-LDA calculations in Sec. 5.2.2 that is consistent with the experimental findings. We will then show that the

---

<sup>2</sup>DFT calculations qualitatively reproduce the observed STM pictures of the switched states, but failed to locate the energy position of the negative ion resonance responsible for the switching [4].

self-energy for films contains additional contributions beyond those for bulk NaCl. These arise from polarisation effects that become important when a charged state is created in a dielectrically heterogeneous environment. It is well known that image effects are taken into account by the  $G_0W_0$  self-energy [30], but their influence on adsorbate levels was demonstrated only very recently for a benzene molecule on graphite [31]. We have studied these effects also for an adsorbate on the supported NaCl films. As suitable adsorbate, we have chosen carbon monoxide, the “fruit fly” of molecular adsorption experiments in surface science.

These calculations not only demonstrate that a whole range of interesting phenomena can now be approached with *ab initio* methods. We can furthermore use the  $GW$  calculations to tailor a dielectric model for the polarisation effects. This in turn can be applied to the  $\text{SiO}_2/\text{Mo}(112)$  system, for which full  $G_0W_0$  calculations are presently not possible due to computational constraints<sup>3</sup>, to estimate the polarisation effects.

The remainder of this thesis is organised as follows. In Sec. 2, we will present the theoretical framework of the *ab initio* calculations performed in this work, namely density-functional theory in Sec. 2.1 and many-body perturbation theory in Sec. 2.3 as well as their implementation in the computer codes employed, **SFHingX** (DFT) and **gwst** (GWA). We will then discuss the implications of the repeated-slab approach for  $G_0W_0$  calculations in Sec. 3. Our DFT investigations of the structural and electronic properties of oxide films are presented in Sec. 4. The self-energy effects for thin films are discussed in Sec. 5 for the free-standing NaCl films and the NaCl/Ge(001) system, where we also present the atomistic model of the interface. We will then summarise our conclusions in Sec. 6.

---

<sup>3</sup>To treat metallic systems reliably in the  $GW$  space-time method, a finite-temperature formalism must be employed, which is currently under development [32].

# Chapter 2

## Theoretical framework

For studying the properties of thin films and surfaces on the atomic scale a wealth of experimental techniques is available. For brevity, we will introduce only those that will be of particular importance for thin insulator films and the questions addressed in this work. For developing a microscopic understanding of a material, the atomic structure is of fundamental interest. Crystalline materials are investigated by the diffraction of photons (X-ray diffraction, XRD), electrons (low-energy electron diffraction, LEED), and ions (low energy ion scattering, LEIS). Further insight, also for non-crystalline materials, can be gained by microscopies that reach atomic resolution, such as scanning tunnelling microscopy (STM) or atomic force microscopy (AFM). Indirect, but often invaluable information is deduced from other experimental data since many properties are strongly correlated to the local atomic structure. For example the vibrational spectrum can exhibit frequencies that are characteristic for certain atomic configurations such as hydroxyl groups. The phonon spectrum is measured by infrared (IR) spectroscopy, Raman spectroscopy, and also by high-resolution electron energy loss spectroscopy (HREELS).

A second key property is the electronic structure. While the knowledge of the electronic ground state is sufficient to describe the techniques described above, the electronic structure reveals itself only when excited. We can group the experimental techniques into those where the excited electrons remain in the sample such as optical spectroscopy in the visible and ultraviolet range (UV-VIS) or electron energy loss spectroscopy (EELS), and electron spectroscopies that change the number of electrons. Of particular importance is photoelectron spectroscopy (PES) where electrons are excited from the sample by electromagnetic radiation. Depending on the energy of the photons, one distinguishes between ultraviolet photoelectron spectroscopy (UPS) for the valence electrons and X-ray photoelectron spectroscopy (XPS) for



the core electrons. The inverse process (inverse photoelectron spectroscopy, IPES), i.e. the emission of photons when electrons are added to the system (bremsstrahlung isochromat spectroscopy, BIS), yields information about the unoccupied part of the spectrum. Also excited He atoms can be used to excite electrons (metastable impact electron spectroscopy, MIES).

The primary experimental data often provide only indirect information about the properties of interest, which must then be deduced from the experimental data by analogy to known systems (“empirical knowledge”). Usually, experimental results from several techniques must be combined to arrive at a consistent picture of the atomic scale. In this situation, theory can supplement the experiment in various ways. In addition to the empirical knowledge, appropriate theories help to assign experimental features to the corresponding microscopic structures or processes. Theoretical simulations can also add data not available from experiment such as the microscopic energetics, the electron distribution, the bonding behaviour, or the character of electronic states. Theory can thereby refine, validate, and sometimes reject the models proposed on the basis of experimental results. The increase in computer power has even made it possible to develop completely new models from extensive theoretical simulations to explain available experiments. The combination of theory and experiment has proved to be a powerful tool to solve complex questions in surface science. For this it is necessary to employ and develop *ab initio* theories that accurately describe the experimental results without relying on experimental input. To address the questions mentioned in the introduction, we have employed density functional theory and many-body perturbation theory in the *GW* approximation. These will be described in the following.

The physics at the atomic scale is governed by the principles of quantum mechanics. A system of atomic nuclei and electrons is described quantum-mechanically by its wavefunction  $\Psi$ , which depends on the coordinates of all electrons (indexed by  $i, j \dots$ ) and all nuclei (indexed by  $\mu, \nu \dots$ ). It solves the time-independent Schrödinger equation

$$H\Psi(\mathbf{x}_i, \mathbf{R}_\mu) = E \Psi(\mathbf{x}_i, \mathbf{R}_\mu), \quad (2.1)$$

where  $E$  denotes the total energy of the system and the Hamiltonian is<sup>1</sup>

$$\begin{aligned} H = & -\frac{1}{2} \sum_i \nabla_i^2 - \sum_\mu \frac{1}{2m_\mu} \nabla_\mu^2 - \sum_{i,\mu} \frac{Z_\mu}{|\mathbf{r}_i - \mathbf{R}_\mu|} \\ & + \sum_{\mu < \nu} \frac{Z_\mu Z_\nu}{|\mathbf{R}_\mu - \mathbf{R}_\nu|} + \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \end{aligned} \quad (2.2)$$

---

<sup>1</sup>In the following, Hartree atomic units are used, i.e.  $\hbar = m_e = 4\pi\epsilon_0 = e = 1$ .

According to the Copenhagen interpretation of quantum mechanics, the absolute square of the wavefunction is proportional to the probability density of finding a particle at each of its arguments. Particles of the same kind (e.g. electrons or identical nuclei) are indistinguishable. When two electrons (or other fermions) are exchanged, the wavefunction changes its sign (anti-symmetry) whereas it is symmetric with respect to boson exchange.<sup>2</sup>

Taking into account that the electrons are lighter than the nuclei by three to four orders of magnitude, we can employ the Born-Oppenheimer approximation [33]: assuming that the electrons adapt instantaneously to any movement of the nuclei, the motion of the electrons can be decoupled from that of the nuclei. For each atomic configuration, the electronic Schrödinger equation then reads

$$\left[ -\sum_i \frac{1}{2} \nabla_i^2 - \sum_{i,\mu} \frac{Z_\mu}{|\mathbf{r}_i - \mathbf{R}_\mu|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \Psi(\mathbf{x}_i) = E^{\text{el}}(\mathbf{R}_\mu) \Psi(\mathbf{x}_i). \quad (2.3)$$

The nuclei appear here only as the point charge sources of the electrostatic potential in which the electrons move. For generality, this potential will be denoted as the external potential  $V_{\text{ext}}$  in the following.

For the movement of the nuclei, only the electronic energy as a function of the nuclear positions is required. Including the internuclear repulsion (and possibly external fields acting on the nuclei), this is called the potential energy surface. An important task of simulations is to find and characterise minima on this potential energy surface since they correspond to the stable atomic structures. The nuclei in the real world are of course never at rest due to their quantum nature and thermal fluctuations. This does not harm the concept of an atomic structure because usually the nuclei fluctuate around the Born-Oppenheimer minimum. The structural parameters obtained from experiments that probe the average positions should thus in a first approximation correspond to the theoretically computed minimum-energy structures. Zero-point vibrations and the anharmonicity of the potential energy surface lead to small ( $\sim 0.5\%$ ) deviations. The movement of the nuclei is also neglected for electronic excitations (Franck-Condon principle), which appears justified for the photoemission processes. We therefore assume that the nuclei remain in their ground-state positions during the excitation process.

Even the electronic Schrödinger equation cannot be solved for realistic systems that contain dozens to thousands of electrons<sup>3</sup>. Moreover, the knowl-

<sup>2</sup>Bosons are particles with an integer spin while fermions have an half-integer spin.

<sup>3</sup>Of course, a macroscopic object has  $\sim 10^{23}$  electrons, but a fully quantum-mechanical description of such an object is neither feasible nor appropriate. The theoretical models must therefore be restricted to a small relevant part such as a cluster or the unit cell of a perfect crystal.

edge of the full wavefunctions is not required since experiment and simplifying theories probe only certain aspects of the wavefunction. Further simplifying approximations have to be made to the solution of the Schrödinger equation to drastically reduce the complexity and simplify the computation. As sketched out above, many properties depend on the electronic ground-state energetics. For this purpose, density functional theory (DFT), which builds on the ground-state electron density as the basic variable, provides an excellent choice. We will describe it in Section 2.1. From the DFT calculation, we extract experimentally observable quantities e.g. the atomic structure, the mechanical and elastic properties, the formation energy of compounds and many more. However, for a comparison with photoelectron spectroscopy we have to go beyond ground-state DFT<sup>4</sup>. An appropriate framework is Green's function theory, in particular many-body perturbation theory (MBPT) in the *GW* approximation (GWA). The GWA has been shown to describe the valence electron spectra in good agreement with experiment for many weakly correlated bulk systems such as the main group semiconductors and simple metals [23]. The GWA has also been applied to the surfaces of these materials with some success. We will describe the GWA and its connection with photoelectron spectroscopy in Section 2.3. In Section 3, the application of *GW* for surface and slab systems will be reviewed, highlighting critical points that have been neglected so far.

## 2.1 Density functional theory (DFT)

In this Section we will present density-functional theory in the Kohn-Sham (KS) formalism for the electronic ground state and furthermore discuss how the outcome of a DFT calculation can be compared to experimental results.

In density-functional theory, the electronic many-body problem (Eq. 2.3) is reduced to finding the electronic ground-state energy without formally including all the electronic degrees of freedom. Instead, the electron density

$$n(\mathbf{r}) = \int d^3r_1 d^3r_2 \dots d^3r_N |\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 \delta(\mathbf{r} - \mathbf{r}_1) \quad (2.4)$$

is used as the basic variable. Hohenberg and Kohn have proved that for a system with a non-degenerate groundstate, there exists a one-to-one mapping between the ground-state density and the external potential (up to a constant) in which the electrons are moving [34]. A more constructive proof

---

<sup>4</sup>Although in DFT every aspect of the electronic system is a functional of the ground-state electron density in principle, the focus in practice is on the ground-state total energy functional and derived quantities.

that also includes degenerate ground-states is due to Levy [35]. Since the external potential defines the electronic Hamiltonian uniquely, all electronic properties are uniquely defined from the ground-state density, too. The electronic energy of the system is thus a functional of the electron density. The true ground-state density minimises this global functional for a given external potential. However, its functional form is not known explicitly.

Kohn and Sham have further shown that the electron density  $n(\mathbf{r})$  can be reproduced by a fictitious system of non-interacting electrons moving in an effective field [36]. We denote the normalised, orthogonal one-electron wavefunctions by  $\phi_i$ . The antisymmetric many-electron wavefunction of  $N$  non-interacting electrons takes the form of a Slater determinant

$$\Psi(\mathbf{r}_1 \dots \mathbf{r}_N) = \sum_P \chi(P) \prod_{i=1}^N \phi_i(\mathbf{r}_{P(i)}) \quad (2.5)$$

where the sum runs over all permutations  $P$  of  $N$  numbers and  $\chi(P)$  is the character of the permutation (+1 for even, -1 for odd permutations). The density is then obtained from the occupied<sup>5</sup> one-particle wavefunctions  $\phi_i$  as

$$n(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2. \quad (2.7)$$

From the density, the Hartree energy, the classical repulsion of the charge distribution,

$$E_H = \frac{1}{2} \int d^3r \int d^3r' \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (2.8)$$

and the potential energy in the external field

$$E_{\text{ext}} = \int d^3r n(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) \quad (2.9)$$

can be obtained.

The advantage of the Kohn–Sham approach lies in the fact that a large part of the kinetic energy can be recovered by the kinetic energy  $T_s$  of the non-interacting electrons

$$T_s[\phi_i] = -\frac{1}{2} \sum_i \langle \phi_i | \nabla^2 | \phi_i \rangle, \quad (2.10)$$

---

<sup>5</sup>At zero temperature, the electron states are either occupied or unoccupied. For metals, it is numerically advantageous to assume an artificial temperature  $T$ . The states are then occupied according to the Fermi partition

$$f_i = \frac{1}{1 + e^{(\epsilon_i - \mu)/(k_B T)}} \quad (2.6)$$

where  $\mu$  is the Fermi energy and  $k_B$  the Boltzmann constant. All state summations then contain the occupation  $f_i$  as additional factors.

which is an explicit functional of the one-particle wavefunctions. Kohn and Sham proposed to separate of the electronic energy into the following contributions

$$E^{\text{el}} = T_{\text{s}}[\phi_i] + E_{\text{H}}[n] + E_{\text{xc}}[n] + E_{\text{ext}}[n], \quad (2.11)$$

where the exchange-correlation energy  $E_{\text{xc}}$  contains everything that is missing from the previous terms, notably the exchange energy, the correlation energy, but also the difference between the true kinetic energy and  $T_{\text{s}}$ . It also cancels the self-interaction contained in the Hartree energy.<sup>6</sup> Approximations to this exchange-correlation functional will be discussed below.

For the Kohn-Sham functional (Eq. 2.11), the variational principle applies, i.e. the ground state assumes the minimum of the functional. The variational derivatives of the explicitly density-dependent terms in Equation 2.11 with respect to the density, i.e. the Hartree potential

$$V_{\text{H}}(\mathbf{r}) = \frac{\delta E_{\text{H}}}{\delta n(\mathbf{r})} = \int d^3r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.12)$$

the exchange-correlation potential<sup>7</sup>

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})}, \quad (2.13)$$

and the external potential can be combined into one local effective potential

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) + V_{\text{ext}}(\mathbf{r}). \quad (2.14)$$

Minimisation with respect to the orbitals  $\phi_i$  under the constraint that these stay orthonormal leads to the Kohn-Sham equations

$$\left\{ -\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}) \right\} \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}), \quad (2.15)$$

where  $\epsilon_i$  are the Lagrangian multipliers that result from the normalisation constraint. They have the dimension of an energy and are therefore often referred to as Kohn-Sham one-particle energies. Since the effective potential depends on the density, which in turn depends on the one-particle wavefunctions, the Kohn-Sham equations have to be solved self-consistently.

---

<sup>6</sup>The self-interaction arises because each electron feels the potential of all electrons instead of all *other* electrons as it should.

<sup>7</sup>We note some formal subtleties with this definition because the variational derivative may not exist. For the approximations to the exchange-correlation functional discussed below, however, the exchange-correlation potential is well defined.

### 2.1.1 The exchange-correlation functional

In order to turn KS-DFT into a practical computational scheme, one the exchange-correlation functional must be approximated by an explicit expression. One of the earliest is the local-density approximation (LDA)

$$E_{\text{xc}}^{\text{LDA}}[n] := \int d^3r n(\mathbf{r}) \epsilon_{\text{xc}}^{\text{HEG}}(n(\mathbf{r})), \quad (2.16)$$

in which the local exchange-correlation energy per electron is approximated by that of an homogeneous electron gas (HEG) of the same density as that of point  $\mathbf{r}$  [37]. For the HEG, the exchange-correlation energy density  $\epsilon_{\text{xc}}^{\text{HEG}}(n)$  are known in the low-density and high-density limits [38, 39]. For intermediate densities it has been very accurately computed from quantum Monte-Carlo simulations by Ceperley and Alder [40].

The LDA should be a good approximation for slowly varying densities, but it has proved to be very successful for a variety of rather inhomogeneous systems such as atoms, molecules, and solids, too. However, the LDA has a tendency to overestimate the strength of covalent chemical bonds, i.e. binding/cohesive energies are found too large, bond lengths and lattice constants too small. This may be attributed to the fact that the LDA underestimates the exchange-correlation energy in regions of strongly varying density and therefore favours compact density distributions. Furthermore, the LDA tends to delocalise electron states since it is not generally self-interaction free. The LDA is also expected to fail for extremely inhomogeneous systems and in cases where the correlation is strongly non-local. For instance, London dispersion interactions between two polarisable, but separated molecules (or solids) result from the correlation of the density fluctuations in the two systems and are not contained in the LDA. Another example are Mott-Hubbard insulators which exhibit partially occupied, localised electron states at different sites. The Coulomb repulsion between the electrons leads to a strong correlation between the sites that is not recovered by the LDA. However, since none of these effects plays an important role for the wide-gap insulators investigated in this work, we found that we can safely employ the LDA.

Many schemes to improve upon the LDA have been suggested, but we will mention only a few of them here. The generalised gradient approximation (GGA) is an attempt to remedy the neglect of the density variations by including the local gradient of the density in the kernel of the exchange-correlation energy functional

$$\epsilon_{\text{xc}}^{\text{GGA}}(n, |\nabla n|) := \epsilon_{\text{xc}}^{\text{HEG}}(n) + \Delta\epsilon_{\text{xc}}(n, |\nabla n|). \quad (2.17)$$

The gradient correction  $\Delta\epsilon_{\text{xc}}$  systematically reduces the energy for density inhomogeneities and increases the bond lengths and lattice constants. How-

ever, this does not always improve the agreement with experiment. A different approach is exact-exchange (EXX). Here, the exact expression for the exchange energy of a Slater determinant is taken as the exchange functional:

$$E_x = -\frac{1}{2} \sum_{ij} \langle \phi_i(\mathbf{r})\phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}')\phi_j(\mathbf{r}) \rangle . \quad (2.18)$$

This exchange energy removes the self-interaction. The exchange functional Eq. 2.18 is an orbital functional rather than a density functional. A direct minimisation with respect to the orbitals would lead to the Hartree-Fock method. In EXX, however, the Kohn-Sham approach is followed: A local exchange potential is constructed following the optimised effective potential (OEP) method to evaluate the functional derivative Eq. 2.13 of the orbital functional Eq. 2.18. EXX is not yet widely employed since it is computationally much more demanding than the LDA or GGA.

For molecular applications, hybrid functionals are often used that linearly combine several exchange-correlation functionals with empirically determined mixing coefficients. Most popular is the B3LYP functional with three independent mixing parameters. It also includes the exact exchange expression Eq. 2.18, but instead of constructing a local potential from it, it is directly used as an orbital functional similar to Hartree-Fock. The additional ingredients are the LDA and the gradient corrections of Becke (for exchange) and Lee, Yang, and Parr (LYP, for correlation). B3LYP gives accurate structural and energetic results for a large number of molecules, but the computational effort combines that of DFT with Hartree-Fock. In essence, none of these functionals is generally superior to the others. Instead, the applicability and accuracy has to be tested for each system.

### 2.1.2 Comparison of DFT results to experiments

The basic output of a DFT calculation is the total energy and the electron density for a given atomic configuration. We can thus explore the nuclear potential energy surface and extract a number of interesting properties from it. The minimisation of the energy with respect to the atomic positions yields the atomic structure. From the curvature of the potential energy surface at the minimum, we can further deduce the vibrational properties of the system in the harmonic approximation. By comparing the total energy between different systems or different minima on the potential energy surface for the same system, we obtain basic thermodynamical data such as the binding, cohesive, formation, and reaction energies. Since the vibrational entropies can be computed from the vibrational frequencies and hence from

the potential energy surface, the combination of statistical mechanics with DFT opens a microscopic approach to the thermodynamical properties (*ab initio* thermodynamics). Likewise, the saddle points between two minima on the same potential energy surface might be used in transition state theory to predict kinetic constants. Alternatively, the potential energy surface can be explored with molecular dynamics techniques.

## 2.2 Solving the Kohn-Sham equations: implementation and additional approximations

In this Section we will summarise how the Kohn-Sham equations are solved in practice in the SFHIngX program package using a plane-wave basis set and pseudopotentials. For a more detailed description of the algorithms employed, we refer to the review article by Payne *et al.* [41], the description of the SFHIngX predecessor fhi96md [42], the pseudopotential generator fhi98pp [43] and the SFHIngX manual [44]. The purpose of this section is to introduce the additional approximations in the calculations. We note that many of the properties of the plane-waves and the  $\mathbf{k}$ -point sampling are relevant for the *GW* space-time method described below, too.

### 2.2.1 Plane-waves

In practical computation, the Kohn-Sham wavefunctions have to be expanded in a finite basis set, which transforms the analytic eigenvalue problem Equation 2.15 into an algebraic one. In this work, plane-waves  $e^{i\mathbf{k}\cdot\mathbf{r}}$  are employed as basis functions which are particularly suitable for periodic systems. Let the real space lattice of the periodic system be given by the basis vectors  $\mathbf{a}_i, i \in \{1, 2, 3\}$ . The reciprocal lattice basis  $\mathbf{b}_i$  is then defined by

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij} , \quad (2.19)$$

where  $\delta_{ij}$  denotes the Kronecker- $\delta$ . In general, we will denote real space lattice vectors with  $\mathbf{R}$  and reciprocal space lattice vectors with  $\mathbf{G}$ .

According to the Bloch theorem [45], the eigenfunctions in a periodic system can be written as

$$\phi(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} , \quad (2.20)$$



where  $\mathbf{k}$  is in the first Brillouin zone<sup>8</sup> and  $u_{\mathbf{k}}(\mathbf{r})$  is a lattice-periodic function, i.e.

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) \quad (2.21)$$

for any lattice vector  $\mathbf{R}$ . When  $u_{\mathbf{k}}$  is expanded in plane-waves, only wave vectors  $\mathbf{G}$  of the reciprocal lattice contribute. We will often denote the plane-wave representation of a function as its 'reciprocal-space' representation.

Plane-waves are advantageous for many of the various steps in a DFT-KS calculation. Plane-waves form an orthonormal basis set, which makes the normalisation and the orthogonalisation very simple operations. The Laplace operator

$$\nabla^2 e^{i\mathbf{k}\cdot\mathbf{r}} = -k^2 e^{i\mathbf{k}\cdot\mathbf{r}} \quad (2.22)$$

(and correspondingly the Coulomb potential) becomes a multiplicative operator in reciprocal space. For the application of local potentials (cf. Eq. 2.15) as well as for the computation of the electron density (Eq. 2.7) a real-space representation of the wavefunctions is required. The transformation to real space on a regular grid as well as the reverse transformation can be efficiently computed with Fast Fourier Transforms (FFTs). For the Hartree potential, the density computed on the FFT grid in real space is transformed to reciprocal space, multiplied with the Coulomb potential

$$v(\mathbf{G}) = \frac{4\pi}{|\mathbf{G}|^2} \quad (2.23)$$

and transformed back to real space.

The plane-wave basis is made finite with a single energy cut-off parameter  $E_{\text{cut}}$  that corresponds to the maximum kinetic energy

$$\frac{1}{2}|\mathbf{k} + \mathbf{G}|^2 \leq E_{\text{cut}} . \quad (2.24)$$

The FFT grids must be large enough to contain plane-waves up to  $2E_{\text{cut}}$  to avoid aliasing effects for the product of two functions. The wavefunctions are therefore stored in their reciprocal space representation, which is about 16 times smaller than the real space representation<sup>9</sup>. However, for describing the oscillations of the wavefunctions and the steep potential close to the nuclei, very high plane-wave cutoffs would be necessary. This motivates the use of pseudopotentials, which are described next.

<sup>8</sup>The first Brillouin zone comprises all points in reciprocal space that are closer to the origin (denoted as  $\Gamma$ ) than to any other reciprocal lattice point.

<sup>9</sup>The circumscribing cube of the cut-off sphere has a volume of  $\sim (2\sqrt{E_{\text{cut}}/2})^3$  compared to the cut-off sphere volume of  $4\pi/3(\sqrt{E_{\text{cut}}/2})^3$ , which gives a ratio of  $6/\pi \approx 2$ . A further factor of  $2^3 = 8$  results from the double grid-size in each dimension to prevent aliasing.

## 2.2.2 Pseudopotentials

In order to obtain smoothly varying wavefunctions and potentials, pseudopotentials are introduced. Only the valence electrons remain in the Kohn-Sham computation, while the effect of the core on the valence electrons is simulated by the pseudopotential and – in certain cases – an auxiliary pseudo-core density to better describe the non-linear behaviour of exchange and correlation between core and valence states. Likewise, the oscillations of the valence orbitals close to the nuclei that result from the orthogonalisation to the core states are replaced by a smooth part. Various flavours of pseudopotentials exist. In this work, *ab initio* norm-conserving pseudopotentials in the Kleinman-Bylander form [46] are used. We will explain these terms in the following. For explicit expressions, we employ spherical coordinates, i.e. the radial coordinate  $\rho$  and the space angle  $\Omega$ . The angular dependence can be expanded in spherical harmonics  $Y_{lm}$ , where  $l$  and  $m$  denote the angular and magnetic quantum number, respectively.

*Ab initio* pseudopotentials are derived from an all-electron calculation for the atom by inverting the Kohn-Sham equation, i.e. the effective potential inside a cut-off radius is computed from a smooth pseudo-wavefunction  $\phi_{nl}^{\text{ps}}(\rho)$ . and the all-electron Kohn-Sham energy of the valence orbitals  $\epsilon_{nl}$ . These potentials  $V_l(\rho)$  depend on the angular momentum quantum number. The norm-conservation implies that the pseudoised part of the valence functions has the same norm as the corresponding all-electron wavefunction.

Pseudopotentials in this semilocal form are not very efficient because the projection onto the local angular momentum is computationally expensive. For numerical convenience, they are transformed into the separable Kleinman-Bylander form. A Kleinman-Bylander pseudopotential consists of a local pseudopotential<sup>10</sup>  $V^{\text{loc}}(\rho)$  and additional pseudopotential projectors  $\chi_{nlm}$ . In Dirac notation, the non-local pseudopotential term in the Hamiltonian reads

$$V^{\text{nl}} = \sum_{nlm} |\chi_{nlm}\rangle E_{nlm} \langle \chi_{nlm}|, \quad (2.25)$$

where the Kleinman-Bylander energies  $E_{nlm}$  describe the strength of the pseudopotential. The projectors for each atom are of the form

$$\chi_{nlm}(\rho, \Omega) = f_{nl}(\rho) Y_{lm}(\Omega), \quad (2.26)$$

where  $f_{nl}$  is a radial function.  $n$  is an index to distinguish projectors for the same angular momentum. The Kleinman-Bylander projectors and energies

---

<sup>10</sup>Usually, one of the  $V_l$  is chosen for this local potential.

are derived from  $l$ -dependent, semilocal pseudopotentials and the corresponding atomic pseudo-wavefunctions:

$$\Delta V_l = V_l - V^{\text{loc}} \quad (2.27)$$

$$f_{nl}(\rho) = \Delta V_l(\rho) \phi_{nl}^{\text{ps}}(\rho) \quad (2.28)$$

$$E_{nlm} = \frac{1}{\langle \phi_{nl}^{\text{ps}} | \Delta V_l | \phi_{nl}^{\text{ps}} \rangle}, \quad (2.29)$$

This definition ensures that the matrix elements of the the original semilocal form and the Kleinman-Bylander separable form agree for the atomic states. The pseudopotentials employed in this work use a single projector per  $l$ -channel. This restriction limits the accuracy which can be obtained with the Kleinman-Bylander pseudopotential. For the bulk systems used in this work, the pseudopotential results were compared to all-electron calculations to ensure that no critical errors are introduced by this approximation.

All the norm-conserving pseudopotentials used in this work were constructed with the `fhi98pp` program [43] according to the Hamann [47] or Troullier-Martins [48] scheme. A pseudopotential further depends on the cut-off radius defining where the pseudo-wavefunctions must coincide with (Troullier-Martins) or differ negligibly from (Hamann) the true all-electron wavefunctions. In addition, the occupation numbers can be varied from the ground-state occupations to improve the performance of the pseudopotentials. For the pseudopotentials used in this work, these parameters were varied to find an optimal compromise between the transferability, numerical efficiency (in terms of the required plane-wave cut-off energy), and reliability of the pseudopotentials. For the relevant bulk systems, the pseudopotentials were tested against all-electron calculations and showed a very good, sometimes excellent agreement. We are therefore confident that the accuracy is sufficient for the questions of this work. In general, however, the limited accuracy due to restrictions in the form of the pseudopotential (norm-conservation, only one projector per  $l$ -channel) may require to go over to more general forms such as ultrasoft Vanderbilt pseudopotentials [49] or projector-augmented waves [50] at the expense of computational simplicity and efficiency.

### 2.2.3 k-points

The Brillouin zone vector  $\mathbf{k}$  is a continuous index; summations then formally become integrals over the Brillouin zone. In practice, these integrals are again approximated by finite summations over grids in the Brillouin zone. A regular  $\mathbf{k}$ -point integration (or rather summation) grid can be obtained by

a procedure due to Monkhorst and Pack [51] (Monkhorst-Pack mesh): The basis vectors  $\mathbf{b}_i$  of the reciprocal lattice are reduced by a certain integer factor  $N_i$  (the “folding”). The reduced basis  $\mathbf{b}_i/N_i$  then defines the grid spacing in the Brillouin zone. The integration grid itself can be centred on the  $\Gamma$ -point  $\mathbf{k} = \mathbf{0}$  or at an offset<sup>11</sup>. The Monkhorst-Pack mesh is then obtained by keeping only one representative for each class of symmetry-equivalent points (a ‘star’).

The  $\mathbf{k}$ -point sampling is an important convergence parameter. The Kohn-Sham band-structure Hamiltonian  $H_{\mathbf{k}}$  and correspondingly its eigenfunctions and eigenenergies vary smoothly over the Brillouin zone. However, this may no longer be the case for the corresponding occupation numbers in the case of metals. While for insulators small  $\mathbf{k}$ -samplings prove to be sufficient, the sampling in metals must be fine enough to resolve the details of the Fermi surface which separates the occupied from the unoccupied regions in the Brillouin zone. Also other entities that involve integrations over the Brillouin zone and depend sensitively on the Kohn-Sham energies may require large foldings even for insulators.

## 2.3 Many-body perturbation theory

In this section, we will briefly address the connection of electron spectroscopy and Green’s function theory, before we show how the Green’s function and in particular its poles can be computed with many-body perturbation theory and the *GW* approximation. At the end of this Chapter, we will present the implementation of the *GW* equations in the *GW* space-time method used throughout in this work.

### 2.3.1 Single-particle excitations in electron spectroscopy

In a direct PES experiment, a sample is irradiated with light of energy  $h\nu$ . The sample then emits electrons with a characteristic kinetic energy  $E_{\text{kin}}$  that is measured. From the energy conservation, the binding energy of the emitted electron

$$\epsilon_i = h\nu - E_{\text{kin}} . \quad (2.30)$$

can be deduced. In an inverse photoemission experiment, the process is reversed: electrons with a (low) kinetic energy are shot at the sample and will finally undergo a radiative transition to a low-lying unoccupied state  $\epsilon_f$ ,

---

<sup>11</sup>The offset is usually given in terms of the grid lattice, the offset coordinates range between 0 and 1

thereby emitting light. The energy  $h\nu$  of this light is measured, which allows to reconstruct the final state energy as

$$\epsilon_f = E_{\text{kin}} - h\nu . \quad (2.31)$$

The experimental observable in PES is the photocurrent. It is given by [52, 53]

$$I \sim \int d\mathbf{x} \int d\mathbf{x}' \phi_{\text{pe}}^*(\mathbf{x}) \delta H(\mathbf{x}) A(\mathbf{x}, \mathbf{x}', E) \delta H(\mathbf{x}') \phi_{\text{pe}}(\mathbf{x}') , \quad (2.32)$$

where  $\mathbf{x}$  comprises a spatial and a spin coordinate.  $\phi_{\text{pe}}$  is a time-reversed damped LEED state describing the photoelectron that reaches the detector.  $\delta H(\mathbf{x})$  is the perturbing field that excites the electron. While these two ingredients are characteristic to the photoemission process, the electronic structure of the sample is contained in the spectral function

$$A(\mathbf{x}, \mathbf{x}', E) = \sum_s f_s(\mathbf{x}) f_s^*(\mathbf{x}') \delta(E - E_s) , \quad (2.33)$$

where  $s$  denotes the excited states of the system. The electron binding energy, or more precisely: electron removal energy,  $E_s$  and the transition amplitude  $f_s(\mathbf{x})$  in Eq. 2.33 are defined from the many-body states of the  $N$  and  $N - 1$  electron systems via

$$f_s(\mathbf{x}) = \langle N - 1, s | \hat{\psi}(\mathbf{x}) | N, 0 \rangle \quad E_s = E_{N,0} - E_{N-1,s} < E^{\text{Fermi}} , \quad (2.34)$$

where  $\hat{\psi}$  denotes the field annihilation operator that creates an excited state of the  $N - 1$  electron system from the ground state of the  $N$  electron system. The electron addition energies relevant for IPES are analogously defined from the  $N + 1$  electron system as

$$f_s(\mathbf{x}) = \langle N, 0 | \hat{\psi}(\mathbf{x}) | N + 1, s \rangle \quad E_s = E_{N+1,s} - E_{N,0} > E^{\text{Fermi}} . \quad (2.35)$$

Inserting Eq. 2.33 into Eq. 2.32 we obtain Fermi's golden rule expression

$$I \sim \sum_s |\langle \phi_{\text{pe}} | \delta H | f_s \rangle|^2 \delta(E - E_s) . \quad (2.36)$$

The magnitude of the transition matrix elements can vary considerably between different states and even become zero, in particular when the symmetry of the system defines selection rules. The photocurrent therefore reflects only a somewhat distorted picture of the electron density of states (DOS)

$$N^{\text{el}}(E) = \sum_s \delta(E - E_s) = \int d^3\mathbf{x} A(\mathbf{x}, \mathbf{x}, E) , \quad (2.37)$$

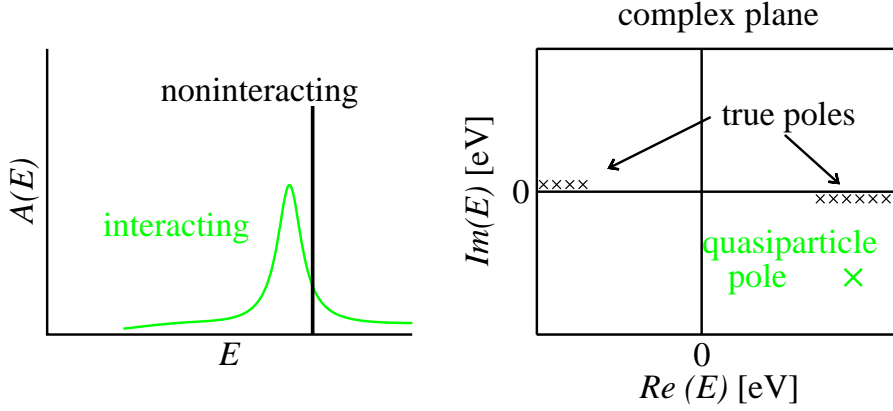


Figure 2.1: Left: schematic representation of a non-interacting and interacting spectral function  $A(E)$  on the real energy axis. Right: the pole structure in the complex plane.

the entity preferably used in theoretical studies to discuss the electronic structure because it does not depend on the experimental setup.

The connection to Green's function theory is given by the Lehmann representation of the one-particle Green's function

$$G(\mathbf{x}, \mathbf{x}', \omega) = \sum_s \frac{f_s(\mathbf{x}) f_s^*(\mathbf{x}')}{\omega - E_s \pm i\eta}, \quad (2.38)$$

where  $i\eta$  is an infinitesimal imaginary part. The positive (negative) sign applies for states above (below) the Fermi energy. The spectral function is then given by

$$A(\mathbf{x}, \mathbf{x}', \omega) = \frac{1}{\pi} \Im G(\mathbf{x}, \mathbf{x}', \omega). \quad (2.39)$$

The poles of the Green's function hence correspond to the one-particle addition and removal energies, also known as single-particle excitations. In an extended interacting system, the infinitely many excitations can merge into peak-like features that can be effectively described by a single pole (cf. Fig. 2.1). This pole has a finite imaginary part and describes the quasiparticle excitation. The real part of the quasiparticle energy corresponds to the peak maximum of the energy distribution when an electron is added or removed, for instance in an (inverse) photoemission experiment. The imaginary part determines the life-time broadening (peak width).

It must be emphasised that the concept of a quasiparticle is an interpretation of the experimentally observed many-body spectrum in terms of one-particle-like excitations. The quasiparticle concept is used for various types of excitations, for instance electron-hole pairs (excitons) or vibrations

(phonons). In the context of one-electron excitations, these quasiparticles can be viewed as a hole or an electron surrounded by its polarisation cloud. Quasiparticles have a finite lifetime due to dephasing, i.e., decay into other quasiparticles. because the true many-body eigenstates may contribute to more than one quasiparticle. In other words, a quasiparticle is not an eigenstate of the system but a superposition of eigenstates. The actual formation and decay of a quasidelectron in silicon has recently been followed in a time-resolved pump-probe experiment for the first time [54].

The connection between the Green's function and the Hamiltonian can be schematically written as

$$G(\omega) = [\omega - H]^{-1} . \quad (2.40)$$

Here,  $H$  denotes the effective one-particle Hamiltonian, which implicitly contains the electron-electron interaction. It can be split into a pure non-interacting part  $H_0$  (in which we include the Hartree potential) and a non-local, energy-dependent self-energy  $\Sigma$ , i.e.

$$H(\mathbf{x}, \mathbf{x}', \omega) = H_0(\mathbf{x}) + \Sigma(\mathbf{x}, \mathbf{x}', \omega) . \quad (2.41)$$

In analogy to Eq. 2.40, a non-interacting Green's function is defined from  $H_0$ . The connection between the non-interacting and the interacting Green's function is given by the non-local, energy-dependent self-energy  $\Sigma$  and will be discussed in the next section. Before we come to this part, we briefly address the use of DFT.

A byproduct of the DFT-KS calculation are the eigenenergies  $\epsilon_i$  and eigenfunctions  $\phi_i$  of the Kohn-Sham Hamiltonian  $H^{KS}$ . Strictly speaking, they are no physical observables except for the energy of the highest (partially) occupied state which equals the chemical potential of the electrons in the system (Janak's theorem [55]). However, we may use them as first approximations to the quasiparticle energies and functions. Correspondingly, a non-interacting DFT-KS Green's function  $G^{KS}$  can be defined in analogy to Eq. 2.38. Such an approach can indeed explain a number of qualitative features of the quasiparticle spectra, owing much to the fact that the true Hamiltonian and the Kohn-Sham Hamiltonian share the non-interacting part  $H_0$ .

However, DFT-KS cannot explain all aspects of the quasiparticle band structure. In addition, the available approximate functionals introduce further errors. The most important failure is the underestimation of the quasiparticle band gap in the LDA by typically 50–100% [23, 56]. This is partially due to inherent deficiencies of the LDA such as the self-interaction, which pushes the occupied states up in energy. However, even the KS band gap

using the exact exchange-correlation potential would differ from the quasiparticle gap because the true xc functional exhibits a derivative discontinuity<sup>12</sup> at integer particle numbers [57]. Nevertheless, it has been found that the LDA wavefunctions are reasonable approximations to the quasiparticle functions in many bulk systems [23, 58, 59]. We will come back to this point later when we discuss quasiparticle corrections.

### 2.3.2 Hedin's equations and the $GW$ approximation

Hedin has shown that the problem of computing the interacting one-particle Green's function can be cast into five coupled equations in the framework of many-body perturbation theory [60]. These equations involve the independent-particle and full Green's function  $G_0$  and  $G$ , the polarisation  $P$ , the bare and screened interaction  $v$  and  $W$ , the self-energy  $\Sigma$ , and the so-called vertex function  $\Gamma$ . For simplicity, we use the notation  $1 = \mathbf{x}_1 t_1$  for every pair of one spatial and one temporal variable. A  $+$  indicates that the time argument has been increased by an infinitesimally small, positive amount. Hedin's equations are:

$$P(1, 2) = -i \int d3d4 G(1, 3)G(4, 1)\Gamma(3, 2, 4) \quad (2.42)$$

$$W(1, 2) = v(1, 2) + \int d3d4 v(1, 3)P(3, 4)W(4, 2) \quad (2.43)$$

$$\Sigma(1, 2) = i \int d3d4 G(1, 3)W(4, 1^+)\Gamma(3, 4, 2) \quad (2.44)$$

$$\begin{aligned} \Gamma(1, 2, 3) &= \delta(1, 2)\delta(2, 3) \\ &+ \int d4d5d6d7 \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)}G(4, 6)G(7, 5)\Gamma(6, 7, 3) \end{aligned} \quad (2.45)$$

$$G(1, 2) = G_0(1, 2) + \int d3d4 G_0(1, 3)\Sigma(3, 4)G(4, 2) \quad (2.46)$$

This set of coupled equations cannot be solved directly due to the presence of the functional derivative in the definition of the vertex function. However, it is amenable to physically meaningful approximations. We will describe the equations in more detail before coming to the most important approximation, the  $GW$  approximation.

We first note that Equations 2.43 and 2.46 have the same structure, known as a Dyson equation. It describes the redressing of independent-particle propagators by interactions with the system, which for the electrons is called self-energy.  $G_0$  and  $G$  correspond to the propagation of electrons or holes in the system, whereas  $v$  and  $W$  can be interpreted as propagators for the

<sup>12</sup>DFT can be extended to fractional electron numbers using ensemble DFT.



quantum particles of the electric field. Thus, the polarisation function plays the role of the self-energy for the electric field particles, changing the bare Coulomb interaction into the screened Coulomb interaction. Correspondingly, the equations for the interaction kernels  $\Sigma$  and  $P$ , Eq. 2.44 and 2.42, have the same structure. They contain two propagators connected to one point of the interaction kernel and are connected via the vertex function to the other point of the kernel. This vertex function is a three-point kernel that describes all possible ways how a (dressed) electron is scattered when a (screened) electric field particle is created or annihilated.

Approximating the vertex function by its first term, i.e.  $\Gamma = \delta\delta$ , we arrive at the random phase approximation for the polarisation

$$P(1, 2) = -iG(1, 2)G(2, 1) \quad (2.47)$$

and the  $GW$  approximation for the self-energy

$$\Sigma(1, 2) = iG(1, 2)W(2, 1^+) . \quad (2.48)$$

When the Dyson equation for the screened interaction is inverted, one obtains schematically

$$W^{-1}(1, 2) = v^{-1}(1, 2) - P(1, 2) . \quad (2.49)$$

For the practical calculation, this equation is usually transformed<sup>13</sup>. Multiplying from left and right with the square root<sup>14</sup> of the Coulomb potential  $v^{1/2}$  and integrating yields the symmetrised dielectric function

$$\tilde{\varepsilon}(1, 2) = \delta(1, 2) - \int d3d4 v^{1/2}(1, 3)P(3, 4)v^{1/2}(4, 2) , \quad (2.50)$$

from which the screened interaction can be computed via

$$W(1, 2) = \int d3d4 v^{1/2}(1, 3)\tilde{\varepsilon}^{-1}(3, 4)v^{1/2}(4, 2) . \quad (2.51)$$

It must be emphasised that for a plane-wave basis, the integrations appear only formally since  $v^{1/2}$  is diagonal in reciprocal space (see also Sec. 2.2.1).

---

<sup>13</sup>Since  $v$  and  $W$  are singular in reciprocal space, the plane-wave representation of  $W$  is “ill conditioned”, i.e. the magnitude of the matrix elements varies strongly. This may lead to numerical inaccuracies or instabilities in the numerical inversion.

<sup>14</sup>The square root of a  $v(1, 2)$  is naturally defined by

$$\int d3 v^{1/2}(1, 3)v^{1/2}(3, 2) = v(1, 2) ,$$

and is given by  $2\sqrt{\pi}/k$  in reciprocal space and by  $\pi^{-3/2}/r^2$  in real space[61].

In practice, it is very common to employ further approximations. To solve 2.47 – 2.51, it would seem reasonable to iterate Eq. 2.46 and Eq. 2.47 – 2.51 to self-consistency. However, this is rarely done. Instead of using the full Green’s function  $G$ , the non-interacting Green’s function  $G_0$  is used as a first approximation in Eq. 2.47 and Eq. 2.48. We use this approach, denoted  $G_0W_0$ , in all the actual calculations. The quality of this approximation has been under debate over the last years. It proves to be very successful for describing the pole structure of the Green’s function in connection with pseudopotentials and Kohn-Sham DFT in the LDA or EXX as the independent-particle starting point [23, 56, 62]. Various flavours of self-consistency have been proposed. The conceptually simplest version is full self-consistency within the GW/RPA scheme, i.e., employing  $G_{GW}$  from the Dyson equation (Equation 2.46) for the construction of the polarisability and the self-energy in the next iteration. This self-consistency scheme was applied for the homogeneous electron gas [63], closed-shell atoms [64], and silicon [65]. However, it was observed that the agreement with experiment for the quasiparticle spectrum does rather worsen than improve. It was suggested that this might be attributed to an inconsistent treatment of higher-order diagrams. Iterating the  $GW$  equations implicitly introduces higher-order diagrams in the Green’s function, which are believed to be cancelled to a large degree by the vertex [64]. In other words, the independent quasiparticle picture may be good for describing the single quasiparticle excitations of the system, it is worse for the two-particle random-phase polarisation. This is in line with findings for the dielectric function computed in the RPA. Computing it from the Green’s function obtained from the  $G_0W_0$  self-energy is usually not better than employing the non-interacting  $G_0$  from the LDA [62]. Instead, explicit particle-hole interactions must be included, e.g. by the Bethe-Salpeter equation [62]. In approximate self-consistent schemes, the self-consistency is imposed only for the construction of  $\Sigma$  or for the quasiparticle energies [66–69]. Since these schemes give different results and the issue of self-consistency is still controversial, we do not go beyond the  $G_0W_0$  approximation.

### 2.3.3 The quasiparticle equation

For the interpretation of PES or IPES in terms of single quasiparticle excitations, the quasiparticle poles of the Green’s function need to be identified. These are obtained as the zeroes of the inverse Green’s function, which is related to  $G_0$  via the the Dyson equation (2.46):

$$G^{-1}(\mathbf{x}, \mathbf{x}', \omega) = G_0^{-1}(\mathbf{x}, \mathbf{x}', \omega) - \Sigma(\mathbf{x}, \mathbf{x}', \omega). \quad (2.52)$$

Since the non-interacting Green's function is the inverse of  $\omega - H_0$ , the (right-hand) quasiparticle wavefunctions corresponding to the quasiparticle pole  $\omega = \epsilon^{\text{qp}}$  are solutions to

$$\left\{ -\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{x}) + V_{\text{H}}(\mathbf{x}) \right\} \phi^{\text{qp}}(\mathbf{x}) + \int d\mathbf{x}' \Sigma(\mathbf{x}, \mathbf{x}', \epsilon^{\text{qp}}) \phi^{\text{qp}}(\mathbf{x}') = \epsilon^{\text{qp}} \phi^{\text{qp}}(\mathbf{x}). \quad (2.53)$$

We note a few things here. Since the self-energy is not Hermitean, the left and right eigenfunctions for a given eigenvalue are different and the eigenvalues are in general complex. The left (or right) eigenfunctions do not form an orthogonal set. Furthermore, the quasiparticle energy  $\epsilon^{\text{qp}}$  appears as the argument of the self-energy and on the right-hand side, i.e. the equation must be solved iteratively.

Eq. 2.53 is reminiscent of the single-particle equations from KS-DFT, the difference being that the local exchange-correlation potential has been replaced by the non-local self-energy. To solve Eq. 2.53, the quasiparticle wavefunctions  $\phi^{\text{qp}}$  are expanded in terms of the KS wavefunctions  $\phi_n^{\text{DFT}}$ , which form a complete orthonormal basis set. By multiplying with  $(\phi_n^{\text{DFT}}(\mathbf{x}))^*$  and integrating over  $\mathbf{x}$ , and then exploiting that the term in curly brackets of Eq. 2.53 can be written as  $H^{KS} - V_{\text{xc}}$ , we arrive at the algebraic equation (in Dirac notation)

$$\sum_{n'} \left( \epsilon_n \delta_{nn'} + \langle \phi_n^{\text{DFT}} | \Sigma(\epsilon^{\text{qp}}) - V_{\text{xc}} | \phi_{n'}^{\text{DFT}} \rangle \right) \langle \phi_{n'}^{\text{DFT}} | \phi^{\text{qp}} \rangle = \epsilon^{\text{qp}} \langle \phi_n^{\text{DFT}} | \phi^{\text{qp}} \rangle, \quad (2.54)$$

where we have inserted  $1 = \sum_{n'} |\phi_{n'}^{\text{DFT}}\rangle \langle \phi_{n'}^{\text{DFT}}|$ .

In practice, the matrix of the operator  $\Sigma(\epsilon^{\text{qp}}) - V_{\text{xc}}$  is usually dominated by the diagonal elements  $n = n'$  [58, 59, 70], which we also found for the systems of this work. Neglecting the small non-diagonal elements yields the quasiparticle energy equation

$$\epsilon_n^{\text{qp}} = \epsilon_n^{\text{DFT}} + \langle \phi_n^{\text{DFT}} | \Sigma(\epsilon_n^{\text{qp}}) - V_{\text{xc}} | \phi_n^{\text{DFT}} \rangle, \quad (2.55)$$

where the second term on the right hand side defines the quasiparticle correction. This result corresponds to applying first-order perturbation theory for the perturbation  $\Sigma(\epsilon^{\text{qp}}) - V_{\text{xc}}$  (note the quasiparticle energy in the argument of  $\Sigma$ ). The quasiparticle correction still contains the quasiparticle energy, i.e., Eq. 2.55 must be solved iteratively.<sup>15</sup>

The derivation above does not require  $V_{\text{xc}}$  to be exact. The “quasiparticle corrections” therefore comprise not only the self-energy (or quasiparticle)

<sup>15</sup>An alternative is to expand the energy-dependence of the self-energy matrix element in a Taylor series around  $\epsilon_n^{\text{DFT}}$ . A linear expansion is usually sufficient and leads to closed

effects absent from the exact KS band structure, but also correct for deficiencies of the exchange-correlation functionals used in practice, notably the self-interaction. It is likely that both effects contribute significantly to the total correction, but in general, they cannot be disentangled easily. The term “quasiparticle correction”, though well established, should therefore be used with care.

When important physical effects are incorrectly described by the chosen density functional, large differences between the DFT-KS and the quasiparticle wavefunctions may arise. This is in particular the case for image states far away from the surface, where DFT-LDA yields a qualitatively wrong potential [30, 71]. The transition from Eq. 2.53 to Eq.2.55 would then introduce significant errors. In such cases, the quasiparticle equation (Eq. 2.54) should be diagonalised. Whether such a diagonalisation is necessary or not can be decided by inspecting the magnitude of the off-diagonal elements of the perturbation operator  $\Sigma - V_{xc}$ . For all the  $GW$  calculations in this work, the off-diagonal elements were therefore computed, too, and found to be negligible in most cases. The diagonalisation may also be important for properties that depend on the quasiparticle wavefunctions. This has been highlighted for the reflectance anisotropy spectrum of GaAs(110) which changes significantly when one employs the quasiparticle wavefunctions rather than the KS wavefunctions although the error in the quasiparticle energies introduced by the perturbation approach is less than 0.1 eV [72].

### 2.3.4 $GW$ implementation: the space-time method

The  $GW$  space-time method exploits the fact that many of the transformations in a  $GW$  calculation can be efficiently performed in either real space and imaginary time or in reciprocal-space and imaginary frequency [73, 74]. The transformations between real space and reciprocal space can be efficiently computed with Fast Fourier Transforms, whereas the time-frequency Fourier transforms are performed via analytically enhanced Gauss-Legendre integrations [75].

An important practical aspect of the space-time approach is the separation of the  $G_0W_0$  self-energy into a static exchange

$$\Sigma_x(\mathbf{r}, \mathbf{r}') = G_0(\mathbf{r}, \mathbf{r}', 0^+)v(\mathbf{r} - \mathbf{r}') \quad (2.56)$$

---

expression

$$\epsilon_n^{\text{QP}} = \epsilon_n^{\text{DFT}} + \frac{\langle \phi_n | \Sigma(\epsilon_n^{\text{DFT}}) - V_{xc} | \phi_n \rangle}{1 - \frac{\partial}{\partial \omega} \langle \phi_n | \Sigma(\omega) | \phi_n \rangle \Big|_{\omega=\epsilon_n^{\text{DFT}}}} .$$

and a dynamic correlation part

$$\Sigma_c(\mathbf{r}, \mathbf{r}', t) = G_0(\mathbf{r}, \mathbf{r}', t)(W_0(\mathbf{r}, \mathbf{r}', t) - v(\mathbf{r} - \mathbf{r}')\delta(t)) . \quad (2.57)$$

In the following, we employ the symbols  $G$ ,  $P$ ,  $\varepsilon$ , and  $W$  to denote the non-self-consistent Kohn-Sham Green's function, polarisation, dielectric matrix, and screened interaction, respectively, in order to improve the readability. Except for the construction of the Green's function, all formulae remain valid for the self-consistent case. If we assume a non-magnetic systems for simplicity (the extension to a spin-dependent Green's function is straight-forward), the computational steps to construct the self-energy matrix elements from the output of a preceding DFT calculation are:

1. Construction of the non-interacting Green's function  $G$  in real space and imaginary time from the Kohn-Sham eigenfunctions  $\varphi_{n\mathbf{k}}$  and eigenvalues  $\epsilon_{n\mathbf{k}}$  (the Fermi level defines the energy zero)

$$G(\mathbf{r}, \mathbf{r}'; i\tau) = i \frac{\Omega}{(2\pi)^3} \int_{\text{BZ}} d^3k \begin{cases} \sum_n^{\text{occ}} \varphi_{n\mathbf{k}}(\mathbf{r}) \varphi_{n\mathbf{k}}^*(\mathbf{r}') e^{-\epsilon_{n\mathbf{k}}\tau}, & \tau < 0, \\ - \sum_n^{\text{unocc}} \varphi_{n\mathbf{k}}(\mathbf{r}) \varphi_{n\mathbf{k}}^*(\mathbf{r}') e^{-\epsilon_{n\mathbf{k}}\tau}, & \tau > 0, \end{cases} \quad (2.58)$$

where  $\Omega$  denotes the unit-cell volume and the integral over  $\mathbf{k}$  runs over the first Brillouin zone,

2. formation of the irreducible polarisability  $P$  in the random-phase approximation in real space and imaginary time

$$P(\mathbf{r}, \mathbf{r}'; i\tau) = -2iG(\mathbf{r}, \mathbf{r}'; i\tau)G(\mathbf{r}', \mathbf{r}; -i\tau) , \quad (2.59)$$

3. Fourier transformation of  $P$  to reciprocal space

$$P_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\tau) = \frac{1}{\Omega} \int d^3r \int d^3r' P(\mathbf{r}, \mathbf{r}'; i\tau) e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}+i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'} \quad (2.60)$$

and to imaginary frequency,

4. construction of the symmetrised dielectric matrix in reciprocal space

$$\tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\omega) = \delta_{\mathbf{G}\mathbf{G}'} - \frac{4\pi}{|\mathbf{k} + \mathbf{G}||\mathbf{k} + \mathbf{G}'|} P_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\omega) , \quad (2.61)$$

5. inversion of the symmetrised dielectric matrix for each  $\mathbf{k}$ -point and each imaginary frequency,

6. subtraction of that part of the dielectric matrix that gives rise to the long-range interaction

$$\tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1,\text{sr}}(\mathbf{k}, i\omega) = \tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{k}, i\omega) - \frac{|\mathbf{k} + \mathbf{G}|^2}{(\mathbf{k} + \mathbf{G})^T \mathbf{L}(i\omega) (\mathbf{k} + \mathbf{G})} \delta_{\mathbf{G}\mathbf{G}'}, \quad (2.62)$$

where  $\mathbf{L}(i\omega)$  denotes the macroscopic dielectric tensor, which is computed at  $\mathbf{k} = \mathbf{0}$ ,

7. calculation of the short-range part of the screened Coulomb interaction in reciprocal space

$$W_{\mathbf{G}\mathbf{G}'}^{\text{sr}}(\mathbf{k}, i\omega) = \frac{4\pi}{|\mathbf{k} + \mathbf{G}| |\mathbf{k} + \mathbf{G}'|} \tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1,\text{sr}}(\mathbf{k}, i\omega), \quad (2.63)$$

8. Fourier transformation of  $W^{\text{sr}}$  to imaginary time and to real space

$$W^{\text{sr}}(\mathbf{r}, \mathbf{r}'; i\tau) = \frac{1}{(2\pi)^3} \int_{\text{BZ}} d^3k \sum_{\mathbf{G}, \mathbf{G}'} W_{\mathbf{G}\mathbf{G}'}^{\text{sr}}(\mathbf{k}, i\tau) e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r} - i(\mathbf{k} + \mathbf{G}') \cdot \mathbf{r}'}, \quad (2.64)$$

9. construction of the long-range screening part  $W^{\text{hrs}} = W^{\text{lr}} - v$  in real space and imaginary time (see also Section 3.2.3),
10. construction of the screening part  $W^{\text{s}} = W - v$  of the screened interaction

$$W^{\text{s}}(\mathbf{r}, \mathbf{r}'; i\tau) = W^{\text{sr}}(\mathbf{r}, \mathbf{r}'; i\tau) + W^{\text{hrs}}(\mathbf{r}, \mathbf{r}'; i\tau) \quad (2.65)$$

11. formation of the correlation self-energy in real space and imaginary time

$$\Sigma_{\text{c}}(\mathbf{r}, \mathbf{r}'; i\tau) = iG(\mathbf{r}, \mathbf{r}'; i\tau) W^{\text{s}}(\mathbf{r}, \mathbf{r}'; i\tau), \quad (2.66)$$

12. computation of the matrix elements of the correlation self-energy

$$\langle \varphi_{n\mathbf{k}} | \Sigma_{\text{c}}(i\tau) | \varphi_{n\mathbf{k}} \rangle = \int d^3\mathbf{r} \int d^3\mathbf{r}' \varphi_{n\mathbf{k}}^*(\mathbf{r}) \Sigma_{\text{c}}(\mathbf{r}, \mathbf{r}'; i\tau) \varphi_{n\mathbf{k}}(\mathbf{r}'), \quad (2.67)$$

13. Fourier transformation of  $\langle \varphi_{n\mathbf{k}} | \Sigma_{\text{c}}(i\tau) | \varphi_{n\mathbf{k}} \rangle$  to imaginary frequency.

The matrix elements are then analytically continued to the real frequency axis by fitting a multi-pole function on the imaginary frequency axis [74]. The matrix element of the static exchange self-energy  $\Sigma_x = Gv$  are obtained separately by constructing the Green's function at  $i\tau = 0^+$  analogous to

Equation 2.58, forming the exchange self energy analogous to Equation 2.66, and computing the matrix elements analogous to Equation 2.67. Finally, the quasiparticle energies  $\epsilon_{n\mathbf{k}}^{\text{qp}}$  are given by the solution of

$$\epsilon_{n\mathbf{k}}^{\text{qp}} = \epsilon_{n\mathbf{k}} + \langle \varphi_{n\mathbf{k}} | \Sigma_c(\epsilon_{n\mathbf{k}}^{\text{qp}}) + \Sigma_x - V_{\text{xc}} | \varphi_{n\mathbf{k}} \rangle, \quad (2.68)$$

where  $V_{\text{xc}}$  is the exchange-correlation potential used in the underlying DFT calculation.

We note that we have improved the numerical implementation of the space-time method, notably the computation of the Green's function (Eq. 2.58), the inversion of the dielectric matrices (Step 5), and the computation of the matrix elements (Eq. 2.67). While leaving the advantageous scaling behaviour of the space-time method unchanged, the modifications greatly improve the numerical efficiency and reduce the overall run-time by a factor 3–5. Further modifications aimed at reducing the main memory and disk space requirements. These modifications are described in Appendix E and have been necessary to make the calculations in this thesis feasible.

# Chapter 3

## *GW* for surfaces and thin films

Plane-wave implementations of the DFT and *GW* schemes introduced in the previous chapter naturally assume three-dimensional periodicity. In this chapter we will describe how they can be used to model two-dimensional systems like thin films and surfaces by means of the “repeated-slab approach”. We will address briefly how it performs for DFT ground-state calculations before coming to the main point of this chapter, the repeated-slab approach in *GW* calculations. We will first show that the assumption of isotropic screening leads to a numerical instability in *GW* calculations and how this can be overcome. We will then demonstrate that the long-range screening leads to a particular, slow convergence behaviour for the  $\mathbf{k}$ -point sampling, which turns out to be one of the most critical parameters. The screening effects in a repeated slab system can be understood with a simple dielectric model. From this model, a robust correction scheme is derived that makes efficient *GW* calculations using the repeated-slab approach possible.

### 3.1 Repeated-slab approach

The use of plane-waves as basis set for DFT-KS and *GW* calculations implies a periodic repetition in all three spatial directions. This is naturally fulfilled for bulk crystals, but in two-dimensional systems like thin films or surfaces the periodicity is broken in one direction. To overcome this difficulty, a structural model for the surface is used that is compatible with three-dimensional periodic boundary conditions. In a first step, the semi-infinite substrate below the surface or the supported thin film is replaced by a finite slab. The thickness of this slab is an important (structural) convergence parameter. Then an artificial periodicity is introduced in the direction of the broken symmetry, i.e. perpendicular to the surface. This is schematically depicted



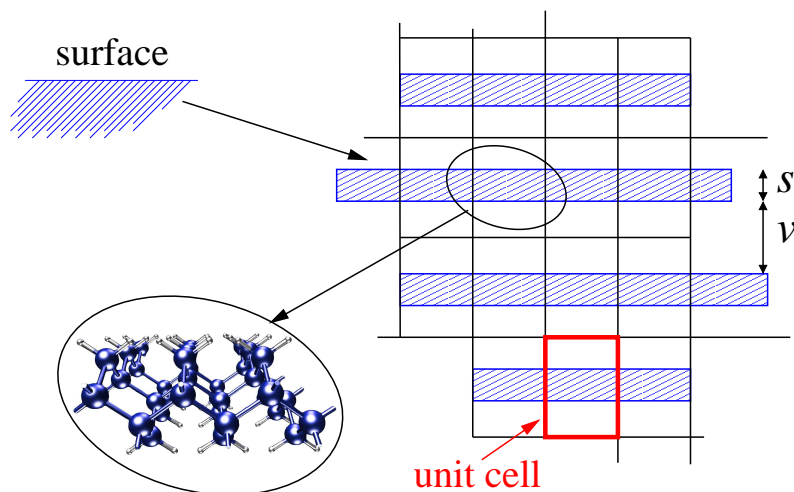


Figure 3.1: The repeated slab approach to describe surfaces.

in Fig. 3.1. The empty region (vacuum) between the individual slabs must thereby be large enough to avoid interactions between the slabs. Since the repeated-slab approach is a structural model, it is conceptually independent of the computational method employed and should work equally well for both DFT-KS and  $GW$  calculations. However, we will show that the convergence with respect to the basic structural parameters, namely the slab thickness  $s$  and the vacuum separation  $v$ , depends critically on the underlying physics of each method.

With density functional theory, we seek to determine the electronic ground-state of a neutral<sup>1</sup> system. The slab thickness controls the convergence to the limit of a semi-infinite system, i.e. it is relevant for the description of the substrate for surfaces or supported thin films. It must be thick enough to prevent an interaction of the surfaces across the slab, which depends strongly on the material and the type of the relevant interactions. For metals or other systems with delocalised, free-electron-like electron states, quantisation effects may become important and require large thicknesses, e.g. more than 20 layers for the Al(111) surface [76]. The total energy differences involved are often small ( $< 0.1$  eV), but the corresponding band structures show larger, sometimes even qualitative changes such as the curvature of the surface resonance at the  $\Gamma$  point for the Si(100)  $p(2 \times 1)$  surface [24]. Other important effects include the charge transfer between nonequivalent surfaces, the coupling of surface resonances of opposite surfaces of the slab, or strain. The

<sup>1</sup>Formally charged periodic systems always include a neutralising background charge density.

convergence behaviour for the properties of interest must therefore be tested in each case.

We now come to the coupling between the slabs through the vacuum. As the density decays exponentially in the vacuum region, the direct overlap between the densities of the slabs quickly becomes negligible with increasing separation. Long-range interactions between the slabs can then result only from electrostatic or dispersion interactions. The latter are however not contained in the commonly employed functionals, in particular the LDA. The decay behaviour of the electrostatic interactions along the direction of the broken symmetry  $z$  depends on the lowest non-zero multipole moment. The 0th-order moment, i.e. the total charge, vanishes for a neutral system.<sup>2</sup> A non-vanishing 1st-order (dipole) moment requires a “dipole correction” to suppress the intrinsic  $z^{-1}$  behaviour [78]. The interaction energy of the 2nd-order (quadrupole) moments decays as  $z^{-3}$ , which is sufficient for achieving a fast convergence to the limit of an isolated slab in practice. In summary, the decoupling of uncharged slabs is usually no problematic issue in DFT.

For  $GW$ , where we deal with charged excitations and include dynamical polarisation effects, the situation is less favourable. As will be shown next, the long-range tail of the screened interaction must be treated carefully to avoid spurious interactions between quasiparticles and their periodic images. In addition, charged quasiparticles polarise the neighbouring slabs which results in a slowly converging dependence of the quasiparticle energies on the vacuum separation between the slabs. We will show that these long-range effects can be modelled with the classic theory of dielectric screening.

## 3.2 The long-range tail of $W$

In this Section the treatment of long-range interaction in  $GW$  calculations is discussed. All actual  $GW$  implementations involve some approximations for this treatment, the validity of which depends on the physics of the system. In particular, long-range screening in repeated-slab system is inherently anisotropic. It is therefore necessary to review the treatment of the screened interaction for periodic systems in  $GW$ . We will do this rather generally and compare to other  $GW$  algorithms, too, to show that this issue is intrinsic to  $GW$  applied to periodic systems rather than being related to the space-time method or the repeated-slab approach.

---

<sup>2</sup>We note here that a constant neutralising background charge density for formally charged slab systems varies with the vacuum thickness and introduces again a slow convergence of the total energy and necessitates corresponding corrections [77].

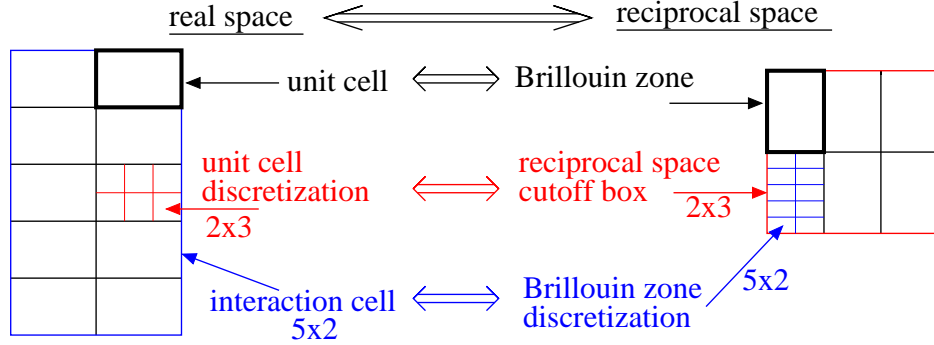


Figure 3.2: Connection between the real and reciprocal space grids. The inner grid in one space corresponds to the outer grid in the other space.

### 3.2.1 $\mathbf{k}$ -point sampling and interaction cell

As an introductory remark, we note that in periodic systems, the two-point functions  $F = G, W$ , or  $\Sigma$  reflect the lattice periodicity

$$F(\mathbf{r}, \mathbf{r}') = F(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R}) . \quad (3.1)$$

We can then introduce the following Wannier-like representation

$$F_{\mathbf{R}}(\mathbf{r}, \mathbf{r}') := F(\mathbf{r} + \mathbf{R}, \mathbf{r}') , \quad (3.2)$$

where  $\mathbf{r}$  and  $\mathbf{r}'$  are restricted to the unit cell. The corresponding Bloch-like representation  $F_{\mathbf{k}}$  (also denoted as mixed-space representation [74]) is obtained from a Fourier transformation

$$F_{\mathbf{k}}(\mathbf{r}, \mathbf{r}') := e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \sum_{\mathbf{R}} F_{\mathbf{R}}(\mathbf{r}, \mathbf{r}') e^{-i\mathbf{k}\cdot\mathbf{R}} . \quad (3.3)$$

The reciprocal-space representation is obtained from  $F_{\mathbf{k}}$  by Fourier transforming also  $\mathbf{r}$  and  $\mathbf{r}'$ , in full analogy to the plane-wave representation for the lattice-periodic part of a Bloch wavefunction.<sup>3</sup>

$\mathbf{k}$  is a continuous index, which is in practice discretised on a regular,  $\Gamma$ -centred grid ( $N_1 \times N_2 \times N_3$ ) as described in Section 2.2.3. The connection to the Wannier-representation provides us with a real-space picture of this approximation, cf. Fig. 3.2. The Brillouin zone discretisation grid  $\mathbf{b}_i/N_i$  is associated with a real space supercell  $\mathbf{a}_i \cdot N_i$ , i.e. it comprises  $N_1 \times N_2 \times N_3$  unit cells. This supercell coincides with the interaction cell of the space-time method, i.e. the range of non-locality for  $G, W$ , and  $\Sigma$  [73]. The

<sup>3</sup>Other implementations choose e.g. a Gaussian [79], LMTO [80], or LAPW [81] basis for the lattice periodic part.

discretisation of  $\mathbf{k}$  for the function  $F_{\mathbf{k}}$  is equivalent to imposing a translational symmetry

$$F_{\mathbf{R}}(\mathbf{r}, \mathbf{r}') = F_{\mathbf{R}+\mathbf{S}}(\mathbf{r}, \mathbf{r}') , \quad (3.4)$$

for its Wannier representation, where  $\mathbf{S}$  denotes a lattice vector of the interaction cell lattice ( $N_i \mathbf{a}_i$ ).

We now turn to the effect that the Brillouin zone discretisation has on the two-point functions in a  $GW$  calculation. For insulators, the Green's function  $G$  decays exponentially for a sufficient separation of its spatial arguments  $\mathbf{r}$  and  $\mathbf{r}'$  and is essentially zero outside the interaction cell. This behaviour is transferred directly to  $P = -iGG$  and  $\Sigma = iGW$ . For these functions, the Wannier and Bloch representations are numerically equivalent, but this is no longer the case for the slowly decaying bare and screened interaction as we will show next.

### 3.2.2 The computation of $W$ in reciprocal space

The computation of  $W$  from the polarisability  $P$  (Eq. 2.50–2.51) is performed in reciprocal space. This has several advantages:

1. The Coulomb potential is diagonal and the real-space convolution becomes a simple multiplication in reciprocal space.
2. The reciprocal-space representation for the lattice-periodic arguments ( $\mathbf{G}$  rather than  $\mathbf{r}$ ) is smaller than the corresponding real-space representation when a spherical cutoff is employed.
3. The construction and inversion of the dielectric matrix and the following construction of  $W$  can be performed for each  $\mathbf{k}$ -point separately.
4. The long-range behaviour  $\mathbf{R} \rightarrow \infty$  is mapped to a finite region in the vicinity of the  $\Gamma$ -point.

While the first two points are specific to plane-wave methods, the other two are independent of the representation of the lattice-periodic arguments. All  $GW$  methods with periodic boundary conditions employ the Bloch representation for the construction of  $W$  and our considerations apply to all of these. However, we restrict our discussion to plane-wave approaches.

Once  $W$  is obtained, the next step is the Fourier transformation to real space in the space-time method, or the reciprocal-space convolution to construct the self-energy matrix elements in reciprocal-space approaches [59, 79, 81]. Both algorithms formally involve an integration over the Brillouin zone, which is replaced by a finite summation over a discrete  $\mathbf{k}$ -point

grid. As pointed out above, the discretisation corresponds to a periodic repetition in real space, i.e. the outcome of this summation corresponds to a periodic array of quasiparticles instead of an isolated one. When no corrections are applied, the quasiparticle interacts with its periodic images. Since  $W$  decays very slowly in real space, the images cannot be simply decoupled by increasing the  $\mathbf{k}$ -point grid size and the computation becomes numerically unstable. This is most obvious for the  $1/r$ -part of the screened interaction that gives rise to a divergence of the average potential, which translates to the  $1/k^2$  singularity in reciprocal space. This singularity is integrable, i.e. the aforementioned Fourier and convolution integrals are well-defined, but cannot be simply replaced by summations in practice.

That the  $1/k^2$  singularity necessitates an explicit treatment has of course been recognised early on [58, 59]. The solution in reciprocal space plane-wave approaches has been to reintroduce the proper integral for the problematic  $\mathbf{G} = \mathbf{G}' = \mathbf{0}$  elements. To this end, a model function with the appropriate singularity, for which the integral can be computed analytically, is subtracted from  $W$ . The remainder is non-singular and can again be computed by finite summations. After subtraction of the singular function, the  $\mathbf{k} = \mathbf{G} = \mathbf{G}' = \mathbf{0}$  element must be set to zero to fix the average potential. The model function is chosen such that the integrals (or approximations to it) can be computed analytically, and a large variety of more or less general models has been proposed [59, 82–85]. In the original space-time method [74], the model function

$$W^{\text{lr,iso}}(\mathbf{k} + \mathbf{G}) = \frac{4\pi \bar{\epsilon}^{-1}}{|\mathbf{k} + \mathbf{G}|^2}, \quad (3.5)$$

where  $\bar{\epsilon}$  denotes the average macroscopic dielectric constant, was employed. We have developed a generalisation to treat anisotropic screening to arbitrary precision.<sup>4</sup> We will summarise the most important steps for the treatment of the  $1/r$  part in the following section but refer to a recent publication for a discussion of the technical details [86].

### 3.2.3 Anisotropy in the screened interaction

It can be shown that the head ( $\mathbf{G} = \mathbf{G}' = \mathbf{0}$ ) and the wings ( $\mathbf{G} = \mathbf{0}$  or  $\mathbf{G}' = \mathbf{0}$ ) of the dielectric matrix at the  $\Gamma$ -point, i.e., for  $\mathbf{k} = \mathbf{0}$ , depend on the direction in which the limit  $\mathbf{k} \rightarrow \mathbf{0}$  is taken [86, 87]. We denote this dependence by the spatial angle  $\Omega_{\mathbf{k}}$ , and the corresponding normalised direction vector by  $\hat{\mathbf{k}}$ . To compute this directional limit, a  $\mathbf{k} \cdot \mathbf{p}$  perturbation

<sup>4</sup>The treatment of anisotropic screening was jointly developed with Philipp Eggert and Arno Schindlmayr [24, 86].

ansatz is employed, leading to

$$\begin{aligned} \tilde{\varepsilon}_{\mathbf{0}\mathbf{0}}(\mathbf{k}, i\omega) \rightarrow & 1 + \frac{2}{\pi^2} \sum_{\alpha, \beta} \hat{k}_\alpha \hat{k}_\beta \sum_{v, c} \int_{\text{BZ}} d^3q \frac{\epsilon_{c\mathbf{q}} - \epsilon_{v\mathbf{q}}}{(\epsilon_{c\mathbf{q}} - \epsilon_{v\mathbf{q}})^2 + \omega^2} \\ & \times \langle \varphi_{v\mathbf{q}} | r_\alpha | \varphi_{c\mathbf{q}} \rangle \langle \varphi_{c\mathbf{q}} | r_\beta | \varphi_{v\mathbf{q}} \rangle \end{aligned} \quad (3.6)$$

and an analogous expression for the wings [86].  $\alpha$  and  $\beta$  are indices for the cartesian components of the vectors  $\mathbf{k}$  and  $\mathbf{r}$ .  $\epsilon_{n\mathbf{q}}$  and  $\varphi_{n\mathbf{q}}$  denote the Kohn-Sham energies and wavefunctions, respectively, and the sum over  $v$  ( $c$ ) runs over occupied (unoccupied) states. The matrix elements of the position operator  $\mathbf{r}$  are computed via the commutator with the Hamiltonian and must include the contribution of the non-local pseudopotential for accurate results. In App. E.4, we present an efficient scheme that we have developed for this purpose with an improved scaling behaviour compared to a previous approach [88].

For simplicity, the imaginary frequency argument  $i\omega$  is omitted in the following. The directional dependence at the  $\Gamma$ -point is present in the whole inverse dielectric matrix, i.e., head, wings, and body. After block-wise inversion [87] the head of the inverse symmetrised dielectric matrix takes the form

$$\tilde{\varepsilon}_{\mathbf{0}\mathbf{0}}^{-1}(\Omega_{\mathbf{k}}) = \frac{1}{\hat{\mathbf{k}}^T \mathbf{L} \hat{\mathbf{k}}}, \quad (3.7)$$

where the matrix  $\mathbf{L}$  is the macroscopic dielectric tensor. We note that in most other implementations where the anisotropy has been considered so far, such as [84, 85], but not Hott [83], the right-hand side of Equation (3.7) was replaced by the expression  $\hat{\mathbf{k}}^T \mathbf{L}^{-1} \hat{\mathbf{k}}$  without formal justification.

Combining Equation (3.7) with Equation (2.63), we obtain the head of the screened interaction for  $\mathbf{k} \rightarrow \mathbf{0}$

$$W_{\mathbf{0}\mathbf{0}}(\mathbf{k}) \rightarrow \frac{4\pi}{\hat{\mathbf{k}}^T \mathbf{L} \hat{\mathbf{k}}}. \quad (3.8)$$

In the space-time method the head of the inverse dielectric matrix is used to define the long-range part of the screened interaction. For this purpose, we extend Equation (3.8) to  $\mathbf{G} = \mathbf{G}' \neq \mathbf{0}$  and define the long-range part as

$$W_{\mathbf{G}\mathbf{G}'}^{\text{lr}}(\mathbf{k}) = \frac{4\pi}{(\mathbf{k} + \mathbf{G})^T \mathbf{L} (\mathbf{k} + \mathbf{G})} \delta_{\mathbf{G}\mathbf{G}'}. \quad (3.9)$$

The short-range part  $W^{\text{sr}} = W - W^{\text{lr}}$  can then safely be Fourier transformed to real space using Fast Fourier Transformations since it is no longer singular. For numerical reasons we subtract the long-range part at the level of the inverse dielectric matrix

$$\tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1, \text{sr}}(\mathbf{k}) := \tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{k}) - \frac{|\mathbf{k} + \mathbf{G}|^2}{(\mathbf{k} + \mathbf{G})^T \mathbf{L} (\mathbf{k} + \mathbf{G})} \delta_{\mathbf{G}\mathbf{G}'} \quad (3.10)$$

and compute  $W^{\text{sr}}$  from this modified entity according to

$$W_{\mathbf{G}\mathbf{G}'}^{\text{sr}}(\mathbf{k}) = \frac{4\pi}{|\mathbf{k} + \mathbf{G}||\mathbf{k} + \mathbf{G}'|} \tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}^{-1,\text{sr}}(\mathbf{k}). \quad (3.11)$$

By expanding the angular dependence of  $W^{\text{lr}}$  into spherical harmonics

$$W_{\mathbf{G}\mathbf{G}'}^{\text{lr}}(\mathbf{k}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l H_{lm} \frac{4\pi}{|\mathbf{k} + \mathbf{G}|^2} \delta_{\mathbf{G}\mathbf{G}'} Y_{lm}(\Omega_{\mathbf{k}+\mathbf{G}}), \quad (3.12)$$

we can perform the Fourier transformation of  $W^{\text{lr}}$  analytically. Only even  $l$  contribute to the sum because the coefficients  $H_{lm}$  vanish for odd  $l$ . The non-vanishing coefficients are obtained by computing the integrals

$$H_{lm} = \int d\Omega_{\mathbf{k}} Y_{lm}^*(\Omega_{\mathbf{k}}) \frac{1}{\hat{\mathbf{k}}^T \mathbf{L} \hat{\mathbf{k}}} \quad (3.13)$$

numerically on a Lebedev-Laikov angular grid [89]. Making use of the expansion of a plane wave [90] in spherical harmonics  $Y_{lm}$  and spherical Bessel functions  $j_l$ ,

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l i^l j_l(kr) Y_{lm}(\Omega_{\mathbf{r}}) Y_{lm}^*(\Omega_{\mathbf{k}}) \quad (3.14)$$

we arrive at

$$W^{\text{lr}}(\mathbf{r}, \mathbf{r}') = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_l i^l H_{lm} Y_{lm}(\Omega_{\mathbf{r}-\mathbf{r}'} ) \frac{1}{|\mathbf{r} - \mathbf{r}'|}. \quad (3.15)$$

The coefficients  $c_l$  for even  $l$  are defined as

$$c_l = \frac{2}{\pi} \int_0^{\infty} dx j_l(x) = \frac{(l-1)!!}{l!!} \quad (3.16)$$

with  $n!! = n(n-2)(n-4)\dots$ . In practice we truncate the sum in Equation (3.15) at finite  $l = l_{\text{max}} = 4$  as discussed in [86].

For numerical convenience the self energy  $\Sigma$  is split into a static exchange part  $\Sigma_x = iGv$  and a frequency-dependent correlation part  $\Sigma_c = iG(W - v)$  in the space-time method [74]. We achieve this by subtracting the bare Coulomb interaction  $v$  from  $W^{\text{lr}}$  in its angular expansion (3.12), i.e., we subtract  $1/\sqrt{4\pi}$  from  $H_{00}$  for each imaginary frequency. Furthermore, the transformation from imaginary frequency to imaginary time is then performed on the expansion coefficients  $H_{lm}(i\omega)$  directly, and we obtain  $(W^{\text{lr}} - v)$  according to Equation (3.15) with the expansion coefficients in imaginary time  $H_{lm}(i\tau)$ .

A proper treatment of the anisotropy in the long-range part of the screened interaction is crucial to obtain converged results. This is easily illustrated in the space-time method: the density of the  $\mathbf{k}$ -point sampling determines the range of the non-locality in real space. If parts of the long-range interaction remain in  $W^{\text{sr}}$  for small but finite  $\mathbf{k}$ , then the tails of  $W^{\text{sr}}$  extend over the boundary of the interaction cell and will be folded back in the numerical Fourier transformation due to the periodic boundary conditions. Since the size of the interaction cell is determined by the  $\mathbf{k}$ -point sampling, an unsatisfactory  $\mathbf{k}$ -convergence behaviour results when the long-range part is treated inadequately.

We have tested this hypothesis with a hydrogen-saturated four layer Si(100) slab [86]. In Figure 3.3 we show the convergence of the quasiparticle energy of the lowest conduction state with respect to the number of  $\mathbf{k}$ -points in the direction perpendicular to the surface. Other states exhibit a similar behaviour. It is obvious that the original isotropic averaging for the screened interaction leads to an unphysical linear increase in the quasiparticle energy. In contrast, the anisotropic treatment converges rapidly. The reason for the linear increase in the isotropic treatment lies in the inadequate treatment of the singularity, which is not fully removed. Integrating  $1/|\mathbf{k}|^2$  numerically on a Cartesian grid yields for  $k_x = k_y = 0$  with  $\Delta k_z = k_{\text{max}}/N_z$

$$\sum_{n=1}^{N_z} \Delta k_z \frac{1}{(n\Delta k_z)^2} \longrightarrow \int_{\Delta k_z}^{k_{\text{max}}} dk \frac{1}{k^2} = \frac{1}{\Delta k_z} - \frac{1}{k_{\text{max}}} = \frac{N_z - 1}{k_{\text{max}}} \quad (3.17)$$

and hence a linearly diverging contribution, whose weight is proportional to  $\Delta k_x \Delta k_y \sim (N_x N_y)^{-1}$ . When the  $\mathbf{k}$ -sampling is increased in all three directions simultaneously, no such linear divergence occurs, but such a restriction is undesirable and inefficient in practice. Therefore, only the proper anisotropic treatment allows us to investigate the importance of the  $\mathbf{k}$ -point sampling in the direction perpendicular to the surface. To our knowledge, the convergence in the perpendicular direction has not been addressed in previous  $GW$  calculations for slab systems. We find an  $1/N_z$  behaviour the magnitude of which scales as  $1/(N_x N_y)$ . In the view of the large  $\mathbf{k}$ -point samplings required in the direction parallel to the surface as discussed below,  $N_z = 1$  proves to be sufficient in practice.

To summarise, the  $1/r$  long-range part of the screened interaction introduces a numerical instability, that can be avoided when appropriate model functions are employed that include the necessary physics. In the space-time approach, the polarisation cloud around each electron is separated into that of a perfectly homogeneous system with the correct average long-range tail and a short-ranged part that contains the microscopic deviations from the



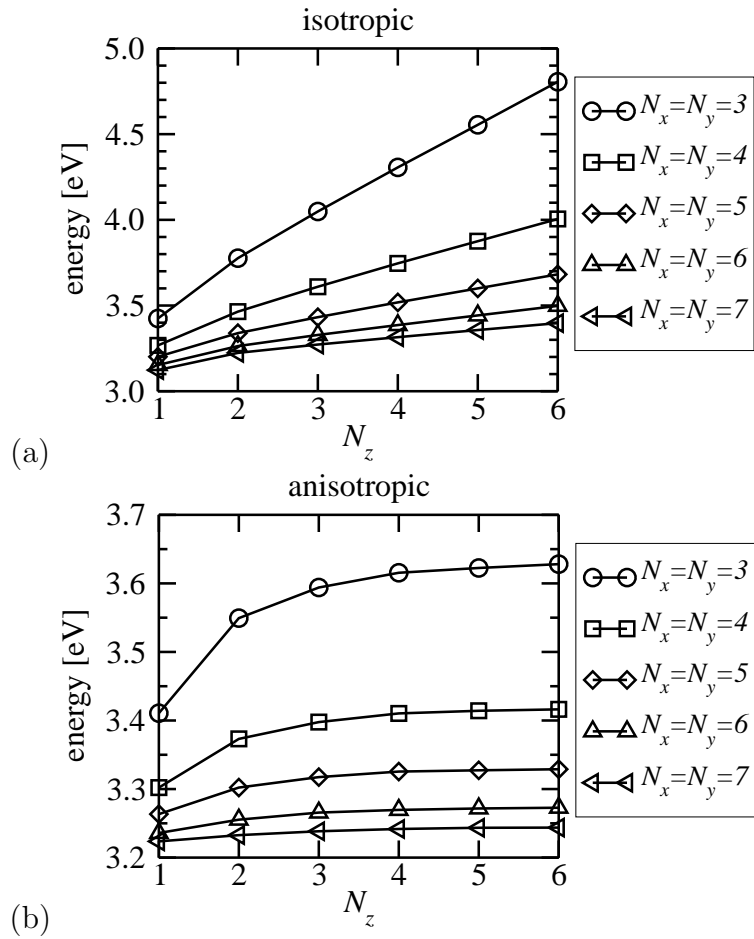


Figure 3.3: Convergence of the lowest conduction-band energy of a  $\text{Si}_4\text{H}_4$  slab with respect to the number of k-points  $N_z$  perpendicular to the surface for (a) the original isotropic implementation and (b) with the anisotropy taken into account. The quantitative behaviour depends on the sampling in the parallel direction ( $N_x = N_y$ ). Note the different scales of the two graphs.

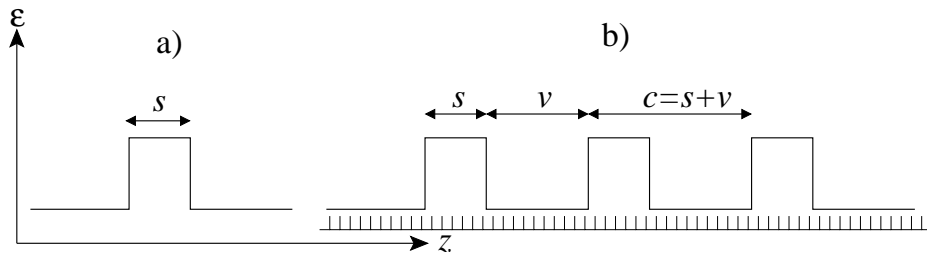


Figure 3.4: Dielectric profiles for a) an isolated slab b) repeated slabs. The discretisation of the  $z$ -axis is indicated for the repeated slabs.

average. For slabs, which are inherently anisotropic system, the anisotropy in the screening must be taken into account, and we have presented an efficient formalism for this in the context of the  $GW$  space-time method.

### 3.3 Screening in slab systems

In this section, the qualitative features of the screened interaction in isolated and repeated slab systems will be discussed. After describing a simplified dielectric model that reproduces the most important characteristics of the screening, we will discuss its implications for the  $\mathbf{k}$ -point sampling in the parallel direction and the role of the periodic repetition of the isolated slabs.

#### 3.3.1 Dielectric model

In order to learn more about the qualitative behaviour of the screened interaction in a slab system, a dielectric model system is considered in which the slab is modelled as a perfectly homogeneous dielectric medium (dielectric constant  $\varepsilon$ ) with thickness  $s$ . The periodic repetition perpendicular to the surface along the  $z$ -direction is simulated by a finite number of dielectric slabs, typically 10–20, with a vacuum separation  $v$ . We note that  $s$  and  $v$  are model parameters, but not independent since the total cell height

$$c = s + v \quad (3.18)$$

must agree with that of the full  $GW$  calculation. In addition to the repeated slabs, we will also discuss the isolated slab as the limiting case for  $v \rightarrow \infty$ . The dielectric profiles  $\varepsilon(z)$  for an isolated slab and a repeated slab system are depicted in Fig. 3.4.

For a practical computation of the screened interaction the  $z$ -coordinate is discretised, i.e. the slab and the vacuum is partitioned into layers. All layers

have the same thickness and are completely homogeneous. The screened potential is then computed by the method of image charges as described in Section B.1. To this end, we place a unit charge in the centre of a layer  $z'$  (an index that we will omit for simplicity in the following). The dielectric discontinuity at the interface to an adjacent layer introduces image charges in both layers. These charges will then induce new image charges due to the next interfaces. In the end, an infinite but discrete set of image charges results from which the screened potential  $W(z, \rho)$  in layer  $z$  at a lateral separation  $\rho$  from the original charge, is computed by<sup>5</sup>

$$W(z, \rho) = \sum_{d=-\infty}^{\infty} \frac{q(z, d)}{\sqrt{d^2 + \rho^2}} \quad (3.19)$$

where  $q(z, d)$  are image charges. They depend on the dielectric constants of all layers and the position  $z'$  of the original charge.

For the discussion below, we will focus mainly on the image potential, i.e.  $z = z'$ ,  $\rho = 0$  and excluding  $d = 0$  from the summation. In doing so, we ignore the interplay of quantum effects, which result from the non-locality of the Green's function, with the long-range screening for the quasiparticles. As we will see this is sufficient to develop a qualitative understanding of the additional screening effects in quasiparticle calculations that are introduced when going from a bulk system to a slab system. Two aspects of practical *GW* calculations are well reproduced by the dielectric model:

1. the convergence behaviour with respect to the parallel  $\mathbf{k}$ -point sampling,
2. the convergence behaviour with respect to the vacuum separation in the repeated-slab approach,

and we will discuss these points in the following.

### 3.3.2 Parallel $\mathbf{k}$ -point sampling

In Section 3.2.3 we have shown how to treat the  $1/r$ -part of the screened interaction by splitting off the macroscopic (average) behaviour. However,

---

<sup>5</sup>For clarity, we use a slightly different notation here than in Section B.1: here, we include the sign  $\sigma$  into the distance variable  $d$ , which can now assume negative values. Also, the dielectric constant of the layer  $z$  is absorbed in the image charges, i.e.

$$q(z, d) \hat{=} q(z, d, \sigma) / \varepsilon(z) .$$

there may be still interactions that exceed the size of the interaction cell dictated by the  $\mathbf{k}$ -point grid. In bulk systems, the deviations from the average arise mainly from the local variations in the polarisation cloud due to the atomic structure and they usually become negligible when the typical length-scale of these structural variations is exceeded. In a slab system, however, the decisive structural variation is the slab itself and we must expect considerable variations at the length scale of the slab thickness. Usually, the parallel  $\mathbf{k}$ -point sampling of a slab is chosen in analogy to that of the bulk, i.e. sufficiently large to average out the lateral variations in the structure. We can then use the model described above to understand the  $\mathbf{k}$ -point convergence parallel to the slab.

For this purpose we take two essential steps: first we identify the model analogue to the “short-range” part of the screened interaction. We then use the real-space interpretation of a discrete  $\mathbf{k}$ -point sampling, i.e. we take into account a periodic repetition along the  $\rho$ -direction parallel to the surface. It is important to clearly distinguish here between the *image charges* and the *periodic images* (cf. Fig. 3.5). The *image charges* are mathematical constructs to compute the screened interaction in our model and are distributed along the  $z$ -direction. The *periodic images* are distributed along the  $\rho$ -direction and simulate the effect of a discrete  $\mathbf{k}$ -sampling. Since the superposition principle applies, the periodically repeated screened interaction can be computed in our model from the periodic images of the image charges for a single quasiparticle. The periodic repetition in the parallel direction due to the finite  $\mathbf{k}$ -point sampling must also not be mistaken for the periodic repetition in the perpendicular direction in the repeated slab approach. The qualitative parallel behaviour is independent on whether the slabs are repeated or not. We will therefore discuss it for both isolated and repeated slabs.

We assume that in the full  $G_0W_0$  calculation the anisotropic  $1/r$  part is taken into account correctly by the long-range treatment of Section 3.2.3. We then have to ask the question how this can be translated to the dielectric model. On this point we note that the image charges  $q(z, d)$  become small for large  $|d|$  and can be neglected above a certain  $d_{\max}$ . For  $\rho \gg d_{\max}$ , we can neglect  $d$  in the denominator in Eq. 3.19 and obtain the long-range  $1/\rho$  behaviour

$$W^{\text{lr}}(z, \rho) = \frac{q_{\text{eff}}}{\rho} \quad q_{\text{eff}} = \sum_{d=-d_{\max}}^{d_{\max}} q(z, d). \quad (3.20)$$

The “short-range” part is then obtained as

$$W^{\text{sr}}(z, \rho) = W(z, \rho) - W^{\text{lr}}(z, \rho)$$

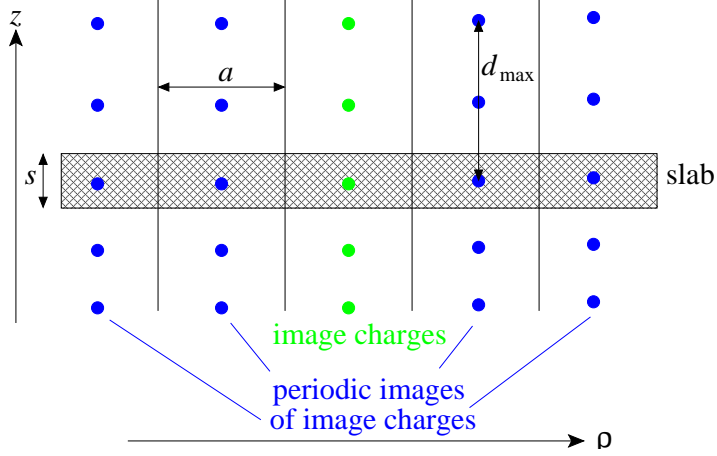


Figure 3.5: Periodic array of image charges to model the artificial periodicity introduced by a finite  $\mathbf{k}$ -sampling, shown schematically for an isolated slab.

$$\begin{aligned}
 &= \sum_{d=-d_{\max}}^{d_{\max}} q(z, d) \left( \frac{1}{\sqrt{d^2 + \rho^2}} - \frac{1}{\rho} \right) \\
 &= \sum_{d=-d_{\max}}^{d_{\max}} \frac{-q(z, d)d^2}{\rho(\rho + \sqrt{\rho^2 + d^2})\sqrt{\rho^2 + d^2}}. \quad (3.21)
 \end{aligned}$$

It is this part that remains in the numerical treatment and may not be as short-ranged as the name suggests. The long-range behaviour of  $W^{\text{sr}}$  is obtained again for  $\rho \gg d_{\max}$  and gives

$$\sum_{d=-d_{\max}}^{d_{\max}} \frac{-q(z, d)d^2}{2\rho^3}, \quad (3.22)$$

and hence a quadrupolar interaction. Since the average potential is set in the  $1/r$  treatment, any shift in the average potential from the quadrupoles is implicitly corrected for.

We will now investigate how the image potential changes when the "short-ranged" part (Eq. 3.21) exceeds the interaction cell for a certain  $\mathbf{k}$ -point sampling in the parallel direction. For this we combine the dielectric model with the idea that a finite  $\mathbf{k}$ -point sampling corresponds to a periodic repetition (cf. Fig. 3.5). We then have to sum the short-range part of the parallel images to obtain the "parallel repetition error" (PRE). For simplicity, we consider only square lattices in the parallel direction. The lattice constant  $a$  is related to the number of  $\mathbf{k}$ -points in the parallel direction  $N_{\mathbf{k}_{\parallel}}$  via

$$a = N_{\mathbf{k}_{\parallel}} \cdot a_{\parallel}, \quad (3.23)$$

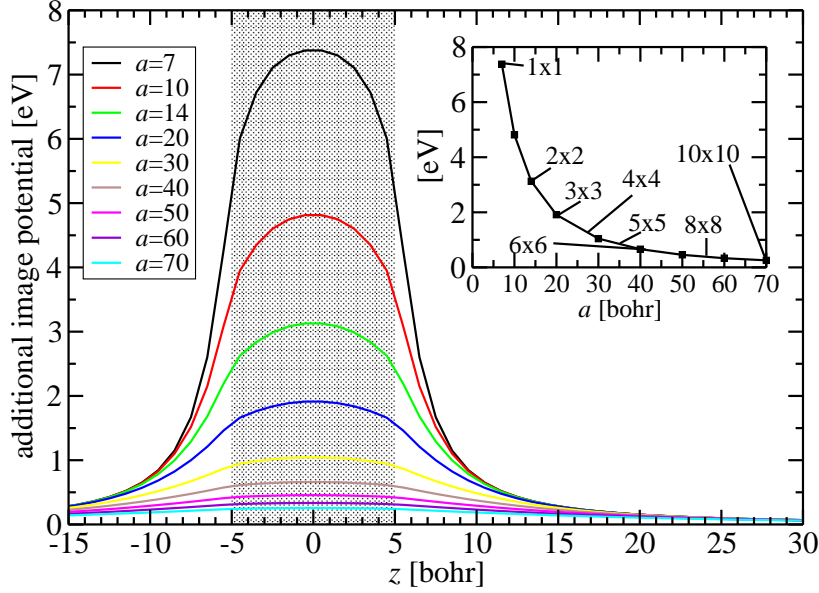


Figure 3.6: Change in the image potential due to a periodic repetition (square lattice, lattice constant  $a$ ) in the parallel direction. The dielectric model slab (10 bohr thick,  $\varepsilon = 2.3$ ) is indicated by the dotted region and corresponds to a 2-layer NaCl film. The inset shows the value at the centre of the slab as a function of  $a$ . The corresponding  $\mathbf{k}$ -sampling ( $a_{\parallel} \approx 7$  bohr) is also indicated.

where  $a_{\parallel}$  is the surface lattice constant of the real systems. The  $a$ -dependent change in the image potential from the periodic images then becomes

$$\begin{aligned} \Delta W(z) &= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} W^{\text{sr}}(z, \sqrt{n^2 + m^2} a) \\ &= \sum_{d=-d_{\text{max}}}^{d_{\text{max}}} q(z, d) \sum_{n,m=-\infty}^{+\infty} \left( \frac{1}{\sqrt{d^2 + (n^2 + m^2) a^2}} - \frac{1}{\sqrt{n^2 + m^2} a} \right) \end{aligned} \quad (3.24)$$

excluding  $n = m = 0$  from the summation. The sums over  $n$  and  $m$  are evaluated using the two-dimensional Ewald summation technique.

We first consider an isolated slab with a dielectric constant  $\varepsilon = 2.3$ , a typical value for NaCl. The influence of the  $\mathbf{k}$ -point sampling on the change in the image potential is investigated by varying the parallel lattice constant  $a$  in our model. In Fig. 3.6 it can be seen that the PRE for the image potential is broader than the slab thickness. For small values of  $a$ , the effect is very large. With increasing  $a$ , the additional image potential quickly becomes smaller and broader. The value at the centre of the slab is then indicative of the magnitude of the PRE within the slab. We will therefore focus on this

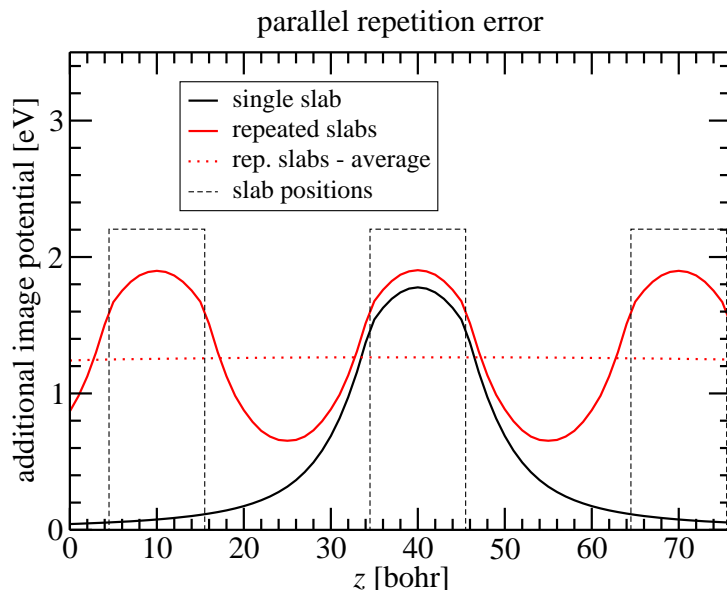


Figure 3.7: Changes in the image potential due to the PRE for an isolated slab and repeated slabs. The model parameters correspond to a 2-layer NaCl slab and a  $3 \times 3$   $\mathbf{k}$ -point sampling.

value for the discussing the  $a$ -dependence.

We now turn to the periodically repeated slabs. The variation in the image potential for an isolated and a periodically repeated slab is shown in Fig. 3.7. The position of the slabs is indicated by dashed lines. It must be emphasised that the average potential is corrected for during the treatment of the  $1/r$  part. This leads to an interesting compensation effect for repeated slabs. The deviation of the additional image potential from its average is much smaller than for an isolated slab.<sup>6</sup> Periodic repetition in the  $z$ -direction and small slab separations reduce the error from the parallel  $\mathbf{k}$ -point sampling. We can quantify this by varying the separation between the slabs, shown in Fig. 3.8. The vacuum separation is denoted by the total cell height (i.e. slab + vacuum) for consistency with realistic systems. Increasing the vacuum separation increases the error. The inset demonstrates that also the qualitative convergence behaviour is altered. For large vacuum

<sup>6</sup>The analogue to the average for an isolated slab is the limit  $z \rightarrow \infty$ , which is zero. Only when the isolated slab is placed in a finite cell, an average can be computed. If it is corrected for by setting the  $\mathbf{G} = \mathbf{G}' = \mathbf{k} = \mathbf{0}$  element of  $W^{\text{sr}}$  to zero, the average compensation reduces the PRE for the isolated case, too, but will depend on the cell size. A cell-size dependent parallel convergence has been indeed observed for  $GW$  calculations where the Coulomb potential has been truncated to decouple neighbouring slabs [25].

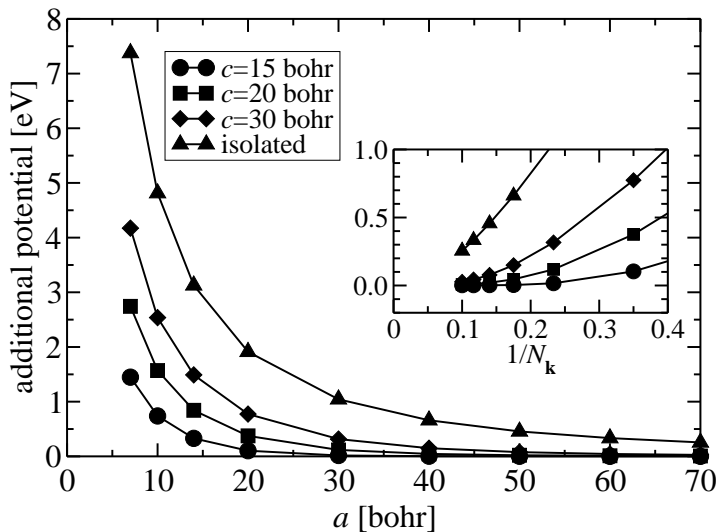


Figure 3.8: Dependence of the PRE at the slab centre on the interaction cell size  $a$  for periodically repeated slabs (10 bohr thick,  $\varepsilon = 2.3$ ) with different total cell heights. The isolated slab values are shown for comparison. The inset shows the data as a function of  $1/N_{\mathbf{k}}$  for a NaCl slab ( $a_{\parallel} = 7$  bohr).

separations or isolated slabs, the error is approximately linear in  $1/N_{\mathbf{k}}$  for small values of  $N_{\mathbf{k}}$ . This dependence becomes increasingly curved when the vacuum separation is decreased, thereby lowering the required  $\mathbf{k}$ -point sampling to achieve a certain accuracy. This is in agreement with Eq. 3.21 that shows a quadrupolar  $1/r^3$  behaviour for very large separations, and a  $1/r$  for shorter distances. We therefore expect a  $1/N_{\mathbf{k}}$  dependence for small  $N_{\mathbf{k}}$  and a  $1/N_{\mathbf{k}}^3$  for large  $N_{\mathbf{k}}$  as the limiting cases.

These qualitative changes have a simple physical reason. The screening in a slab system, which gives rise to the observed behaviour, depends on the length scale. At lateral separations much larger than the slab thickness, we observe an average anisotropic screening with an average dielectric constant  $\varepsilon_{\parallel}$  for the parallel direction. This is treated correctly, and the remainder is dominated by the quadrupolar part decaying as  $1/r^3$ . At lateral separations much smaller than the slab thickness<sup>7</sup> the screening inside the slab is essentially that of the bulk material. The PRE then reflects the difference  $(1/\varepsilon - 1/\varepsilon_{\parallel})/r$ . In the repeated slab approach,  $\varepsilon_{\parallel}$  is the weighted average of the dielectric constant of the bulk and of the vacuum, i.e. the difference between  $\varepsilon$  and  $\varepsilon_{\parallel}$  increases with the vacuum. Small vacuum separations therefore reduce the  $1/r$ -part of the PRE. This prediction is fully confirmed

<sup>7</sup>or more precisely: the distance from the closest interface



by the convergence behaviour of the quasiparticle energies in repeated slab systems with different amounts of vacuum (e.g. Fig. 3.9).

Unfortunately, the simple models are not sufficient to compute reliable quantitative corrections for the PRE introduced by the finite  $\mathbf{k}$ -sampling. On the one hand, the corrections depend rather sensitively on the model parameters for the dielectric slab, particularly its thickness, and straightforward estimates of the model slab thickness from the dielectric tensor of the repeated-slab system tend to yield overestimated PREs. On the other hand, the additional potential varies over the slab and the COHSEX approach described in Section B.2 may not reduce to a scissors correction.

We can however derive a useful expression for the qualitative convergence behaviour of the quasiparticle energies with respect to the  $\mathbf{k}$ -point sampling. The error in the quasiparticle energy is likely to be proportional to the error in the screened interaction. Assuming that only the closest lateral images play an important role, which is justified by the fast decay of  $W^{\text{sr}}(\rho)$ , the main effect can be described by a single term in the sum of Eq. 3.24. The proposed fitting function with parameters  $Q$  and  $Z$  is therefore

$$\Delta E(N_{\mathbf{k}}) = \frac{QZ^2/N_{\mathbf{k}}^3}{1 + \sqrt{1 + Z^2/N_{\mathbf{k}}^2 + Z^2/N_{\mathbf{k}}^2}} \quad (3.25)$$

and turns out to be a reasonable choice. It not only smoothly interpolates between a  $1/N_{\mathbf{k}}$  behaviour for small  $N_{\mathbf{k}}$  and a  $1/N_{\mathbf{k}}^3$  behaviour for large  $N_{\mathbf{k}}$  as the limiting cases, but also serves as a robust extrapolation function.

We demonstrate this for NaCl slabs of varying thickness with a  $N_{\mathbf{k}} \times N_{\mathbf{k}} \times 1$  sampling in Fig. 3.9. For the 2-layer slab, a large number of  $\mathbf{k}$ -samplings was tested. The good agreement proves that the curvature of the fit function describes the numerical data very well. For thicker slabs there is additional numerical noise, which is however below the expected accuracy of our numerical calculation (0.05 eV). We note also that the vacuum thickness is approximately the same for all slabs. The improvement of the convergence behaviour for thick slabs is therefore exclusively due to the average compensation effects described in this Section. For isolated slabs, the opposite behaviour is expected: thicker slabs should show larger PREs than thin slabs since the closest image charges define the length scale  $Z$  for the lateral convergence in Eq. 3.25. This hypothesis can be tested within the dielectric model and is fully confirmed (Fig. 3.10).

It must be emphasised here that the  $\mathbf{k}$ -point sampling has an influence on more than just the long-range part of the screened interaction. Therefore, also other effects contribute to the observed convergence behaviour of the quasiparticle gap. In particular, we observed that the overall convergence

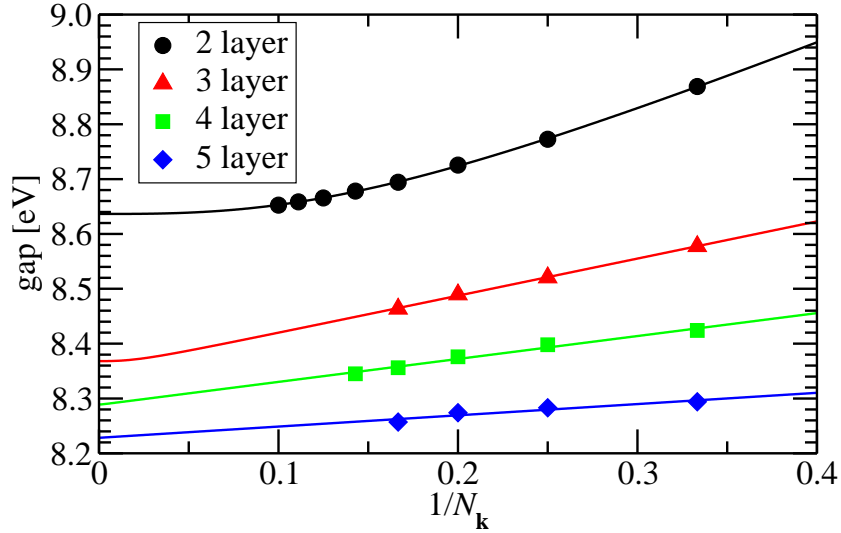


Figure 3.9:  $\mathbf{k}$ -point convergence of the quasiparticle gap for NaCl slabs of varying thickness. The solid lines represent fits to Equation 3.25 (see text).

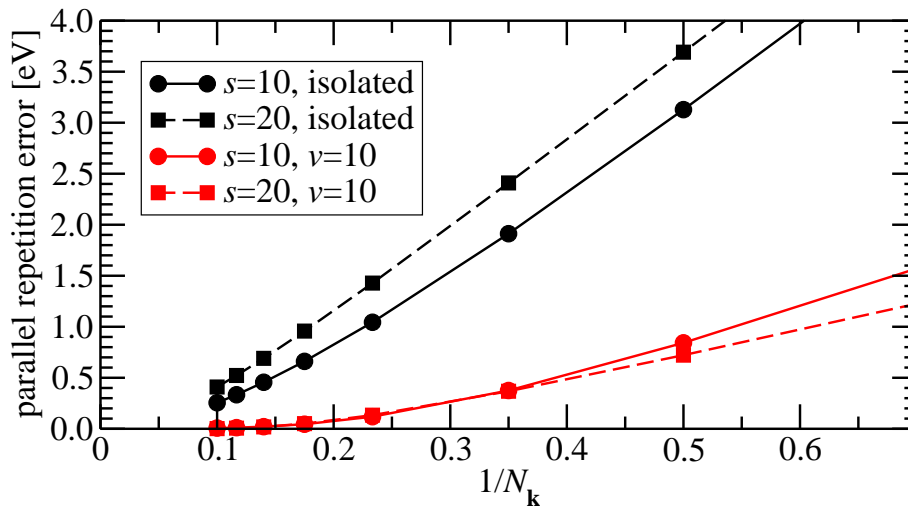


Figure 3.10: Simulated  $\mathbf{k}$ -point convergence behaviour for isolated and periodically repeated (vacuum thickness  $v=10$ ) slabs (NaCl,  $a_{\parallel} = 7$  bohr,  $\epsilon=2.3$ ) of different thicknesses  $s$ .  $s$  and  $v$  are given in bohr.

with respect to the  $\mathbf{k}$ -sampling orthogonal to the surface depends on the sampling in the parallel direction<sup>8</sup> and becomes very small for sufficiently large  $\mathbf{k}$ -samplings, cf. Fig 3.3. When the convergence in the parallel direction is extrapolated for a fixed sampling in the orthogonal direction, the extrapolated value does practically not depend on the  $\mathbf{k}$ -sampling in the orthogonal direction, i.e. the variation is below the required accuracy. The fitting procedure does therefore also deal with the “under-convergence” in the orthogonal direction.

The qualitative behaviour of the  $\mathbf{k}$ -point convergence described here is not restricted to the  $GW$  space-time approach. It is an intrinsic property of the reciprocal-space approach for computing the screened interaction when the subsequent Brillouin zone integrals are replaced by summations. We are not aware of an alternative approach in any of the other existing  $GW$  implementations for periodic systems and would thus assume that these should show a similar behaviour.<sup>9</sup> However, the space-time method may be more appropriate than others to investigate the  $\mathbf{k}$ -point convergence, since the linear scaling of the computation time with respect to the  $\mathbf{k}$ -points is advantageous. The computational effort between the first ( $3\times 3$ ) and last point ( $10\times 10$ ) of the 2-layer NaCl test increases by a factor of  $100/9 \approx 11$  for the space-time method. A convolution approach (i.e. conventional reciprocal-space approaches) scales quadratically with the  $\mathbf{k}$ -point sampling and the last point of the curve would be  $\sim 121$  times more expensive to compute than the first point! This may be one of the reasons why the  $\mathbf{k}$ -point convergence behaviour has rarely been addressed for slab systems before [25, 84]. In addition, isotropic screening models for the treatment of the  $1/r$  part in the screened interaction do not allow to perform such convergence studies in two dimensions since the errors would introduce a logarithmic divergence similar to the linear divergence that we have demonstrated in Section 3.2.3 for the one-dimensional convergence study.

In summary, the screening in slab system shows specific deviations from the macroscopic  $1/r$  behaviour. In the parallel direction, the difference function behaves like  $1/r$  for short distances and like  $1/r^3$  for longer distances.

---

<sup>8</sup>It is likely that this is largely due to the neglect of the wing contributions to the  $\Gamma$ -point dielectric matrix.

<sup>9</sup>This may however depend on the details of the implementation. The improved integration technique by Pulci *et al.* [84] for instance is conceptually based on a summation, but employs a modified Coulomb potential which might improve the convergence behaviour in the parallel direction as a side effect. The Coulomb truncation technique of Ismail-Beigi on the other hand includes a different model function for the reciprocal space singularity [25]. Its real-space form is  $\beta/\sqrt{\alpha^2 + r^2}$ , where  $\alpha$  and  $\beta$  are parameters, equivalent to our fitting functions.

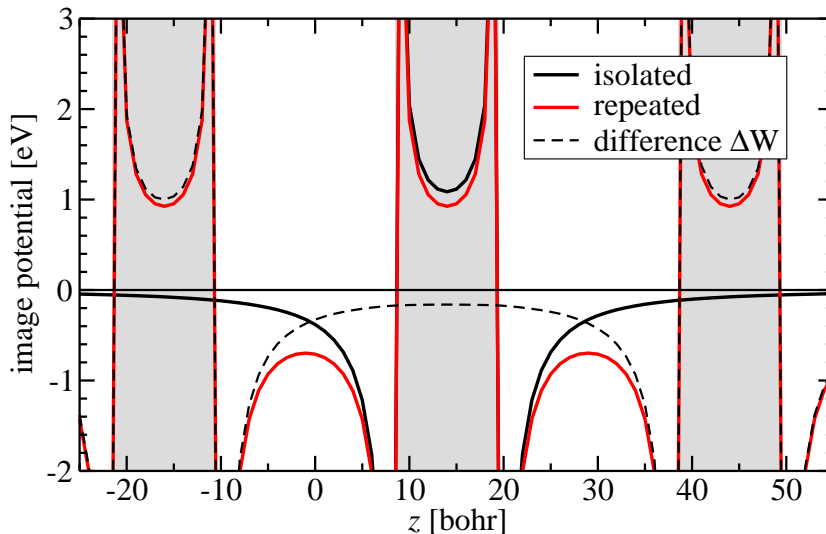


Figure 3.11: Comparison of the image potential for an isolated and a repeated slab (NaCl,  $\epsilon=2.35$ ,  $s=11$ ,  $c=30$ ).

As this difference remains in the numerically treated part of the screened interaction, a peculiar convergence behaviour with respect to the  $\mathbf{k}$ -point sampling parallel to the surface results. The qualitative behaviour can be understood by a simple dielectric slab model. The periodic repetition of the slabs and small vacuum separations reduce this parallel repetition error fortuitously. A robust extrapolation function was derived from the model that can be used to gauge the convergence behaviour with respect to the  $\mathbf{k}$ -point sampling and to extrapolate to the infinite  $\mathbf{k}$ -point limit if necessary.

### 3.3.3 Periodic repetition in the repeated-slab approach

Until now we have shown how it is possible to obtain numerically converged *GW* results for a repeated slab system, in particular how the description of the long-range interaction must be improved to prevent a coupling between the quasiparticle and its periodic images. However, since we are interested in an isolated thin film or a surface, this is still not the physical system that we want to describe. In other words, the quasiparticle will still see the influence of the periodic images of the slab in the perpendicular direction. When a single quasiparticle is placed in one slab, it will polarise the neighbouring slabs, which will induce a change in the image potential.

It has recently been suggested to prevent this polarisation during the computation by truncating the Coulomb potential [25–27]. All the schemes presented so far truncate  $v$  as a function  $\mathbf{r} - \mathbf{r}'$  in order to maintain the

reciprocal-space diagonality. To include all interactions with the slab, but to exclude interactions between the slabs, the vacuum thickness must then be (at least) as large as the slab thickness, i.e. the cell size must be twice as large as the slab itself. We argue that this is numerically inefficient. Not only the increased size makes these calculations up to 8 times<sup>10</sup> more expensive than a normal calculation with a small vacuum, but also very large  $\mathbf{k}$ -point samplings may be required for these quasi-isolated slabs as demonstrated above and also found by others [25, 26]. We therefore do not attempt to suppress this effect in the  $GW$  calculation, but rather correct for it *a posteriori*. The corrections are derived from the dielectric slab model using the dielectric profiles shown in Fig. 3.4. For this purpose, we compute the image potential (without the PRE) for a single slab  $W^{\text{iso}}(z)$  and repeated slabs  $W^{\text{rep}}(z)$ . The resulting image potentials (cf. Fig. 3.11) exhibit non-negligible changes when going from the isolated to the repeated case. However, the difference

$$\Delta W(z) = W^{\text{rep}}(z) - W^{\text{iso}}(z) \quad (3.26)$$

varies very smoothly across the slab [24]. We can therefore apply the simplified COHSEX scheme described in Section B.2 to estimate its contribution to the quasiparticle energies. In this scheme, a constant image potential  $\Delta W$  would induce a scissor-like change

$$\epsilon^{\text{qp}}(\Delta W) = \epsilon_{\text{iso}}^{\text{qp}} + \begin{cases} +\frac{1}{2}\Delta W & \text{unoccupied} \\ -\frac{1}{2}\Delta W & \text{occupied} \end{cases} \quad (3.27)$$

in the quasiparticle spectrum. We can then correct for instance the quasiparticle gap for this finite-vacuum effect by the change in the image potential, determined at the centre of the (model) slab. In some cases (in particular for surface and image potential states), we may reach the region of non-negligible deviations from the slab centre value. In order to obtain a best guess for the constant part of  $\Delta W$  we compute its expectation value  $\langle \psi_n | \Delta W | \psi_n \rangle$  for each of these states separately.

Since our dielectric layer model (Appendix B.1) allows to compute the image potential for arbitrary dielectric profiles, we are not restricted to integer ratios between vacuum and slab thickness as in previous work in our group [24]. This also poses the question for the optimal model parameters. Assuming the simplest possible model – dielectric slabs with thickness  $s$  and

---

<sup>10</sup>The computational effort scales cubically with the system size when the empty space is covered by basis functions as is the case for plane-waves. The size of the two-point functions then scales quadratically with the system size. The construction of  $G$  (space-time approach) or  $P$  and  $\Sigma$  (reciprocal-space approaches), that involve a summation over unoccupied states, and the inversion of  $\tilde{\epsilon}$  then scale cubically.

a dielectric constant  $\varepsilon$ , separated by a vacuum of thickness  $c - s$  (where  $c$  is the known height of the repeated unit cell), we can link the parallel<sup>11</sup> and orthogonal components of the dielectric tensor  $\varepsilon_{\parallel}$  and  $\varepsilon_z$  of the repeated slab system to the model parameters via effective medium theory [90]:

$$\varepsilon_{\parallel} = \frac{(c - s) + \varepsilon s}{c} = 1 + (\varepsilon - 1) \frac{s}{c} \quad (3.28)$$

$$\varepsilon_z^{-1} = \frac{(c - s) + \varepsilon^{-1} s}{c} = 1 - (\varepsilon - 1) \frac{s}{\varepsilon c} \quad (3.29)$$

Solving for  $\varepsilon$  and  $s$  yields

$$\varepsilon = \frac{\varepsilon_{\parallel} - 1}{1 - \varepsilon_z^{-1}} \quad (3.30)$$

$$s = c \cdot \left[ \frac{1}{1 - \varepsilon_{\parallel}} + \frac{1}{1 - \varepsilon_z^{-1}} \right]^{-1} \quad (3.31)$$

Since the dielectric tensor is required for the anisotropy treatment, it is computed in the  $G_0W_0$  calculation [86]. Using Eq. 3.30 and 3.31, we can then obtain model parameters for correcting the contribution of the neighbouring slabs to the quasiparticle energies that are consistent with the  $G_0W_0$  calculation. When the thickness of the vacuum is varied, the computed dielectric tensors perfectly agree with the expectations from effective medium theory (not shown here, but in [24]). Correspondingly, the derived model parameters are independent of the vacuum thickness. Moreover, the change in the image potential is not very sensitive to the model parameters, which guarantees a robust correction scheme.

We will now demonstrate that the correction scheme allows to extract the isolated slab quasiparticle energies from a repeated-slab calculation with an arbitrary vacuum thickness. We focus on the quasiparticle gap since the finite vacuum influences mostly the separation between the occupied and unoccupied levels, whereas level shifts within the occupied or unoccupied part of the spectrum due to the weak curvature of  $\Delta W$  are minor. In Fig. 3.12 we show the results of the  $\mathbf{k}$ -point extrapolation and the resulting gap for a 2-layer NaCl slab with varying amount of vacuum. The graph clearly underlines that for ultrathin slabs, a  $\mathbf{k}$ -point extrapolation for each vacuum separation is crucial to obtain converged results. The convergence behaviour depends on the vacuum thickness, in full agreement with the predictions from Section 3.3. The lower graph on the right shows that the extrapolated quasiparticle correction converges only slowly with increasing cell size to the

---

<sup>11</sup>possibly averaged

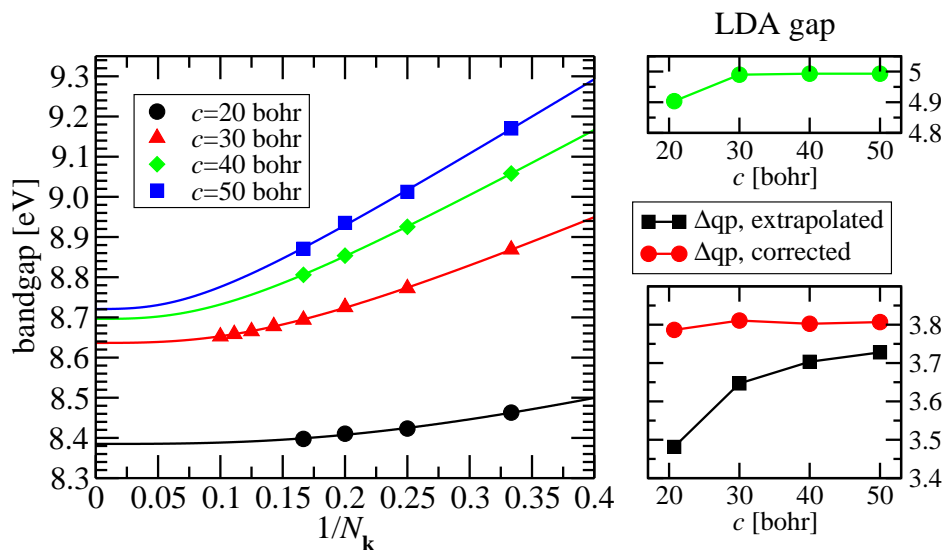


Figure 3.12: Left:  $\mathbf{k}$ -point convergence for the quasiparticle gap of a 2-layer NaCl slab for different total cell heights  $c$ . Right: DFT-LDA gap (top) and extrapolated quasiparticle correction  $\Delta_{\text{qp}}$  as well as its value corrected for the finite vacuum (bottom) as a function of  $c$ .

isolated slab limit. At a finite  $\mathbf{k}$ -point sampling, the increase would be even more dramatic due to the PRE described in the previous section. When the data are corrected for the finite-vacuum effect, however, the quasiparticle correction becomes practically independent of the vacuum thickness. In this very case, the quasiparticle correction including the finite-vacuum correction converges even faster with respect to the vacuum thickness than the underlying LDA calculation. Thus, our *a posteriori* correction even reverses the order of convergence in favour of *GW*.

Earlier results from our group indicated that the dielectric model does not describe the finite-vacuum effect for ultrathin slabs well, which was attributed to the breakdown of a macroscopic description for these systems [24]. We show here that once the  $\mathbf{k}$ -point convergence is carefully extrapolated, the correct physical behaviour is recovered. The success of the modified approach can be attributed to the following improvements:

- The  $\mathbf{k}$ -point convergence is fully taken into account.
- The slab model parameters (thickness and dielectric constant) are extracted from the *GW* calculation and contain possible changes in the dielectric properties of an ultrathin slab.
- The refined dielectric layer model allows to compute the effect for non-

integer vacuum/slab ratios and does not rely on additional interpolation/extrapolation assumptions.

Since the dielectric layer model is not specific to the space-time method, but describes the physical behaviour of repeated slab systems, other *GW* implementations that treat surfaces as periodically repeated slabs must observe the same behaviour. In turn, any implementation that does not recover the slow convergence with respect to the slab separation employs additional implicit or explicit approximations for the long-range behaviour of the screened interaction. We believe that this will in general lead to uncontrollable systematic errors. One example is the treatment of long-range interactions in Rohlfing’s Gaussian orbital *GW* implementation [61], which builds – though physically motivated – on assumptions for the screening behaviour at the surface. The improved *GW* space-time method with finite-vacuum corrections yields quasiparticle surface band gaps for semiconductor surfaces which are systematically larger by 0.1 eV to 0.2 eV compared to the Gaussian orbital results [24]. Although other differences between the implementations like the use of plasmon-pole models or the different basis sets might contribute to this discrepancy, the systematic errors due to implicit assumptions increase the theoretical uncertainty considerably compared to the usual estimates of 0.05 eV to 0.1 eV.

### 3.4 Summary

In this chapter we have reviewed *GW* calculations for two-dimensional systems in the repeated-slab approach. The long-range part of the screened interaction and its role in *GW* calculations were discussed. We have shown that an appropriate treatment of the  $1/r$  part is essential to avoid numerical instabilities, and we have presented a treatment that includes anisotropic screening for the space-time method. We have then studied the  $\mathbf{k}$ -point convergence in slab systems, and found that it is dominated by long-range screening effects beyond the anisotropic  $1/r$  behaviour. We have proved that the repeated slab approach and *small* vacuum separations are advantageous for this question, and we have developed a reliable and robust extrapolation scheme. Finally, we have demonstrated that the influence of the finite vacuum in the repeated slab approach can be quantitatively corrected for *a posteriori*. It must be emphasised that these issues arise from the construction of the correlation part  $\Sigma_c$  of the self-energy. The exchange self-energy  $\Sigma_x$  and the exchange-correlation potential depend only on the occupied DFT-KS states. Thus, they show essentially the same convergence behaviour as the ground-state calculation [24]. The removal of the self-interaction of the



underlying DFT-LDA calculation is due to these contributions and therefore not affected by the repeated-slab approach. Furthermore, the changes in the self-energy due to the polarisation effects do not vary significantly in the direction parallel to the surface at the length-scale of the unit cell. The corresponding variations of the  $G_0W_0$  corrections to the band structure vary therefore little over the Brillouin zone.<sup>12</sup> We conclude that the combination of an anisotropic treatment, a careful convergence with respect to the  $\mathbf{k}$ -points, and an *a posteriori* correction of the finite-vacuum effect provides an accurate and numerically efficient scheme to compute  $GW$  corrections for two-dimensional systems in the repeated-slab approach.

---

<sup>12</sup>The band structure  $\mathbf{k}$ -point must not be confused with the  $\mathbf{k}$ -point sampling for the construction of  $\Sigma$ . The difference is most obvious in the convolution approach. For a band at point  $\mathbf{q}$  in the Brillouin zone, the relevant self-energy is

$$\Sigma_{\mathbf{q}}(\mathbf{r}, \mathbf{r}', i\tau) = \sum_{\mathbf{k}} G_{\mathbf{q}-\mathbf{k}}(\mathbf{r}, \mathbf{r}', i\tau) W_{\mathbf{k}}(\mathbf{r}, \mathbf{r}', i\tau),$$

where  $\mathbf{q}$  and  $\mathbf{k}$  are independent.

# Chapter 4

## Ultrathin oxides: representative surface models or unique films?

### 4.1 Introduction

In this Section we will discuss how thin oxide films compare to the corresponding bulk materials and their surfaces. We will approach this question from a ground-state perspective and employ density-functional theory for this purpose. As already mentioned in the introductory Chapter, we will focus mainly on the implications for heterogeneous model catalysts, in which ultrathin oxide films of often only one or two monolayers (ML) thickness are used to represent an oxide support material of macroscopic thickness. This is a drastic approximation and requires a systematic investigation of the relevant aspects to learn more about the significance of surface science experiments for the situation on a real oxide surface.

In order to analyse the differences between an ultrathin oxide film on a metal substrate and the surface of a bulk oxide, it is helpful to consider intermediate steps between the two limiting cases. We have sketched such a “pathway” in Fig. 4.1 where we take apart the supported thin film system in a controlled manner and “transform” it into the surface of a bulk oxide. In contrast to experiment, this is easily achieved in theory and helps to disentangle the mutual influence of the substrate/film interface, the unique atomic configurations within the thin film, and the evolution of a bulk-like behaviour when going from a two-dimensional film to a three-dimensional bulk system. First we ask the question what role the substrate plays for the properties of the thin film. When the supported film is linked to the metal substrate

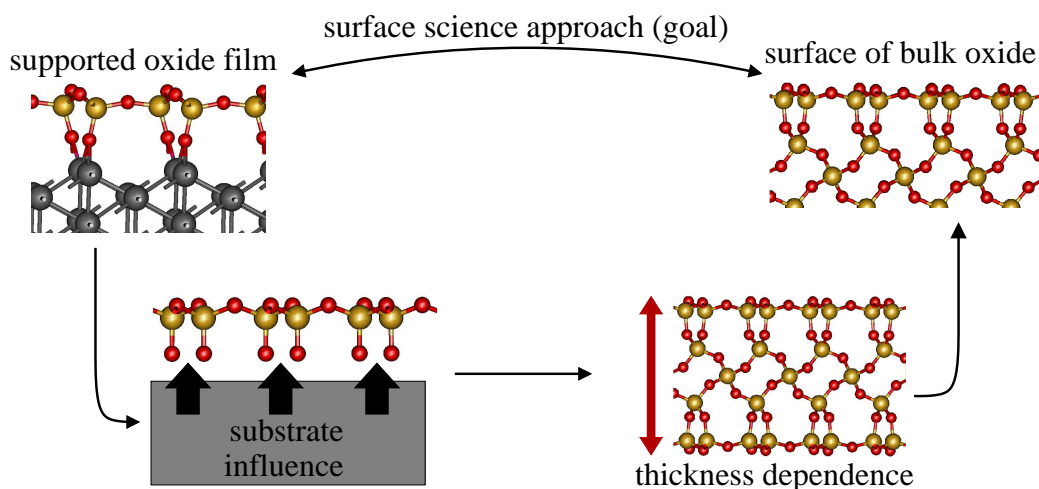


Figure 4.1: Important aspects of the comparison between supported thin films and bulk surfaces.

by strong chemical bonds (which is the case for e.g. silica and alumina), this implies a strong influence of the interface on the local bonding arrangements. Moreover, the surface lattice constants of commensurate well-ordered films are determined by the lattice constant of the respective substrate.<sup>1</sup> In Section 4.3, we will study the substrate influence for the example of the thin silica film on Mo(112) since this film fortuitously resembles an  $\alpha$ -quartz (0001) surface and is hence not specific to the Mo(112) substrate.

The thickness of the film is the second key aspect and will be the topic of Section 4.4. It is well known that the properties of nanosized materials may differ significantly from their bulk counterparts, to which various factors contribute. For instance, films that are only a few atomic layers thick expose a significant part of their atoms at the surface. The surface may then influence or even dominate the properties of the film. Furthermore, the charge separation to form ions and hence the ionic bonding is a cooperative effect: the local charge transfer from a cation to an anion is enhanced by the Madelung field of the surrounding ions, which in turn depends on the charge of these ions. In a film of finite thickness, this cooperativity is disturbed.

<sup>1</sup>We note here that this provides a simple explanation for the structural wealth of the thin silica and alumina films. The bulk oxide structure does usually not match well with the substrate and would experience considerable strain, even when taking into account that the optimized lattice constant for a (freestanding) thin film may in general deviate from its bulk value [91]. This strain can be released either by the introduction of lattice defects – up to the point where the film becomes essentially amorphous – or by reconstructions at specific thicknesses. This point will be addressed in more detail in Sec. 4.3.4.

For ultrathin films also quantisation and confinement effects should be considered. They usually become important when the thickness of the film falls short of the typical length scale of the electron or hole states, but our studies show that they play no important role for the oxides studied in this work. To ensure that the chosen films systematically converge towards the bulk we study these effects for free-standing films in a surface unit cell compatible to the bulk. The thickest films then correspond to the slab models routinely used to model surfaces in theoretical simulations. The goal of studying the thickness dependence with free-standing films without a substrate is not to judge the suitability of a particular film for a particular purpose, but serves mainly to identify general systematic trends. We therefore consider three characteristic oxides (silica, alumina, and hafnia) not only from the field of heterogeneous catalysis, but also for electronic device applications where the possible thickness dependence may set a limit in the on-going trend to smaller and smaller devices. Silica and alumina are important support materials for noble-metal catalysts. All three oxides are current materials for the insulating layer in field-effect transistors, which are the heart of every modern integrated circuit, and we thereby touch another area in which the properties of ultrathin oxide films are of great interest.

For both steps of our pathway, we will focus the comparison on the differences in the atomic and electronic structure, which are the key properties to which many others such as the chemical, spectroscopic, or dielectric properties are correlated. The atomic structure is well described with DFT-LDA. We will assess the changes in the electronic structure from a ground-state perspective, too. The DFT-LDA density of states (DOS) reflects the influence of several relevant physical effects, such as quantum confinement, the electrostatic potential, or the hybridisation of different atomic states, but may not directly be compared to experiments which probe the quasiparticle spectrum such as photoelectron spectroscopy. For a direct comparison with these techniques, quasiparticle corrections must be included. This will be discussed in Chapter 5, where we will show for a computationally feasible system that the quasiparticle spectra of supported insulator films will include additional image potential effects absent from the DFT-LDA ground-state calculations. However, we will show here in Sec. 4.3.3 that the DFT-LDA electronic structure is sufficient to qualitatively explain several aspects of the experimental spectra from ultraviolet photoelectron spectroscopy (UPS) and metastable impact electron spectroscopy (MIES).

The physical and chemical principles that govern the ionic bonding in the bulk materials apply to the thin films, too. It is therefore instructive to recapitulate the properties of the bulk materials, which will be done in Section 4.2.

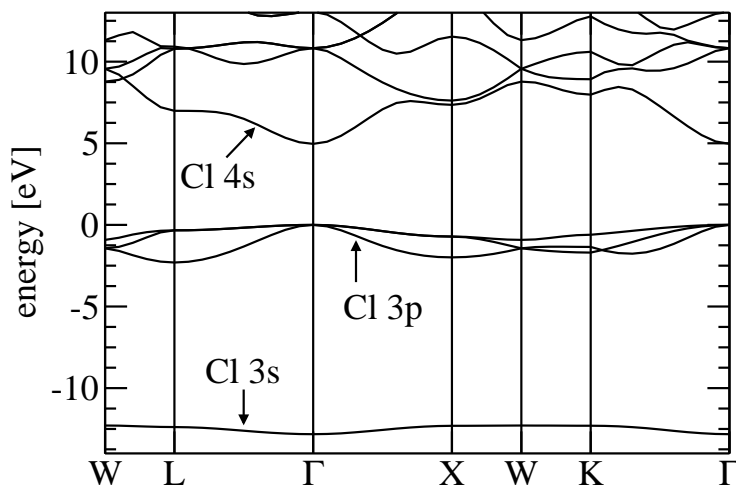


Figure 4.2: DFT-LDA band structure of bulk NaCl at the theoretical lattice constant (5.49 Å).

## 4.2 Properties of the bulk materials

In this Section we will briefly summarise the atomic and electronic structure of the bulk materials from which the thin films in this work are derived. In addition to the oxides that are the subject of this Section (silica, alumina, and hafnia) we will also include sodium chloride here since it exemplarily demonstrates the characteristic properties of wide-gap insulators and will be used to investigate  $G_0W_0$  effects in Chapter 5.

### 4.2.1 Sodium chloride

Sodium chloride is probably *the* prototypical ionic material. It crystallises in the rocksalt structure, where each ion is octahedrally surrounded by six ions of opposite charge. The experimental lattice constant is 5.63 Å, compared to our theoretical equilibrium constant at the DFT-LDA level of 5.49 Å which agrees with all-electron results. We attribute this underestimation by about 3% mainly to the description of exchange and correlation between the  $\text{Na}^+$  core and the valence electrons in the LDA. Also metallic sodium (which consists of  $\text{Na}^+$  and an essentially homogeneous electron gas) shows a similar underestimation of the lattice constant [92]. This is further corroborated by the comparison to a GGA functional (PBE in this case) for the  $\text{Na}^+$  ion: while the density is almost indistinguishable from the LDA, the effective potential is noticeably less attractive in the GGA than in the LDA. Indeed, the GGA functionals give significantly improved lattice constants for metallic sodium

[92].

The electronic structure is simple, cf. Fig. 4.2. In the following bulk band structures, the valence band maximum defines the energy zero. We will focus on the DFT-LDA electronic structure at the theoretical lattice constant in the following. Applying  $G_0W_0$  corrections mainly affects the gaps between the bands<sup>2</sup> while leaving the band dispersion essentially unchanged. The valence part of the bandstructure consist of two bands, a 0.5 eV wide band with mainly Cl 3s character and a 2.3 eV wide band dominated by the Cl 3p orbitals with some admixture of Na 3s and 3p states. Using the atomic orbital (AO) projection technique described in App. D.1, we finde that the Na-centered orbitals contribute, depending on the  $\mathbf{k}$ -point, only 0–12% to the norm in the AO basis for these states, whereas the Cl orbitals account for 85–99%. The lowest conduction band begins 5.0 eV above the valence band maximum (DFT-LDA). Its character is more difficult to assess since different approaches give different results. A detailed discussion is given in Sec. D.2. In brief, we find that it is a hybridisation of the Na 3s and 3p states with a chlorine scattering state (which may be denoted Cl 4s). The chlorine 4s plays a crucial role for the dispersion of this band, while its energetic position is determined by the sodium states.

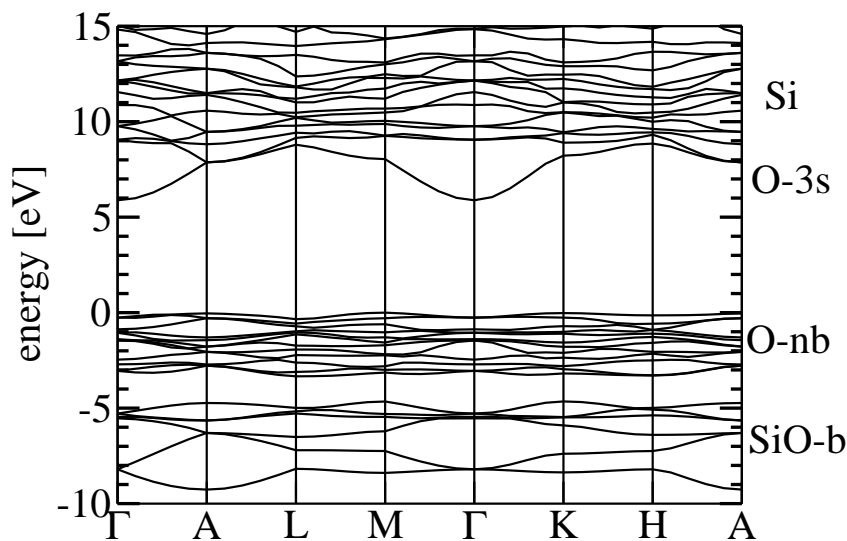
A highly dispersive conduction band similar to the one discussed here is found in all binary main group compounds that are ionic insulators, for instance the alkali halides and earth-alkali chalcogenides. Notably, the wide-gap oxides considered next all show such a O 3s-derived, free-electron-like band, which hybridises with the cation states to a lesser extent than what we observe for the conduction band in NaCl. The valence part of the oxides is – as in NaCl – dominated by the anion states (i.e. O 2p) with a few percent admixture of the cation states. We will now proceed to the discussion of the individual oxides.

### 4.2.2 $\alpha$ -Quartz

Our first example will be silica, or silicon dioxide ( $\text{SiO}_2$ ). Silica has a large number of polymorphs. The ratio of the tabulated ionic radii [94] of  $\text{Si}^{4+}$  (0.54 Å) and  $\text{O}^{2-}$  (1.26 Å) is 0.429, very close to the critical value for octahedral coordination (0.414, [94]). Most polymorphs therefore assume a tetrahedral coordination for the silicon centres and a two-fold coordination for the oxygen, often denoted as 4:2 coordination. Only the metastable stishovite

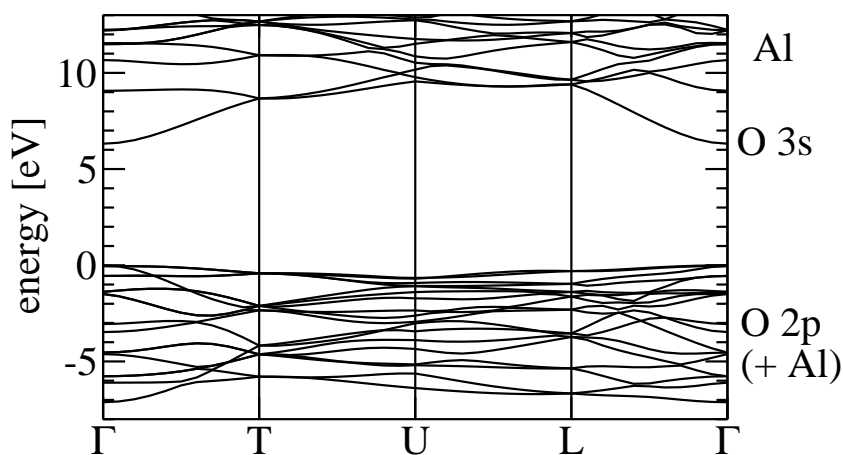
---

<sup>2</sup>The conduction band is shifted upward by 3.3 eV relative to the valence band maximum. The quasiparticle gap then becomes 8.3 eV in reasonable agreement with experiment (8.5 eV [93]). The Cl 3s band is shifted upward by 0.5 eV. We also find that the  $G_0W_0$  corrections change only negligibly when the lattice constant is varied by a few percent.

Figure 4.3: DFT-LDA band structure of bulk  $\alpha$ -quartz.

phase with a rutile structure shows an octahedral 6:3 coordination. In the tetrahedral phases, the Si-O bond length is  $\sim 1.6 \text{ \AA}$  and the Si-O-Si angle varies between  $\sim 140^\circ$  and  $180^\circ$  [95]. The bond length is shorter than the empirical estimates for a purely ionic bond ( $1.80 \text{ \AA}$ ) and lies between the values for a covalent single bond ( $1.77 \text{ \AA}$ ) and a double bond ( $1.53 \text{ \AA}$ ). A considerable covalent character has therefore been ascribed to the chemical bonds in  $\text{SiO}_2$ . The structure of the 4:2-coordinated phases can be characterised as a network of  $[\text{SiO}_4]$  tetrahedra connected at their corners. The polymorphs then differ in the long-range order of the  $[\text{SiO}_4]$  units. Correspondingly, their energetic stability is very similar [95]. The most stable silica polymorph is  $\alpha$ -quartz, where the tetrahedra form interconnected helical chains.

The electronic band structure of  $\alpha$ -quartz is shown in Fig. 4.3 and can be understood as follows: The occupied states are derived from oxygen  $2p$  orbitals with some admixture of the Si  $3s$  and  $3p$  states. Since oxygen is only two-fold coordinated, the O  $2p$  orbitals split into two subbands: those that point towards the Si neighbours form the Si-O bonding (SiO-b) bands, which have some Si character. The other two orbitals form very flat, oxygen non-bonding (O-nb) bands, which are separated from the SiO-b subband by a gap of  $\sim 1 \text{ eV}$ . This separation into SiO-b and O-nb subbands exists for all 4:2-coordinated silica bulk phases. The lowest conduction band is the O-3s band, which hybridises to some extent with Si states. Most of the very flat bands at higher energies derive from Si. At the optimised lattice constant, we find

Figure 4.4: DFT-LDA band structure of bulk  $\alpha$ -alumina

an indirect band gap of 5.9 eV between M<sup>3</sup> and  $\Gamma$  (DFT-LDA). The total O-2p band width is about 9.3 eV.

### 4.2.3 $\alpha$ -Alumina

The second oxide that we will consider is alumina. Other than silica, its structure is easily understood from close-packing arguments. In accordance with the ratio of the ionic radii ( $0.68 \text{ \AA} : 1.24 \text{ \AA} = 0.55$ , [94]), the aluminium ions in  $\alpha$ -alumina, or corundum, have an octahedral coordination. The oxide ions form a close-packed hexagonal structure and are surrounded tetrahedrally by Al.  $\alpha$ -alumina is the thermodynamically most stable phase at ambient conditions. A second polymorph,  $\gamma$ -alumina, is a high-temperature phase with a lower density than  $\alpha$ -alumina. In  $\gamma$ -alumina, the oxide ions form a close-packed cubic structure, where the aluminium ions statistically occupy both tetrahedral and octahedral sites [94]. We point out that the lower density in  $\gamma$ -alumina is accompanied by a lower average coordination of the aluminium atoms, a motif that plays a role also for supported thin films. Since many alumina thin films exhibit Al ions in tetrahedral configurations, too, they are often classified as  $\gamma$ -like (e.g. on Ni<sub>3</sub>Al(111) [96], NiAl(001) [97], or NiAl(110) [98]). However, the lack of long-range order makes  $\gamma$ -alumina unsuitable for our theoretical investigations. We therefore focused our investigations on  $\alpha$ -alumina.

The band structure of  $\alpha$ -alumina is presented in Fig. 4.4. The oxygen 2p-orbitals form the valence band with a width of 7.1 eV, somewhat smaller

<sup>3</sup>However, the valence band at K is only 0.02 eV lower



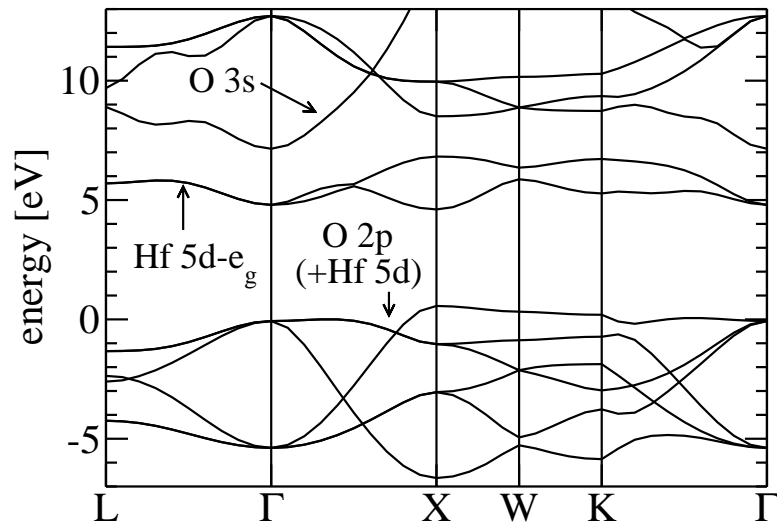


Figure 4.5: DFT-LDA band structure of bulk cubic hafnia

than in  $\alpha$ -quartz. A separate oxygen non-bonding subband like in silica does not exist due to the tetrahedral configuration of the oxide ions. The band gap is direct at  $\Gamma$  and amounts to 6.3 eV (DFT-LDA), similar to the direct band gap in  $\alpha$ -quartz (6.1 eV). As in silica, the lowest conduction band has a very strong oxygen 3s character with some admixture of Al 3s.

#### 4.2.4 Cubic hafnia

Hafnium has the largest cation of the three oxides considered (ionic radius ratio  $0.97 \text{ \AA} : 1.24 \text{ \AA} = 0.78$ ), which makes a cubic coordination of the hafnium ions possible. The stable bulk form at room temperature is monoclinic hafnia with a 7-fold coordination. However, we focus on the high-temperature polymorph cubic hafnia here because it appears to play an important role in hafnia thin films [99, 100]. It has a  $\text{CaF}_2$  structure with 8-fold cubic coordination of the hafnium ions, and correspondingly a tetrahedral configuration for the oxygen atoms.

The electronic structure of hafnia (see Fig. 4.5) shows a stronger influence of the metal than silica or alumina. The upper valence band has metal-oxygen bonding character and is 7.2 eV wide, similar to the band width in alumina (7.1 eV) which has the same oxygen coordination. The band gap (direct at X) is 4.0 eV (DFT-LDA)<sup>4</sup>. The conduction band derives from the Hf-5d states, which are split by the cubic crystal field into  $t_{2g}$  and  $e_g$  states. The

<sup>4</sup>Experimental data is not available since the bulk cubic hafnia is not stable at room temperature

latter form the bottom of the conduction band. A band with mainly O-3s character can be identified close to the  $\Gamma$  point at about 7.2 eV above the top of the valence band. At other  $\mathbf{k}$ -points, it hybridises with the  $t_{2g}$  states.

### 4.2.5 Comparison of the oxides

Having presented the different bulk materials, we will now discuss how the ionic properties may be modified in a thin film. For this purpose let us imagine the (hypothetical) case of forming a film directly from the bulk material. To achieve this, several ionic bonds have to be broken. Since the ionic bonds are not directed, the ions rearrange to either increase their coordination or to reduce the repulsive interactions between the ions of equal charge, thereby increasing the local symmetry of the coordination shell. The magnitude of these relaxations or reconstructions depends on the relative number of broken bonds. Therefore, the small coordination numbers in silica will lead to significant reconstructions because already a single broken bond represents a large disturbance. For cubic hafnia, on the other hand, reconstructions are unlikely to occur (and indeed not observed for other examples of this structure type such as  $\text{CaF}_2$  and  $\text{CeO}_2$ ) because most of the bulk environment is still present, and also the relaxations are small (see Sec. 4.4.3). A second important point is the density of the bulk structure, since it is connected to the flexibility of the oxide. Very open structures like that of silica undergo reconstructions involving several layers. In alumina, the close-packed oxide layers tolerate only minor changes within the layers and only the inter-layer separation is a flexible degree of freedom.

What can we then expect for the electronic structure? The bulk bands of all three oxides are very similar in their character, width, and also the gaps between them. Characteristic are the localised O  $2p$  states<sup>5</sup>, forming a 7–9 eV wide valence band, and a dispersive O  $3s$  band in the conduction band starting 6–7 eV above the top of the valence band. Only in hafnia, additional metal states appear between these oxygen-derived states. The localised states will not be affected strongly by perturbations in the long-range order, but change sensitively when the local order is modified. The formation of a thin film can be seen as a perturbation of the long-range order. If the most of the atoms thereby remain in a bulk-like local environment, the overall features of the electronic structure of a thin film are therefore expected to be similar to those of the bulk oxide. On the other hand, changes in the local coordination shell due to the subsequent relaxation are very important for the details of the electronic structure.

---

<sup>5</sup>Localisation here refers to the extent of the atomic orbitals contributing to a band.

## 4.3 Substrate influence for silica films

In this Section we will investigate how a metal substrate influences and modifies the properties of a thin film for a specific example: silica on Mo(112). The system represents an experimental breakthrough in the growth of crystalline silica films. While silica films on Mo(100) and Mo(110) are amorphous [101–103], the silica film on Mo(112) shows a very sharp  $c(2\times 2)$  LEED pattern. It has been studied with a variety of experimental techniques, including low-energy electron diffraction (LEED) [16, 17, 104–106], Auger electron spectroscopy (AES) [16, 17, 104, 105], X-ray photoemission spectroscopy (XPS) [104, 105], ultraviolet photoemission spectroscopy (UPS) [17, 106], metastable impact electron spectroscopy (MIES) [17], IRAS (infrared reflection absorption spectroscopy) [17, 106], high resolution electron energy loss spectroscopy (HREELS) [13, 16], and scanning tunneling microscopy (STM) [13, 106]. The structure of this film has been a matter of debate [16, 104–107] that has been recently solved by means of DFT calculations [13, 14]. We will first show that the structure of the well-ordered silica film on Mo(112) is closely related to the reconstructed  $\alpha$ -quartz (0001) surface. This fortuitous similarity motivated us to study the influence of the substrate for this system, while for instance alumina films usually strongly deviate from any known alumina surface. We employ DFT to compare the atomic and electronic structure between the silica film on Mo(112) and the  $\alpha$ -quartz (0001) surface. The key aspects of the substrate influence identified here are then later used to discuss other oxide films. We will also compare our computed electronic structure to UPS and MIES experiments for the silica film on Mo, demonstrating that the observed features are characteristic for the  $\text{SiO}_2/\text{Mo}(112)$  system, but not for silica surfaces. Finally we will draw some conclusions on the use of supported monolayer-thin films in heterogeneous catalyst studies.

### 4.3.1 Atomic structure: the siloxane surface

A reliable atomistic model of the thin silica film on Mo(112) has only recently been found by a combination of DFT simulations and several experimental techniques, including STM, XPS, and HREELS [13, 14, 108]. The local atomic structure of the film corresponds to that of the tetrahedral silica polymorphs, but its long-range order is notably different. The  $[\text{SiO}_4]$  tetrahedra form a two-dimensional, perfectly bonded hexagonal network, which is linked to the substrate via oxygen atoms (cf. Fig. 4.6). The film is one monolayer thick, and it is not possible to add further silica layers without destroying the surface. This warrants the conclusion that the film corresponds to a surface

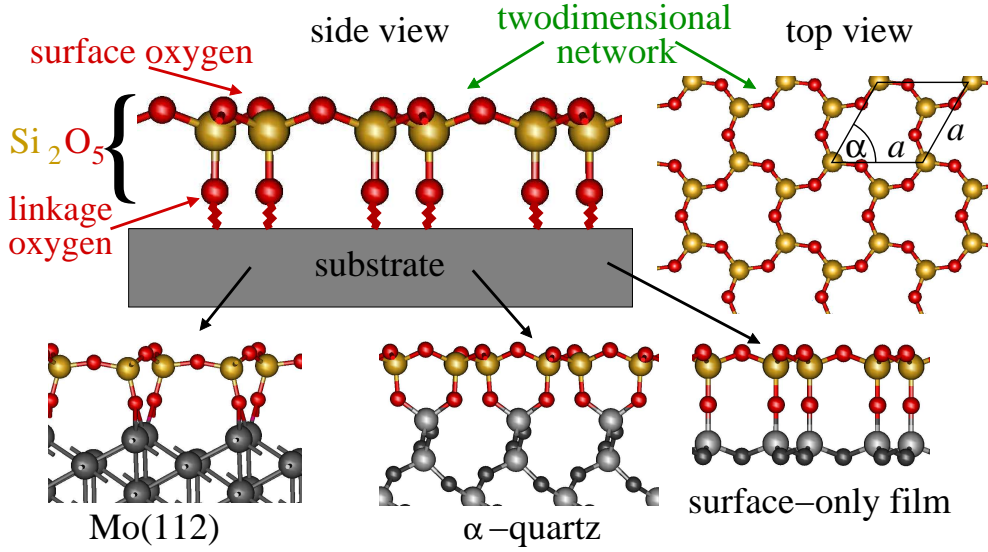


Figure 4.6: Structure of the siloxane surface on Mo(112),  $\alpha$ -quartz, and the hypothetical freestanding surface-only film. Red spheres indicate oxygen atoms, yellow spheres silicon atoms.

form of silica. A comparison to known silica surface and thin film structures reveals that the bonding pattern, i.e. the connectivity and long-range order of the  $[\text{SiO}_4]$  units, is identical to the “dense” reconstruction of the  $\alpha$ -quartz (0001) surface [109], the “silicate adlayer” on 4H and 6H SiC [110], and the “siloxane surface” of clay minerals [111]. In the following, we will refer to this surface form as “siloxane surface” because it emphasises the “siloxane bond” Si-O-Si as characteristic surface motif as opposed to hydroxyl groups or unsaturated bonds.

Having established the similarity between the silica film on Mo(112) and the  $\alpha$ -quartz(0001) surface, we will now compare the atomic structure in detail. The main question we pose is how well the silica film found in experiment represents the corresponding  $\alpha$ -quartz surface. To our knowledge a well-ordered, dry  $\alpha$ -quartz surface has not yet been observed in experiment, where amorphous and partially hydroxylated surfaces prevail. However, theoretical simulations predict it to be the most stable surface structure in the absence of hydrogen [109] in full agreement with our own results (see Sec. C.1). We will denote the two systems as the siloxane surface on a Mo and on a quartz substrate, respectively. As starting point for the structure optimisation of the  $c(2\times 2)$  silica film on Mo(112) in our DFT-LDA calculation, we assume the most stable configuration identified by Todorova *et al.* [13, 108], where the linkage oxygens occupy a two-fold bridge position on the Mo rows of the

	Mo(112)	$\alpha$ -quartz(0001)	Si <sub>4</sub> O <sub>8</sub>
a [Å]	5.21 (5.22)	4.85 (4.92)	5.24
$\alpha$ [°]	63	60	60
area [Å <sup>2</sup> ]	24.2 (24.3)	20.4 (21.0)	23.8
Si-O (surface) [Å]	1.61–1.63	1.60–1.61	1.61
Si-O (bridge) [Å]	1.63	1.62–1.63	1.61
Si-O-Si (surface) [°]	134, 162	118–125	141
[SiO <sub>4</sub> ] rotation [°]	0.1–0.2	21–22	0
[SiO <sub>4</sub> ] tilting [°]	14	4	0
O buckling [Å]	0.4	0.2	0

Table 4.1: Structural parameters (DFT-LDA) of the siloxane surface on Mo(112) and  $\alpha$ -quartz and the relaxed pure-surface Si<sub>4</sub>O<sub>8</sub> model in comparison. Numbers in brackets refer to experimental data.

(112) surface. The small differences in the structural parameters reported here and in [13] can be traced to the different functional (GGA) employed and reflect the well-known trends towards larger bond lengths in the GGA. The Mo substrate is modelled by 5 layers<sup>6</sup> and the surface unit cell corresponds to a c(2×2) superlattice of the (112) surface of Mo at the theoretical bulk lattice constant (3.139 Å). The  $\alpha$ -quartz (0001) surface is modelled by a symmetric Si<sub>8</sub>O<sub>16</sub> slab at the theoretical lattice constant of  $\alpha$ -quartz. To learn more about the intrinsic structural properties of the “ideal” siloxane surface when the lattice constant is varied, we have additionally studied a pure-surface Si<sub>4</sub>O<sub>8</sub> film (cf. Fig. 4.6).

The structural parameters of the siloxane surface for the three cases are summarised in Tab. 4.1. As we will show in the following, the influence of the substrate on the structure of the siloxane surface can be attributed to two key factors: the lattice constant of the substrate and the linkage of the film to the surface of the substrate. The siloxane surface adapts to these constraints easily by reorienting the tetrahedral building blocks, a relaxation mode that mostly affects the rather flexible Si-O-Si angles. The harder degrees of freedom such as the Si-O bond length remain almost identical to those of bulk  $\alpha$ -quartz (cf. Tab. 4.1).

Let us focus on the role of the lattice constant first. The structural properties of an “ideal” siloxane surface independent of its substrate were studied with the free-standing Si<sub>4</sub>O<sub>8</sub> film that consists of two siloxane-surfaces directly linked together. When the lattice constant of this film is varied, the behaviour below and above the optimum lattice constant of 5.24 Å differs (cf.

<sup>6</sup>Using 7 Mo layers shows negligible differences.

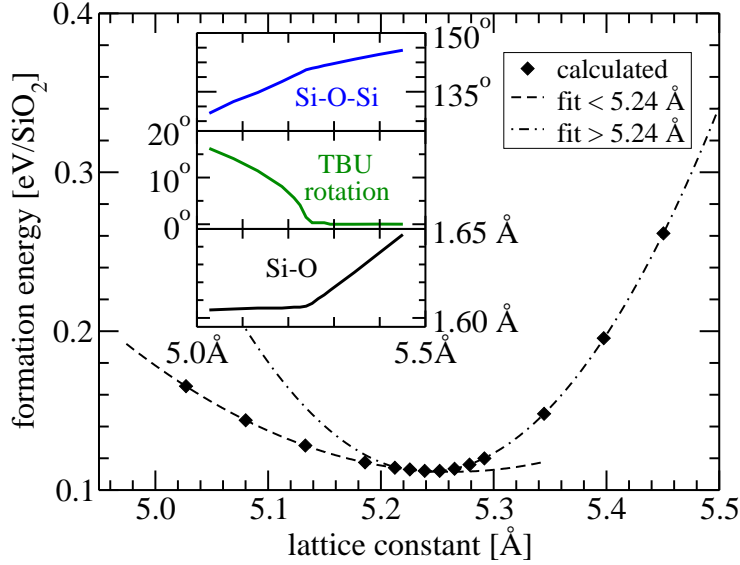


Figure 4.7: Formation energy and important structural parameters as a function of the lattice constant for the siloxane pure-surface  $\text{Si}_4\text{O}_8$  film. To demonstrate the different behaviour above and below  $5.24 \text{ \AA}$ , separate quadratic functions have been fitted to the two regions.

Fig 4.7). The formation energy on either side of the minimum is a quadratic function of the lattice constant, but the curvature differs by a factor of  $\sim 3$ . Under compression, the energy rises more slowly since the surface network adapts by rotating the  $[\text{SiO}_4]$  tetrahedra along their vertical axis, a very soft degree of freedom. We use the in-plane component of the Si-Si-O angle to monitor this rotation (schematically illustrated in Fig. 4.8 a). This rotation mechanism is known from phyllosilicates [112] and results in a reduction of the Si-Si distance (and hence the lattice constant) with only moderate changes at the oxygen atoms. At the energy minimum, the rotation angle has reached zero. Under further expansion, the  $[\text{SiO}_4]$  tetrahedra are distorted and the Si-O bonds are stretched. Since harder degrees of freedom are involved in this case, the energy rises more quickly than under compression. The relaxation of the siloxane surface on the Mo and  $\alpha$ -quartz substrates follows the same pattern. The lattice constant of the quasi-hexagonal  $c(2 \times 2)$  Mo(112) superlattice ( $a=5.21 \text{ \AA}$  and  $\alpha=63^\circ$ ) is very close to that of the “surface-only”  $\text{Si}_4\text{O}_8$  film ( $5.24 \text{ \AA}$ ). Correspondingly, the  $[\text{SiO}_4]$  rotation about the surface normal is essentially zero. We note that the lattice constant of the  $(\sqrt{3} \times \sqrt{3})R30^\circ$  pattern observed for the SiC substrates [110] is  $5.25 \text{ \AA}$  (DFT-LDA [113]) and matches very well with the relaxed surface, too. This provides a simple explanation why the siloxane surface forms readily on

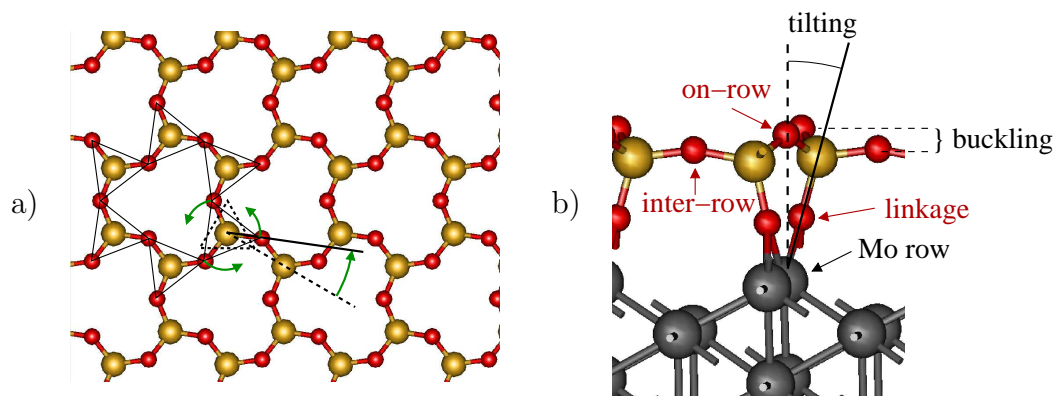


Figure 4.8: a)  $[\text{SiO}_4]$  rotation, schematically. b) Definition of tilting and buckling for the siloxane surface.

the lattice-matched substrates Mo(112) and 4H/6H-SiC (0001). The lattice match is of course only one of several factors determining the stability. The linkage to the substrate is another one as discussed below for Mo(112). In the case of SiC, this is almost ideal (hexagonal pattern of dangling Si bonds). The lattice mismatch to  $\alpha$ -quartz ( $4.85 \text{ \AA}$ ) on the other hand amounts to 8%. As in the pure-surface model, the compression of the siloxane surface on the quartz substrate is accompanied by a large  $[\text{SiO}_4]$  rotation ( $\sim 20^\circ$ ). The considerable lattice mismatch probably introduces a significant number of defects in real  $\alpha$ -quartz surfaces and explains the lack of long-range order. The defects may then be responsible for its propensity towards hydroxylation.

Apart from the lattice constant, the oxygen linkage to the substrate is an important factor, too. The Mo(112) surface exhibits rows to which the linkage oxygen atoms bind in a two-fold bridge position [13]. The three surface oxygen atoms of the two-dimensional network thus become nonequivalent: two of them link  $[\text{SiO}_4]$  units along the same row and one forms the inter-row connections (cf. Fig. 4.8 b). In order to achieve the bonding to the substrate, the trifold symmetry axis of the  $[\text{SiO}_4]$  tetrahedra is tilted by  $14^\circ$  against the surface normal towards the inter-row oxygen. The bonding angle at this atom ( $162^\circ$ ) is considerably larger than along the row ( $134^\circ$ ). The tilting also introduces a height difference (buckling, cf. Fig. 4.8 b) for the surface atoms of  $\sim 0.4 \text{ \AA}$ . The siloxane surface on quartz, however, tilts only by  $4^\circ$  and buckles by  $0.2 \text{ \AA}$ . The tilting and buckling on  $\alpha$ -quartz are smaller because not only the bulk-truncated quartz substrate matches better with the preferred hexagonal arrangement of the linkage oxygens, but it also relaxes more than the Mo(112) surface. The mismatch-induced strain is hence distributed between the siloxane surface and the  $\alpha$ -quartz substrate, in contrast

to the Mo-substrate where only the siloxane surface is affected.

In summary, we have shown that the connectivity of the network found for the silica film on Mo(112) is not unique to this substrate, but appears to be common to a variety of hexagonal substrates. We suggest that it might exist on more substrates than currently known, provided that the lattice constant of the (quasi)hexagonal substrate is close to 5.25 Å and that the substrate offers oxygen adsorption sites in an approximately hexagonal arrangement. While the surface lattice of a Mo(112)  $c(2 \times 2)$  structure is very close to that of the ideal siloxane surface, the linkage of the surface to the Mo rows introduces significant distortions. The siloxane surface can also be found on  $\alpha$ -quartz when the surface is compressed by  $\sim 8\%$ . From our structural analysis, we conclude that the siloxane surface on Mo(112) found in the surface science experiments exhibits noticeable deviations from the ideal  $\alpha$ -quartz surface although the chemical connectivity is identical. As we will see in the following, these differences induce changes for instance in the electronic structure and we expect that also the chemical properties are affected. We note also that the different relaxation behaviour of the two substrates may become important for the creation and reactivity of point defects in the film. While the  $\alpha$ -quartz substrate readily supports relaxations around a point defect, this is not the case for Mo(112). Point defects in silica films on Mo(112) are indeed rare [106, 114].

The key factors for the substrate-induced structural changes identified here, namely the lattice constant and the interface constraints, can be used to discuss other thin film structures, too. We will take the two-layer alumina film on NiAl(110) as an example. This film is easily obtained by exposing the clean metal surface to oxygen at elevated temperatures [98]. The self-limited oxidation results in a well-defined film that has been intensively studied over the last 15 years [15, 98, 115–117]. Only very recently, however, its complex structure could be finally solved by combining DFT calculations and STM investigations [15]. The film has a non-stoichiometric  $\text{Al}_{10}\text{O}_{13}$  composition and exhibits Al atoms at the surface surrounded by three or four O atoms. These motifs have been interpreted as tetrahedral and truncated octahedral configurations. Kresse *et al.* have shown that the structure of the bottom (Al) layer is dictated by the preferential adsorption sites of the aluminium atoms on the NiAl(110) surface, i.e. by the interface to the substrate [15]. The substrate lattice also plays a considerable role for this film. The NiAl(110) surface lattice does not match with bulk alumina polymorphs. A comparison of the alumina bulk polymorphs reveals that changes in the lattice constant are always linked to changes in the coordination of the aluminium atoms. The varying coordination of the aluminium atoms found in the  $\text{Al}_2\text{O}_3/\text{NiAl}(110)$  film can thus be seen as a way of adapting the film to



the lattice constant of the substrate.<sup>7</sup> Both for silica and alumina, however, the interplay between the interface structure and the lattice relaxation stabilises structures of a particular thickness that do not correspond to a piece of the bulk material.

An example where the bulk structure is maintained and the thickness of the film can be varied is MgO on Mo(100) [19]. This film easily grows bulk-like despite a lattice mismatch of  $\sim 5\%$  because none of the dominant lattice relaxation mechanisms enforce deviations from the rocksalt structure. A complex interface structure does not develop because the flat MgO(100) surface matches reasonably with the Mo(100) substrate.

We conclude that crystalline oxide films with a bulk-like structure can only be grown on substrates where the surface lattice constant and the surface structure are compatible with the bulk material. How deviations of the ideal parameters are accommodated by the film then depends on the structural chemistry of the oxide. Silica and alumina, which have a very rich structural variety, will require very well tuned substrate properties to stabilise the bulk structures, whereas other oxides tolerate considerable deviations.

### 4.3.2 Electronic structure

We will now come back to the siloxane surface and compare the electronic structure of the siloxane surface on the Mo(112) substrate, i.e. the film found in experiment, and the  $\alpha$ -quartz (0001) surface, which represents a perfect, dry silica surface. For the comparison of the different systems, we use the vacuum energy as a common energy zero.<sup>8</sup> Two important questions will guide us through the comparison: the more fundamental aspect if and how the hybridisation between the oxide surface and a metal substrate will affect the insulating character of the oxide surface, and the more case-specific question of how the changes in the atomic structure modify the features visible in the electronic structure. We will focus on the DFT-LDA electronic density of states (DOS) for this purpose. To compare to UPS or MIES experiments, the LDA band energies should be corrected for self-interaction and quasiparticle effects (e.g. with  $G_0W_0$ ), and method-specific transition amplitudes should be taken into account. We will estimate the influence of these corrections when we discuss the connection to the experiments in Section 4.3.3.

---

<sup>7</sup>We note that the film is commensurate only along one direction (row-matching), while the incommensurate lattice along the other one introduces line defects.

<sup>8</sup>Alternative alignment strategies such as aligning the top of the valence band, or localised core states, yields similar results.

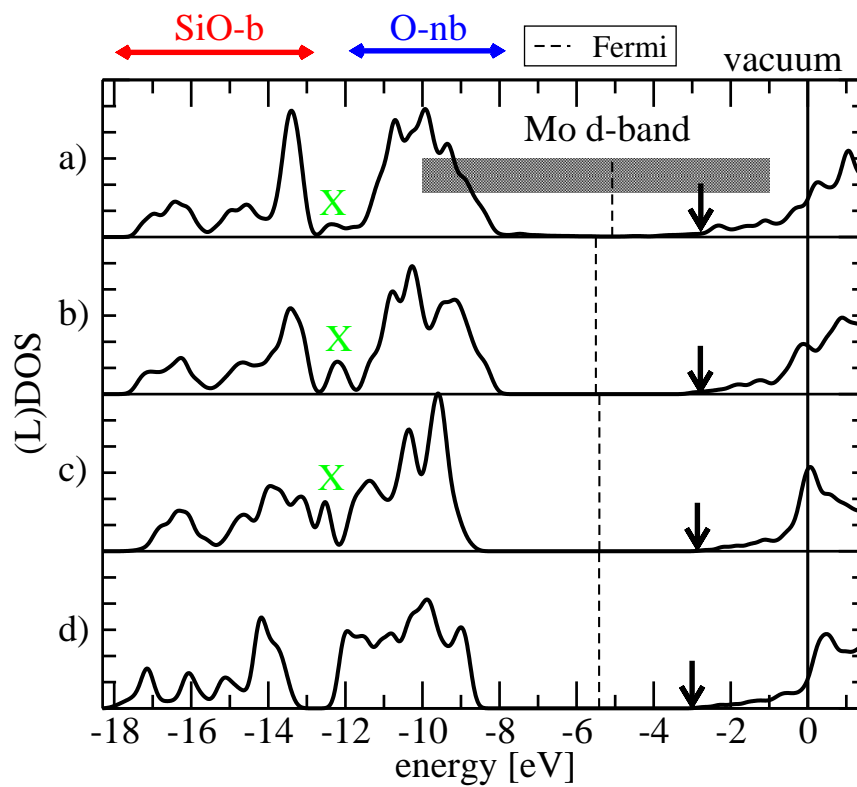


Figure 4.9: Surface LDOS for a) the silica film on Mo(112), b) the same, but the Mo atoms replaced by quartz (cf. text), and c) a relaxed  $\alpha$ -quartz surface. d) Total DOS of bulk  $\alpha$ -quartz. The X indicates the O-3mr feature between the SiO-b and O-nb subbands (cf. text). The bottom of the (surface) conduction band is marked by an arrow.

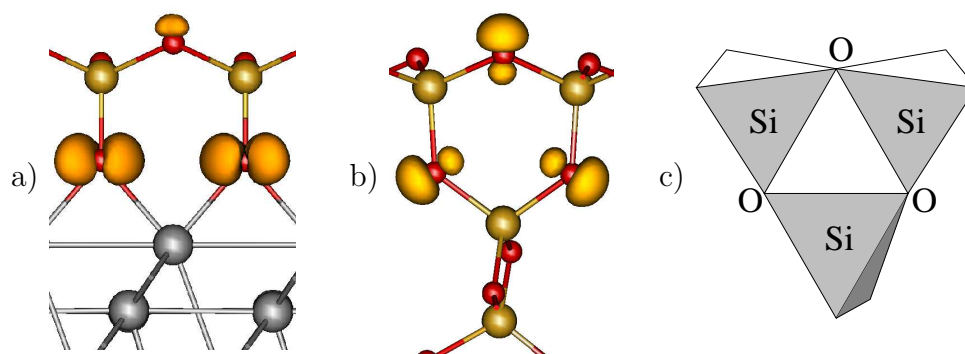


Figure 4.10: Isodensity surface for the partial density from the O-3mr peak on a) the Mo(112) and b)  $\alpha$ -quartz substrates. c) Schematic representation of the 3-membered ring on  $\alpha$ -quartz.

The surface LDOS (see Sec. C.2) for the siloxane surface on Mo(112) and quartz is presented in Fig. 4.9 a) and c). It can be understood on the basis of the  $\alpha$ -quartz bulk bands, and we therefore include the total DOS of  $\alpha$ -quartz as a reference. Surprisingly, the band edges at the surface agree with those of  $\alpha$ -quartz to within 0.5 eV for both substrates. There are no surface states in the large band gap. The Mo metal states, which energetically lie within the surface band gap, play no important role at the surface. This indicates that the electronic structure of the defect-free surface and the substrate are essentially decoupled. The overall wide-gap insulating character of the siloxane surface resembles that of any other tetrahedral silica polymorph, independent of the substrate. However, an additional state (marked by X in Fig. 4.9) appears for the siloxane surface between the SiO-b and O-nb subbands. It is connected to the presence of almost planar “three-membered rings”<sup>9</sup> formed by two  $[\text{SiO}_4]$  units of the surface which are linked to the same substrate atom (Mo or Si). The state is composed of oxygen  $p$ -like orbitals (cf. Fig. 4.10) within the plane of the three-membered ring and will be denoted O-3mr in the following. It is thus not specific to the siloxane surface, but to a common motif of the interface to the Mo(112) and  $\alpha$ -quartz substrates.

The peak structure in the valence region of the surface DOS, however, exhibits clear differences for the two substrates and deviates also from the DOS of bulk  $\alpha$ -quartz. On Mo(112), there is a sharp feature at the top of the SiO-b subband, while the O-nb subband is relatively unstructured with three peaks and a broad flank. The tail of the unoccupied states shows

<sup>9</sup>The terminology in silica structures [112] refers to the number of  $[\text{SiO}_4]$  units in ring structures, not the number of atoms; cf. Fig. 4.10 c).

(small) resonances with the Mo states. On quartz, on the other hand, the O-nb region shows three strong features, whereas the top of the SiO-b band is less peaked than for Mo. The question now is whether these differences are related to the electronic structure of the substrate and/or the interface (the “chemical” influence) and what role the different atomic structure plays (the “structural” influence).

In order to disentangle the changes due to structural variations from other factors such as the chemical differences of the two interfaces, we have replaced the Mo(112) substrate by quartz (4 bulk-like layers + a siloxane surface). These additional atoms were fully relaxed while the original surface structure including the bridge oxygen atoms was kept fixed. The resulting surface DOS is shown in Fig. 4.9b. The differences between spectrum a) and b) in Fig. 4.9 thus reflect the chemical influence when the substrate is changed. The most important changes occur in the centre of the O-nb region and can be traced to the shifts in the mixed surface-interface states. Most of the spectrum, however, remains unaffected. The influence of the structure can be seen by comparing spectrum b) and c). When the lateral lattice is compressed to match  $\alpha$ -quartz and the surface structure is fully relaxed, the SiO-b peak splits into two peaks and partially merges with the O-3mr feature at  $-12.5$  eV. The O-nb region becomes more clearly structured and the band edges become steeper since the non-equivalence of the surface oxygens is reduced. We conclude from this that the changes in the local atomic structure induced by the substrate have a stronger influence on the shape of the electronic structure than the hybridisation with the substrate states.

Our findings can be summarised as follows: the global electronic structure of the siloxane surface is comparable to that of other tetrahedral silica polymorphs. It maintains a very ionic band structure with a large band gap. The band edges agree well with those of bulk  $\alpha$ -quartz. A metallic substrate below the surface does not significantly alter the fundamental, strongly ionic characteristics of the electronic DOS. However, the substrate-induced variations in the atomic structure strongly modify the peak structure within the bands.

Before we address the implications of this for model catalysts and the surface science approach, we will first make a connection to available spectroscopic data.

### 4.3.3 Comparison to UPS and MIES experiments

The surface electronic structure of the siloxane surface has been characterised experimentally with UPS and MIES [17, 106]. The spectra show significant

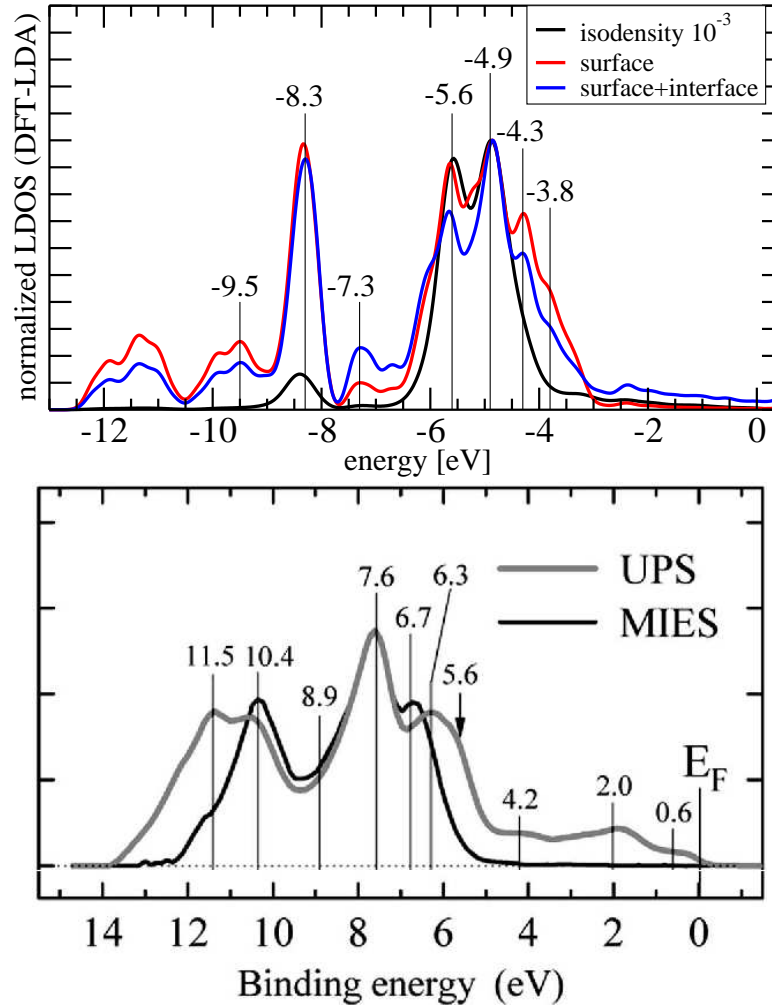


Figure 4.11: Top: Comparison of the local DOS in the LDA for different regions of the siloxane surface on Mo(112), see text. The spectra are normalised to the same maximum peak height. The (substrate) Fermi energy here defines the energy zero, i.e. the spectrum is shifted by 5.1 eV with respect to the vacuum-aligned spectra of Fig. 4.9. Bottom: Experimental spectra, taken from Fig. 4c) in [17].

differences due to the surface sensitivity of the two methods. In UPS, the electrons are excited by light which penetrates deep into the sample. In MIES, optically excited He atoms are shot at the surface which then emits electrons. While the probing depth of UPS is determined by the inelastic scattering of the emitted photoelectrons in the sample with typical mean-free paths in the 1-10 nm range, the MIES signal is believed to result mainly from Auger deexcitation of the incident helium atoms when electrons from the surface tunnel into the empty He 1s state [118]. Since the helium atoms never enter the surface in this case, the MIE spectrum is highly sensitive to the electronic structure of the surface, in particular to the tails of the electronic states at the surface. We have simulated this sensitivity by evaluating the DOS at an isodensity surface at low electron densities ( $10^{-3}$  e bohr $^{-3}$ ). In Fig. 4.11 we compare such an isodensity LDOS with the LDOS for two regions that may correspond to the surface sensitivity of the UP spectra, namely the topmost layer and the first two layers. We display both regions to demonstrate that the major differences occur between the isodensity LDOS and the surface-layer LDOS and not – as suggested in the original interpretation of the experiment [17] – between different oxygen layers.

The experimental MIE spectrum is considerably narrower than the UP spectrum, which was explained by the presence of an interface state at the top of the O-nb band [17]. The theoretical counterparts, i.e. the isodensity LDOS and the surface-region LDOS, reproduce the smaller band width, but suggests a different mechanism for the narrowing: the state at the top of the valence band that is visible only in UPS can be identified with a non-bonding  $p$ -orbital at the inter-row oxygen atom, oriented parallel to the surface plane (not shown). It is shadowed by the out-of-plane oxygen orbital at the same atom which extends further out into the vacuum and will thus be preferentially probed by the approaching He atom. At other surface atoms, the differently oriented orbitals cannot be associated with specific peaks in the spectrum. However, while the width of the spectra in the different techniques agrees well with our LDOS plots, the spectral intensities of the experimental spectra as well as the absolute position are not well reproduced by this simplified approach. In particular, the intensity at higher binding energies in the SiO-b region is significantly underestimated. Computing the spectral intensities, which depend on the state-specific transition amplitudes for each spectroscopic method, is outside the focus of this work. We therefore compare only the peak positions of the spectrum with experiment (cf. Tab. 4.2).

Since quasiparticle corrections are neglected in the computation of the DOS at the level of DFT-LDA and also due to inherent deficiencies of the LDA (self-interaction), the energetic position of the silica features with re-

	Si-O	O-3mr	O-nb				
UPS <sup>a</sup>	-11.5	-10.6	–	-7.6	–	-6.3	-5.6 <sup>c</sup>
MIES <sup>a</sup>	-11.6 <sup>c</sup>	-10.4	–	-7.6	-6.7	–	–
DFT-LDA <sup>b</sup>	-9.5	-8.3	-7.3	-5.6	-4.9	-4.3	-3.8 <sup>c</sup>

<sup>a</sup> From Fig. 4c) in Wendt *et al.* [17]; <sup>b</sup> this work; <sup>c</sup> shoulder

Table 4.2: Peak positions relative to the Fermi level for the electronic surface DOS of the siloxane surface on Mo(112) in comparison with experiment. The theoretical data is not corrected for the estimated quasiparticle shift of 2 eV (see text).

spect to the Fermi energy is not well described. A scissor shift for the silica surface states can be estimated from bulk silica when possible local-structure, finite-size, surface, interface, and Fermi alignment effects are ignored. Using this crude approximation, we estimate that the computed position of the silica features in the valence region is too high by  $\sim 2$  eV, half of the quasiparticle corrections for the band gap of bulk silica<sup>10</sup>. Taking into account a quasiparticle shift of 2 eV, a good agreement with experiment for the silica film on Mo(112) is obtained. This agreement warrants the assignment of the peaks made in this section. It is interesting to note that Wendt *et al.* have interpreted these spectra as supportive evidence for their isolated SiO<sub>4</sub> cluster model [16], whereas we find that the electronic spectra are consistent with the siloxane surface model. It must be emphasised that the agreement between theory and experiment includes the features that were predicted to differ between the Mo(112) and the  $\alpha$ -quartz substrates. We consider this strong evidence that the experimental spectra indeed reflect film-specific features rather than generic  $\alpha$ -quartz features as claimed e.g. in [106]. The comparison to the DFT-LDA LDOS can help to understand the origin of the experimental signals, but to improve the agreement between theory and experiment regarding the absolute position and the spectral intensity,  $G_0W_0$  should be applied to obtain a better quasiparticle band structure and also transition probabilities should be included.

<sup>10</sup>Chang *et al.* [119] report a quasiparticle correction for  $\alpha$ -quartz of 4.5 eV, whereas our own  $G_0W_0$  calculation give 3.4 eV. Ramos *et al.* [120] obtain corrections for  $\beta$ -cristobalite of 4.3–4.8 eV including a 0.2–0.3 eV overestimation due to certain approximations made.

### 4.3.4 How well do supported monolayer films model realistic surfaces?

As we have seen in this Section, a supported silica film of one monolayer thickness develops a structure that is directly influenced by the substrate. The ionic bonding within the film is not altered fundamentally compared to the situation in the bulk, but applying the same principles that govern the structure in bulk materials leads to a different result for thin films. For oxide films that develop a “chemical” interface to the substrate, i.e. which are linked to it by chemical bonds, the lattice is no longer determined by the oxide material itself, but by the respective substrate. The film seeks to maintain the local coordination of its ions wherever possible. A different lattice can therefore only be achieved by

1. compressing or stretching inter-ionic distances by small amounts,
2. distorting coordination shells,
3. introducing defects, or
4. developing new arrangements of the ions.

These effects can be recognised in the experimentally grown films. The first two points on the list become increasingly costly when the film becomes thicker. As suggested in point 3 of our list, thicker films grow often defective or even completely amorphous. New regular arrangements, on the other hand, can often only be stabilised as ultrathin films on a substrate. The required charge neutrality may involve the metallic substrate or the interface. Therefore, deviations from the bulk stoichiometry become possible such as the  $\text{Si}_2\text{O}_5$  stoichiometry of the silica film on Mo(112), or also the  $\text{Al}_{10}\text{O}_{13}$  stoichiometry [15] of the thin alumina film on NiAl(110). Moreover, the different bonding situation at the interface and the surface already induces different local structures. We can consider these as alternative building blocks from which an ultrathin film structure is built. This may often lead to unique structures that are specific to one particular substrate and one particular thickness. However, even in the fortunate situation where the film structure is closely related to a bulk surface like the silica film on Mo(112), the local bonding arrangement is altered by the substrate.

While these changes play no important role for the ionic character of the chemical bonds, other properties may be more sensitive. We have seen in Section 4.3.2 that the peaks in the electronic DOS are strongly affected by the structure of the thin film. This has important consequences for defects created in the film. Since active defects in the band gap derive from bands



of the perfect material, the shifts observed for these bands should translate to similar shifts in the derived defect states. In addition, the existence of a Fermi level in the metal substrate may modify the charge state of the defects when electrons tunnel from or to these states [121]. We note that defects are very important for the chemical properties of the film. Since the perfect materials are chemically quite inert, the chemical reactivity of real materials is determined by the defects in the film. Our observations for the influence of the substrate on the electronic properties therefore suggests that also the chemical reactivity of supported monolayer films is modified by the substrate.

In summary, we find that oxide films that are only a few monolayers thick are strongly influenced by the substrate. They will in general not be representative for a bulk material or its surfaces. The changes in the local atomic structure, which are induced by the lattice constant and the interface to the substrate, play a crucial role: they induce further changes e.g. in the electronic structure and the chemical properties. The question now is if the comparability to the bulk can be improved by growing thicker films. We will therefore consider next the possibility to create films that do have bulk properties. We focus on one key aspect, namely the thickness.

## 4.4 Thickness dependence in freestanding films

We now address the surface science approach from a different angle and seek to establish a connection between ultrathin films and thick slab models for the surface of a bulk material. In order to see if ultrathin films can in principle have the same properties as the bulk oxide we eliminate the influence of the substrate by considering free-standing films (cf. step 2 in Fig. 4.1). By varying the film thickness we can tune our system between the two limits: ultrathin films and bulk surfaces. We have chosen three characteristic wide-gap oxides for these investigations: silica, alumina, and hafnia. The orientation of the films is chosen such that their surfaces correspond to the most stable, simple surface of the respective bulk material. We focus mainly on symmetric, stoichiometric films.<sup>11</sup> The thickness of the films is increased until an essentially bulk-like behaviour establishes. These thick films then correspond to the slab systems commonly employed in theory to model surfaces, i.e. we arrive at the situation that is sought in surface science experiments.

---

<sup>11</sup>For silica, also non-stoichiometric terminations were considered, but were found to be far less stable, see Sec. C.1.

### 4.4.1 $\alpha$ -Quartz

As in the previous section, we will focus our discussion on the development of the atomic structure first before we investigate the electronic structure for the most stable films obtained. For silica, several stoichiometric and non-stoichiometric terminations have been considered, see Appendix C.1 for details. In summary, quartz slabs with various surface terminations were constructed and their geometries were optimised. At all thicknesses, stoichiometric slabs with fully saturated bonds were found to be most stable. To achieve saturation, the two outermost quartz layers merge into one to form a siloxane surface or – in the smallest case – a similar structure. We will discuss here only the most stable structures, containing two to ten formula units ( $\text{SiO}_2$ ) per  $(1 \times 1)$  surface unit cell. Films with more than five formula units, i.e.  $\text{Si}_6\text{O}_{12}$  to  $\text{Si}_{10}\text{O}_{20}$ , are best characterised as  $\alpha$ -quartz-like layers sandwiched between two siloxane surfaces. They correspond to two-sided<sup>12</sup> slab models of the (0001) surface of  $\alpha$ -quartz. Thinner films deviate from this pattern in one way or another and we will briefly describe them one by one. The  $\text{Si}_5\text{O}_{10}$  structure contains only one intermediate layer between the siloxane surfaces and is highly strained. To saturate the bonds, the central Si atom must bind to two  $[\text{SiO}_4]$  units of the upper surface and to two of the lower surface. A tetrahedral conformation at this atom would require a relative orientation of the two surfaces of  $90^\circ$ , but the three-fold symmetry of the ideal surface is only compatible with steps of  $120^\circ$ . The result (cf. Fig. 4.12a) is a distorted structure in which the  $[\text{SiO}_4]$  tetrahedra at the surface are tilted by  $28.5^\circ$ . The most stable structure found for  $\text{Si}_4\text{O}_8$  is the siloxane pure-surface film already discussed in Section 4.3 (cf. Fig.4.6), in which two siloxane surfaces are directly linked together. We exclude the  $\text{Si}_3\text{O}_6$  case from the discussion here since it inevitably exhibits dangling bonds at variance with all other films. The most stable structure for  $\text{Si}_2\text{O}_4$  turns out to be a single-layer two-dimensional network with hexagonal rings, where two oxygen atoms lie in the Si plane, while the two others form an out-of-plane two-membered silica ring (cf. Fig. 4.12b), a motif known from other silica surfaces [14, 122].

In the previous section we have already presented the siloxane surface as the most stable reconstruction of the  $\alpha$ -quartz (0001) surface. A remarkable point of this surface reconstruction is that it restores the  $[\text{SiO}_4]$  units with fully saturated bonds at the expense of their long-range order and the Si-O-Si

---

<sup>12</sup>The two surfaces are equivalent, but the slabs are not symmetric. The only non-trivial symmetry (i.e. which is not a pure translation) of bulk  $\alpha$ -quartz is a screw axis along (0001), which is necessarily lost when then translational symmetry along (0001) is broken.

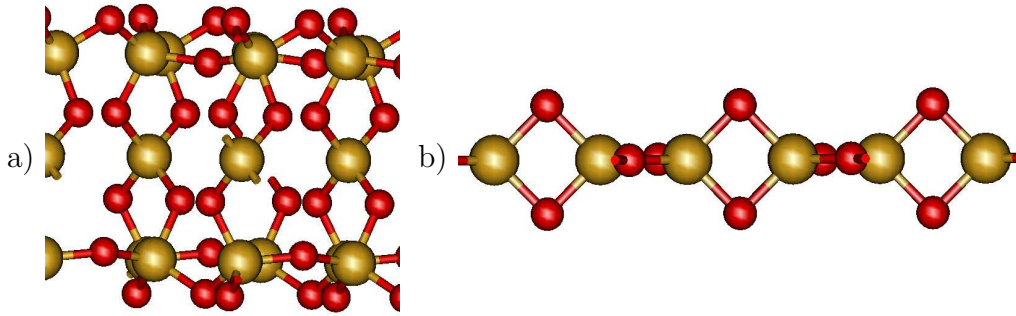


Figure 4.12: Side-view of reconstructed films: a)  $\text{Si}_5\text{O}_{10}$  and b)  $\text{Si}_2\text{O}_4$ .

angles. In other words, the long-range structure is strongly modified to reduce or even remove the perturbations in the local atomic structure. We find the same trend for the ultrathin films. The saturation of the dangling bonds appears to be a very strong driving force for the reconstructions observed. It is supported by the flexibility of the Si-O-Si angles, i.e. the connections of the  $[\text{SiO}_4]$  units in the silica network. The strain induced by the reconstructions is distributed in the network. Correspondingly, we find a strong variation of the Si-O-Si angles with the film thickness.

What does this imply for the use of silica films in the surface science approach? First, we note that the reconstructions required to saturate dangling bonds drastically modify the atomic structure at small thicknesses. A thickness of six formula units is necessary before a bulk-like central part covered by a siloxane surface forms. We note that in real systems the presence of water will offer an alternative to the reconstructions observed here, namely the saturation by hydroxyl groups. We therefore expect that the tendency to form hydroxyl groups will strongly depend on the thickness. Second, the strain induced by the reconstructions is distributed over the network. This has important consequences for the geometrical relaxation around defects and the strain induced by it. Only thicker films can include  $\alpha$ -quartz-like layers, which in general are very flexible, and can thus reduce strain at the expense of long-range relaxations. In thin films, distortions will be more localised leading to higher strain. Last, we have seen that the reconstructions at the surface include the two outermost layers. This indicates that reactions at the surface may involve one or more subsurface layers, and these are not present or behave differently in ultrathin films.

We now come to the discussion of the electronic structure for these films. The total DOS for the different reconstructed slabs is displayed in Fig. 4.13. For two or more quartz-like layers, i.e. from  $\text{Si}_6\text{O}_{12}$  on, the changes in the DOS become rather small and reflect mainly the varying relative contribu-

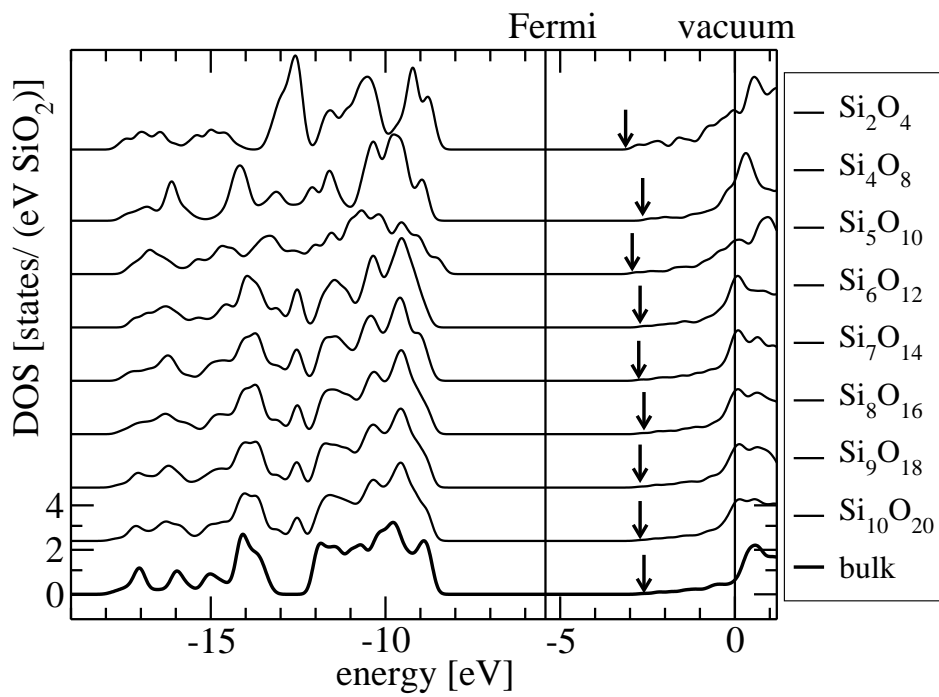


Figure 4.13: Thickness dependence of the DOS for stoichiometric thin silica slabs. The bottom of the conduction band is marked by an arrow.

tion of the thickness-independent surface features<sup>13</sup> to the total DOS. The thinner films exhibit stronger differences, which we attribute to the structural variations described above. In all these ultrathin films, a gap between the Si-O bonding and the O non-bonding does not exist, in contrast to all 4:2-coordinated silica bulk polymorphs.<sup>14</sup> Surprisingly, the band edges and band widths are affected very little by the thickness variation, even for the thinnest films. While this is to be expected for localised states<sup>15</sup>, notably the O-nb states, the insensitivity of the lowest conduction band which is a highly delocalised state in the bulk seems contradictory. For such a state, one could expect similar confinement effects as they are observed in ultrathin semiconductor and metal films [123, 124]. We attribute the absence of such an effect to a surface resonance that traces the edge of the projected bulk bands. It couples to the bulk bands at  $k_z \approx 0$  as is visible from the flat envelope function along the  $z$ -direction. In Fig. 4.14, we demonstrate this for the conduction band minimum. Since  $k_z = 0$  implies a constant phase, the coupling between the surface resonance and a  $k_z = 0$  bulk band is independent of the film thickness, giving rise to the thickness-independent feature in the band structure. This highlights that the magnitude of possible confinement effects cannot be estimated from the bulk bands alone, but also the scattering properties at the surface play a decisive role.

We conclude that ultrathin silica films will usually not resemble any bulk-like structure even if the lattice constant is chosen in accordance to the bulk. To saturate the dangling bonds, strong relaxations occur that have a drastic effect on the shape of the electronic structure. Even the characteristic gap between the SiO-b and O-nb regions of the spectrum disappears. The band edges, however, are almost independent of the thickness. This demonstrates that the band gap is no useful parameter to assess the changes in the electronic structure for ultrathin silica films. Although the perfect films discussed here are expected to be chemically inert due to the large band gap, realistic, defective films will differ in their chemical properties because the changes observed within the perfect bands will translate to corresponding changes in the defect states derived from them. However, we note that relatively few bulk-like layers in the centre of the slab (two in the case of  $\alpha$ -quartz) are

---

<sup>13</sup>The surface LDOS presented in Fig. 4.9 c) gives an impression of these features.

<sup>14</sup>Other than for the siloxane surface, no particular structure motif can be made responsible for this observation.

<sup>15</sup>We note that “localised state” refers here mainly to the character of the atomic orbitals from which the state derives, showing only little inter-atomic overlap. The Kohn-Sham eigenstates, however, are not localised in one part of the slab, but typically extend over the slab. The envelope function typically shows a sine-like, confined shape, but this has no dominant influence on the energy of this state.

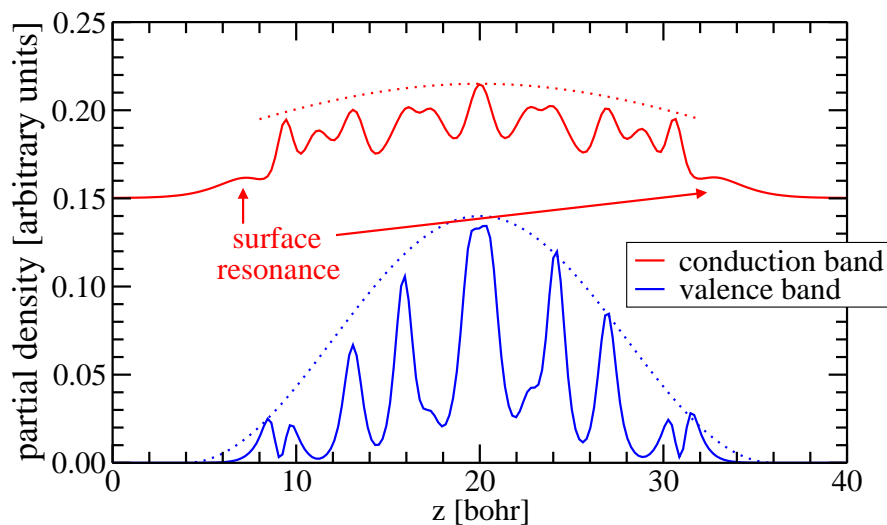


Figure 4.14: Partial electron density profiles (normalised to 1) for the valence band maximum and the conduction band minimum of the  $\text{Si}_8\text{O}_{16}$  slab. While the envelope of the valence state shows a typical confined shape, the conduction band is very flat due to its coupling with a surface resonance.

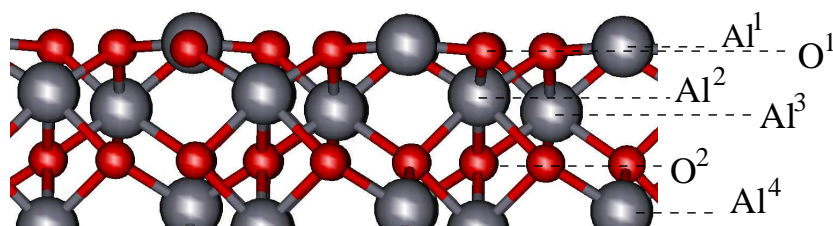


Figure 4.15: Side view of the (0001) surface of  $\alpha\text{-Al}_2\text{O}_3$ . The letters denote the vertical position of the topmost Al and O layers.

sufficient to approach the bulk properties.

#### 4.4.2 $\alpha$ -Alumina

As a second example we consider alumina, which is another important material for heterogeneous catalysts. Thin alumina films have been grown on a wide variety of substrates [11], but the well-ordered epitaxial films are generally found to exhibit large and complex unit cells. The detailed atomic structure has been recently solved for the 5 Å thick film on NiAl(110) [15], but it is often not known for other substrates. We therefore do not try to approach one particular film, but restrict ourselves to the most stable bulk polymorph,  $\alpha$ -alumina. We have chosen the  $1 \times 1$  (0001) surface with an

	Al <sub>4</sub> O <sub>6</sub>	Al <sub>6</sub> O <sub>9</sub>	Al <sub>8</sub> O <sub>12</sub>	Al <sub>10</sub> O <sub>15</sub>	Al <sub>12</sub> O <sub>18</sub>
$\Delta$ (Al <sup>1</sup> -O <sup>1</sup> ) [%]	-62.4	-96.3	-86.9	-85.1	-85.7
$\Delta$ (O <sup>1</sup> -Al <sup>2</sup> ) [%]	+28.3	+3.5	+2.5	+4.2	+4.0
$\Delta$ (Al <sup>2</sup> -Al <sup>3</sup> ) [%]	-100.0	-52.7	-40.3	-44.6	-44.4

Table 4.3: Thickness dependence of the inter-layer relaxation  $\Delta$  between the first layers at the surface of  $\alpha$ -alumina slabs.

Al-O<sub>3</sub>-Al-termination, which is the most stable termination in the absence of hydrogen [125]. Stoichiometric slabs of varying thickness can be constructed by adding Al-O<sub>3</sub>-Al layers. We investigated slabs with a thickness of up to six of these triple layers, i.e. Al<sub>12</sub>O<sub>18</sub>.

The topmost aluminium atom at the (0001) surface shows a very strong inward relaxation, thereby becoming almost coplanar with the oxygen layer (cf. Fig. 4.15). This is easily explained: compared to the bulk, the surface aluminium ion has lost three of its six oxygen neighbours. The ideal local arrangement for the reduced coordination would be trigonal. To approach this configuration, the Al ion relaxes strongly inward, thereby pushing the oxygen ions aside. These relaxations at the surfaces then induce further changes in the deeper layers. Interestingly, we find in our DFT-LDA calculations that the magnitude of the surface relaxation depends on the thickness of the slab. To monitor these changes, we focus on the relative change  $\Delta(L-L')$  in the separation between the layers L and L'. The layers are denoted by the atom type, i.e. Al or O, and are then simply counted by increasing depth, i.e. the first layers are Al<sup>1</sup>-O<sup>2</sup>-Al<sup>2</sup>-Al<sup>3</sup>-O<sup>2</sup>-Al<sup>4</sup>... as indicated in Fig. 4.15. The computed interlayer separations at the surface are summarised in Tab. 4.3 and demonstrate that there is a delicate, non-monotoneous balance of the relaxations in the different layers.

These changes in the atomic structure are very important for the adsorption of atoms or small molecules. On ionic surfaces with low-coordinated atoms, electric fields play an important role not only for the energetics of the adsorption [126, 127], but they may also activate or deactivate bonds in subsequent reactions [128]. The strength of these fields strongly depends on the local geometry. We therefore have to expect a considerable influence of the film thickness on the adsorption properties when only a few monolayers of the oxide are present. For the  $\alpha$ -alumina slabs a thickness of four oxygen layers ( $\sim 10$  Å) is required to obtain a surface relaxation close to that of a bulk substrate (cf. Tab. 4.3).

Are these variations in the relaxation also reflected in the electronic structure, and will they exert a similar influence on the shape of the electronic spectrum as we had observed for silica? The total DOS for the relaxed slabs,

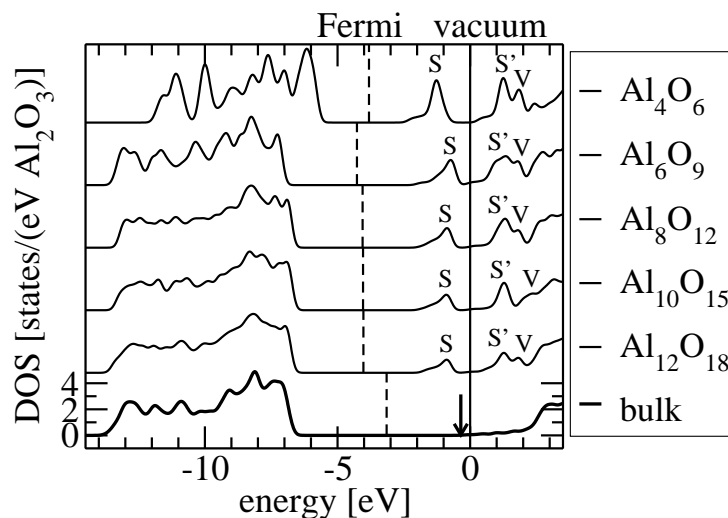


Figure 4.16: Thickness dependence of the DOS for stoichiometric thin  $\alpha$ -alumina slabs in comparison to the bulk DOS. The vacuum energy defines the energy zero. The arrow marks the bottom of the conduction band in bulk alumina. S and S' denote the Al-derived surface states. V is a confined vacuum state, a meaningless artifact of the repeated-slab approach.

shown in Fig. 4.16 in comparison to bulk alumina, gives an answer to these questions. The most striking difference to the silica films is the appearance of a surface state in the band gap of the bulk material. It derives from the unoccupied aluminium states at the surface and overlaps energetically with the bottom of the O-3s-derived bulk conduction band. A second Al-derived surface feature is positioned energetically above the vacuum energy. The lower energy of the aluminium states at the surface relative to the bulk results from the reduction in the Madelung potential at the surface when going from a 6-fold to a 3-fold coordination [129].

We now turn our attention to the valence region. The DOS of the thinnest film differs dramatically from all others: the valence states are shifted by more than 1 eV relative to the conduction states and the vacuum level. A similar effect has been observed earlier for freestanding MgO films by Pacchioni *et al.* [91]. They concluded that the Madelung potential, which stabilises the ionic state, evolves only after a few monolayers. What can be observed for the Al<sub>4</sub>O<sub>6</sub> film is a breakdown of the bulk ionicity, which reduces the energy separation of the anion and cation states. The thickness-dependent relaxation leads additionally to a variation of the surface dipole and hence the alignment of the DOS with respect to the vacuum energy. The calculated Fermi energy then varies strongly between the thinnest films because it coincides with the



centre of the band gap and thus reflects the gap changes from the Madelung effect as well as the changes in the surface dipole. The atomic structure in the central part of the thicker films, i.e. from  $\text{Al}_6\text{O}_9$  on, corresponds to almost perfect corundum and we find an essentially bulk-like valence electronic structure for these films. Only the density of states of the thinnest film differs significantly from the bulk. The quick, almost sudden emergence of a bulk-like shape with increasing thickness differs from the trend observed for silica. We can relate this to the structural relaxation mechanism in the two materials. The local coordination of the oxygen atoms is very important for the valence electronic structure, which is dominated by the oxygen  $2p$  orbitals. In silica, the structure relaxation affects mostly the oxygen atoms. The structural changes therefore visibly modify the electronic structure. In alumina, on the other hand, the deformation of the coordination shell of the aluminium ions is the dominant structural relaxation mechanism, which has no important influence on the O- $2p$ -derived valence states.

The observed bulk-like behaviour for the films beyond  $\text{Al}_6\text{O}_9$  is relevant for a second area of application of thin alumina films, namely its possible use in electronic devices. In field-effect transistors (cf. Fig. 1.3), from which the basic logic elements in integrated circuits are built, the gate electrode must be electrically insulated from the conducting channel that the gate electrode controls. This is achieved by a thin insulating film, the gate dielectric. The capacitance of the film must be low to reduce the switching currents, but the film must be thick enough to avoid leakage currents due to defects or – in particular for ultrathin films – due to tunneling. The dielectric constant (often denoted as  $\kappa$  in this context) should be as large as possible to increase the physical thickness of the insulating film. In addition, the band alignment between the semiconductor and the insulator must create large barriers for the tunneling currents. In silicon-based semiconductors, amorphous silica has been the material of choice for this purpose. The on-going miniaturisation of the electron devices and new semiconducting materials, however, requires alternative high- $\kappa$  materials. Among others, alumina and hafnia (which we will discuss below) have been considered as a possible replacement. Our results indicate that the relevant macroscopic properties of bulk alumina are preserved down to thicknesses of about 10 Å. Only then, size-effects and a break-down of the bulk behaviour must be taken into account.

### 4.4.3 Cubic hafnia

We now turn to hafnia as the last example, which will complete our study of the thickness dependence. In contrast to alumina and silica, hafnia plays no important role for catalytic applications. However, it has attracted consid-

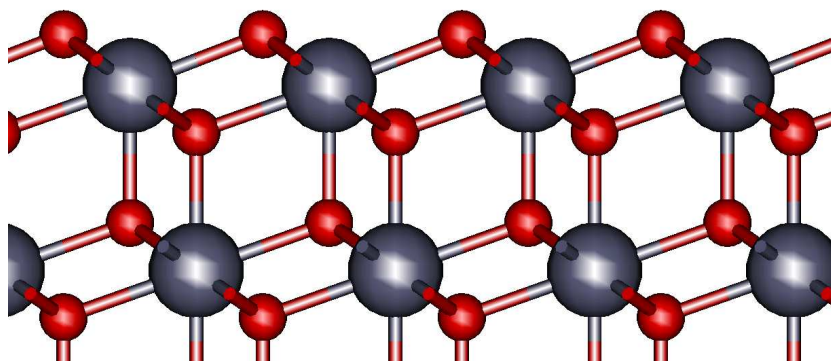


Figure 4.17: Side view of the (111) surface of cubic  $\text{HfO}_2$ .

erable interest for electronic devices due to its large dielectric constant [22]. We have therefore chosen to include this material in our investigations of the thickness dependence. In experiment, hafnia films are usually grown epitaxially on semiconductor substrates, in particular silicon. X-ray diffraction studies have indicated that these epitaxial films may consist of cubic hafnia [99, 100]. However, the structural characterisation at the atomic level is less detailed than for the alumina films.

Electrostatic considerations predict the (111) surface to be the most stable unreconstructed surface, and stoichiometric slabs can then be constructed by adding O-Hf-O layers. The (111) surface is very dense, i.e. only few ionic bonds are broken. The outermost oxygen atoms miss one of their four bulk neighbours, and the Hf ions only one out of eight. Correspondingly, the surface relaxations are much smaller than in alumina and show almost no dependence on the thickness. We can compare the situation to the (001) surface of rocksalt oxides such as MgO, that also show very small relaxations at the surface [129]. We can identify here a clear trend. The higher the coordination numbers in the bulk and the denser the surface termination, the smaller the relaxations at the surface and their sensitivity to the film thickness. This trend has been already visible in the comparison of the reconstructing, 4:2-coordinated silica to the more densely packed, 6:4-coordinated alumina, and continues to the 8:4-coordinated cubic hafnia.

We find a similar result for the electronic structure, cf. Fig. 4.18. From three (triple) layers on, i.e.  $\text{Hf}_3\text{O}_6$ , the DOS in the valence region is bulk-like. The thinnest film  $\text{Hf}_2\text{O}_4$ , which contains only undercoordinated Hf ions, shows again a different shape of the electron spectrum than the bulk which mainly reflects the surface electronic structure. We note that the trifold-coordinated oxygen atom introduce a state right at the top of the bulk valence band which is responsible for the steeper band edge of the thinnest films. In

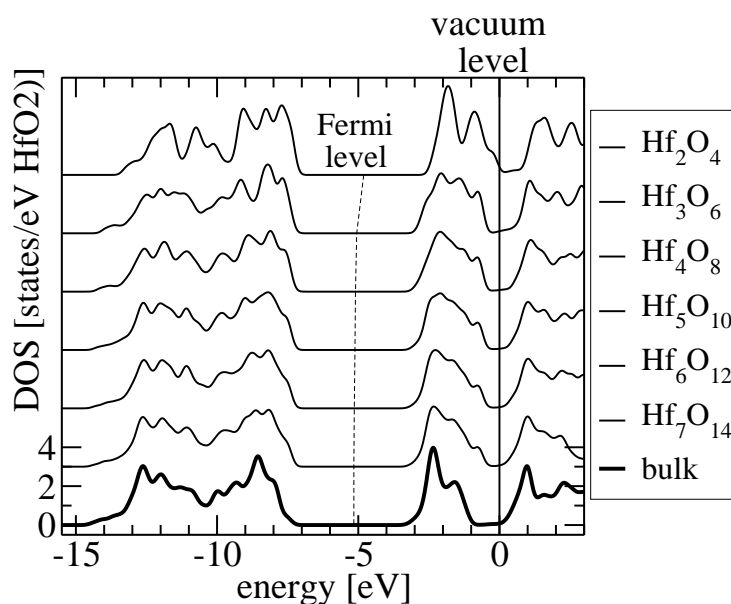


Figure 4.18: Thickness dependence of the DOS for stoichiometric cubic hafnia slabs in comparison to bulk  $\text{HfO}_2$ .

the conduction band, the Hf  $e_g$  states are shifted towards higher energies at the surface, i.e. the opposite direction than for the aluminium surface state in alumina. For hafnia, the change in the Madelung potential at the surface is apparently less important than the reduction of the crystal field splitting. Like for alumina, the evolution of the DOS of thicker films can be understood in terms of a simple mixture of bulk-like and surface-like features. When the thickness is increased, a bulk-like DOS appears quite suddenly above a minimum thickness of four oxygen layers. For both materials, we then find the atoms in the central part of the slabs in a bulk-like environment. We conclude that the macroscopic electronic properties of both alumina and hafnia are essentially identical to those of the bulk materials above  $\sim 10 \text{ \AA}$  thickness. For thinner films, however, we find considerable changes.

## 4.5 Summary

In this Chapter, we have studied the atomic and electronic structure of epitaxial thin films of wide-gap oxides by means of density-functional theory. We have made contact to two important fields of application for oxide films, namely the “surface science approach” to study heterogeneous model catalysts and to electronic devices. We have shown that the structural and elec-

tronic properties of ML-thin oxide films systematically deviate from thicker films and bulk oxides. On the one hand, the substrate crucially influences and modifies the atomic structure. We have shown that the structural relaxations can be explained in terms of the same principles that govern the relaxation in the corresponding bulk structure(s), but the particular situation at the individual surface must be taken into account. We have identified the substrate lattice for commensurate thin films and the interface to the substrate as two key factors that determine the structural constraints for supported thin films, and demonstrated that these key factors can explain many of the structural trends observed for epitaxial oxide films grown on metals. We have then turned our attention to the role of the thickness of the film. Again, we find considerable differences between monolayer-thin films and the bulk materials even if we explicitly exclude any influence from a possible substrate. We have seen that the atomic structure sensitively depends on the thickness below a critical size. Close-packed structures and high coordination numbers in the bulk reduce the thickness sensitivity of the film properties. The largest changes are found for silica due to its high structural flexibility. While the band gaps and band widths are usually very close to the corresponding bulk values even at very small thicknesses, peaks in the electronic spectrum are very sensitive to the structural changes for monolayer-thin films. However, comparatively few bulk layers in the centre of the films are sufficient to obtain a bulk-like electronic structure and converged surface properties. This suggests that already films with a few more monolayers than the currently prevailing one- and two-ML films may be sufficient to improve the comparability to the surfaces of bulk materials. It must be emphasised that a larger total thickness of the film alone will not be sufficient to achieve this; it is decisive that bulk-like layers are present below the surface. A crucial point, however, is to find substrates that are compatible with the corresponding bulk structures not only in terms of the lattice constant, but also for a smooth interface to the bulk structure.

Before concluding this Chapter, we should briefly return to the outset. The original motivation for studying ultrathin epitaxial films in the context of heterogeneous catalysis was to learn more about the relevant microscopic properties of catalysts from simple model systems. We have then seen in this Chapter that epitaxial monolayer films are neither simple, nor are they representative models for the surfaces of the bulk oxides. However, even the ideal surface of a bulk oxide itself differs drastically from a real support material's surface and may lack decisive features. Of crucial importance is the influence of the gas phase. Under atmospheric pressure, the surface will be completely covered by adsorbates, in particular by hydroxyl groups created by the dissociation of water. Apart from defects, hydroxyl groups

are a second type of chemically reactive surface species, in particular for proton transfer reactions. Likewise, they can act as anchors for metal ions by exchanging the hydroxyl proton for the metal ion. The presence of water can also modify the thermodynamical stability as has been demonstrated for the  $\alpha$ -alumina (0001) surface [125]. While the dry surface has a positive surface energy (in agreement with our freestanding films), the presence of hydrogen at the surface makes the surface energy negative.<sup>16</sup> We note that the density of hydroxyl groups on a crystalline film can differ substantially from that of a realistic surface under the same conditions: A film grown in the absence of water may not be very reactive when exposed to water afterwards [130], while a real surface is usually created in the presence of water and will most likely immediately incorporate it into the surface.

For the chemical reactivity of a surface under real conditions, point defects play an important role. Their stability and reactivity however depends crucially on the charge state, which is determined by the Fermi energy of the system. For film systems, the metal substrate defines the Fermi energy. In the absence of a metal, i.e. in real catalysts, the Fermi energy depends on the defects and dopants (the “dirt”) and is thus influenced by the treatment of the sample. This implies that the substrate of the epitaxial films in the surface-science approach fixes a possibly important degree of freedom. In addition, the defect levels themselves may be modified by the structure of the epitaxial film as we have discussed in Sec. 4.3.4.

Bearing these restrictions in mind, the epitaxial films are invaluable for improving our understanding of heterogeneous catalysts and related surface phenomena. Experiments on thin epitaxial films allow us to take a much more detailed look at insulator surfaces than would otherwise be possible at the moment. Theory then offers a complementary view on the experimental observations and how they relate to the true systems of interest.

---

<sup>16</sup>Thermodynamically, the crystal should thus split into small particles with hydroxyl-covered surfaces. In practice, this reaction runs at negligible speed and  $\alpha$ -alumina is very inert towards water [94].

# Chapter 5

## Excited-state perspectives of ultrathin films

In the previous chapter we have investigated the role of the finite thickness and the substrate for ground-state properties of thin oxide films and how they influence the film's electronic structure. However, new physical effects come into play for excited states that are not visible from the DFT ground-state perspective. These are contained in the self-energy  $\Sigma$  that connects the DFT-KS electronic structure to the quasiparticle spectrum observed in a photoelectron spectroscopy experiment as described in Sec. 2.3. Here we use the  $G_0W_0$  approximation for  $\Sigma$  introduced in Sec. 2.3.2. We will show in this chapter that the  $G_0W_0$  self-energy corrections for thin films differ from the bulk corrections mostly by polarisation effects that become important when a charged excited state is created in a dielectrically heterogeneous environment. Insulator films supported on a polarisable material, e.g. metal-supported oxide films, are prime representatives of this material class. We therefore expect that the substrate-induced image potential changes the photoelectron spectra and other charged-state properties beyond the ground-state effects discussed so far.

Unfortunately, reliable  $G_0W_0$  calculations for metallic slab systems are not yet possible in the  $GW$  space-time method.<sup>1</sup> We have therefore chosen to study these effects for a prototypical insulator/semiconductor interface: NaCl on Ge(001). This system has been intensively studied with LEED [9, 29, 131, 132], SXRD [132], UPS [9], ELS-LEED [133], and STM [29]. In contrast to the metal-supported oxide films discussed in the previous Chapter, the thickness of the NaCl films on Ge(001) can be varied because the films assume

---

<sup>1</sup>Metallic systems require a different treatment of the time/frequency dependence in the space-time approach, which is currently under development [32].

the rocksalt structure of the bulk materials. We will demonstrate that also the  $G_0W_0$  corrections are thickness-dependent. Most experiments, however, focussed on the surface properties of thick films (10–20 ML) [134–136] and no thickness-dependent measurements for thin films are available to date. For reference, the effects in free-standing NaCl slabs will be discussed, too. Before we present our results, we will briefly sketch out the type of  $G_0W_0$  effects that are expected for thin films beyond the scissor shifts known from bulk systems [23, 56].

The dominant effect is introduced via the screened interaction  $W_0$  and the long-range screening effects contained therein. We have already seen in Sec. 3 that the long-range screening and the image potential arising from dielectric discontinuities play a crucial role in practical  $G_0W_0$  calculations for slab systems, where we have regarded them as artifacts of the periodic boundary conditions. Here we will use the same dielectric models to analyse the effect of the intrinsic dielectric steps at the surface of the thin film and its interface to the substrate. Using simplified models like this, Delerue *et al.* have previously investigated image potential effects for free-standing nanoparticles, wires, and films [137]. We will show that free-standing and supported films behave differently, although the physical basis of the effect is the same. Isolated films have two equivalent surfaces. The image potential induced by these surfaces will then shift all states of one type into the same direction: downwards for the valence states, upwards for the conduction states. Only the magnitude of the shifts might differ between the states, but we will see that this is not significant for NaCl. In a supported film, on the other hand, the image potential is asymmetric and will introduce a strong position-dependence of the self-energy that varies from one side of the film to the other.

Image effects are well-known in the context of core-level spectroscopy for atoms in the vicinity of dielectric interfaces e.g. in  $\text{SiO}_2/\text{Si}$  [138, 139]. They have also been discussed for the valence electron spectra of metal-supported oxide films of varying thickness [21, 140], but reliable theoretical calculations for a quantitative comparison are not available. The first  $G_0W_0$  calculation that clearly demonstrated an image potential effect for an adsorbate was performed only very recently for a benzene molecule on graphite [31], where a drastic reduction of the molecular gap was found. As we will show in the following, the image effects can induce additional shifts of  $\sim 0.5$  eV for NaCl/Ge, thereby modifying the shape and the peak position of the electronic DOS.

The remainder of this chapter is organised as follows: first, we will discuss the free-standing NaCl films. Then we will turn to the supported NaCl films on a Ge(001) substrate. This system has not been theoretically investigated

before, in particular no atomistic structural model of the interface to Ge is available. We will therefore present our atomistic model that we developed from the experimental knowledge and DFT-LDA calculations, before we discuss the quasiparticle spectrum of ultra-thin films as well as its thickness dependence. Finally, we will use the experience gained in this section to reconsider screening effects in the electronic structure of the silica film on Mo(112), that had been discussed in Section 4.3.3 from the DFT-LDA perspective only.

## 5.1 Freestanding slabs

In this first part, we will investigate the  $G_0W_0$  quasiparticle spectrum of free-standing NaCl films. Although such films cannot be prepared experimentally, they provide a simple system for illustrating the changes in the self-energy when going from a bulk material to a thin film. We thereby extend our study of the thickness dependence for free-standing films (Sec. 4.4) to excited-state effects. We use NaCl here instead of the oxides in order to compare to the supported NaCl films presented in the next Section. We will show in the following that the band gap of free-standing ultra-thin NaCl films is larger than the corresponding bulk band gap. The increase of the band gap due to a surface contribution to the self-energy has been predicted from model calculations [137]. Mainly long-range screening effects are responsible for this, and we will show that a dielectric model can semi-quantitatively reproduce the results from our *ab initio*  $G_0W_0$  calculations. However, we find that the model calculations overestimate the effect systematically.

For our calculations, we consider free-standing NaCl slabs with a  $(1\times 1)$  (001) surface containing two to five NaCl layers, using the theoretical bulk lattice constant to define the surface unit cell. As shown in Sec. 3.3.3, we correct for the finite vacuum effect in  $G_0W_0$  calculations. We then find that a vacuum thickness of  $10 \text{ \AA}$  is sufficient to decouple the films. The films are fully relaxed in DFT-LDA, but the relaxations are small, as is well known for the (001) surface of sodium chloride [141]. The Na ions relax inward by  $0.046 \text{ \AA}$  (1.7% of the inter-layer spacing) and the Cl ions outward by  $0.040 \text{ \AA}$  (1.4%) for films with 5 or more layers. Thinner films show slightly deviating absolute values, but the same sign (e.g. 2 layers: Na -2.7%, Cl +2.2%). The values for the Cl relaxation agree well with experiment, but those for the sodium relaxation are smaller than the experimental results from LEED (Na -2.9%, Cl +1.4% [141]). However, the surface relaxation does not play any role for the  $G_0W_0$  corrections. The quasiparticle energies presented in the following include all corrections and extrapolations discussed in Sections



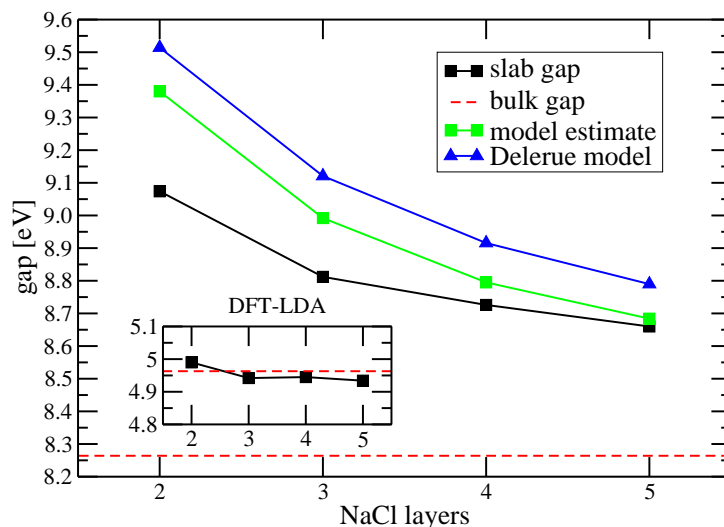


Figure 5.1: Quasiparticle gap of free-standing NaCl slabs of varying thickness. The slab data is corrected for the bcut error (cf. Section F.3.3). The model estimate refers to the image potential at the centre of the slab, see text.

3.3.2, 3.3.3, and F.3.3.

The electronic DOS at the LDA level is indistinguishable from bulk NaCl at all thicknesses, and we therefore refrain from showing it here. The observation that the electronic structure is not affected by the thickness is easily explained. As mentioned above, the atomic structure deviates only very little from the ideal bulk positions and does therefore not introduce significant changes in the electronic structure. Furthermore, the localised character of the valence states makes them insensitive to the slab thickness. The delocalised, dispersive conduction band on the other hand hybridises with a surface resonance (cf. Fig. 5.5), pinning the gap at about 4.94 eV close to the bulk gap at 4.96 eV. When  $G_0W_0$  corrections are applied, this invariance of the gap to the thickness is removed (cf. Fig. 5.1). At a thickness of two layers, the quasiparticle gap is 0.8 eV larger than in the bulk. With increasing thickness, it slowly reduces towards the value of the bulk band gap. The other states behave similar to the band edges states, i.e. the main effect of the  $G_0W_0$  correction is a scissor shift of the bands. This is well known from bulk  $G_0W_0$  calculations [23, 62]. Other significant changes besides this thickness-dependent scissor shift are not observed.

We will now show that the change of the dielectric constant at the surface and the resulting image potential is responsible for the thickness dependence of the quasiparticle gap. For this purpose, we compute the image potential for a homogeneous model dielectric slab (thickness  $s$ , dielectric constant  $\epsilon$ ).

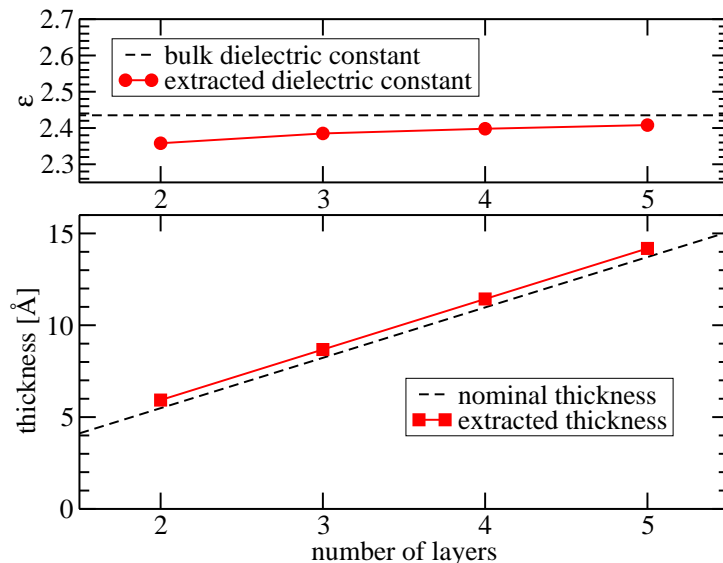


Figure 5.2: Model parameters extracted from the dielectric tensor of the repeated-slab systems.

This model is the basis of Delerue’s “surface contribution to the self-energy” [137], given by

$$\Delta = \frac{2}{\varepsilon s} \left( 0.21883 \frac{\varepsilon - 1}{\varepsilon + 1} + \ln \frac{2}{\varepsilon + 1} \right), \quad (5.1)$$

which corresponds to the expectation value of the image potential for a sine model function. Alternatively, we employ the value of the image potential in the slab centre to estimate the magnitude of the image potential effect. The necessary parameters, i.e. the slab thickness  $s$  and its dielectric constant  $\varepsilon$ , are extracted from the dielectric tensor of the repeated slab system as described in Section 3.3.3. The extracted parameters are shown in Fig. 5.2. The dielectric constant depends only weakly on the slab thickness and is close to its bulk value. Also the effective thickness is in very good agreement with the nominal thicknesses of 2.74 Å per NaCl layer. We thus find that the computed model parameters are very close to intuitive estimates, indicating that the dielectric properties of the slabs are bulk-like.

In Fig. 5.3 we show the image potential computed from the parameters for 2–5 NaCl layers. The image potential is positive inside the slab and becomes smaller when the thickness is increased. At the surfaces, the image potential diverges, which is an artifact of the dielectric discontinuities in the step profile chosen as model here. A smoothly varying dielectric constant yields a smoothly varying image potential. Since the image charge method that we have developed (cf. Sec. B.1) allows us to use arbitrary profiles

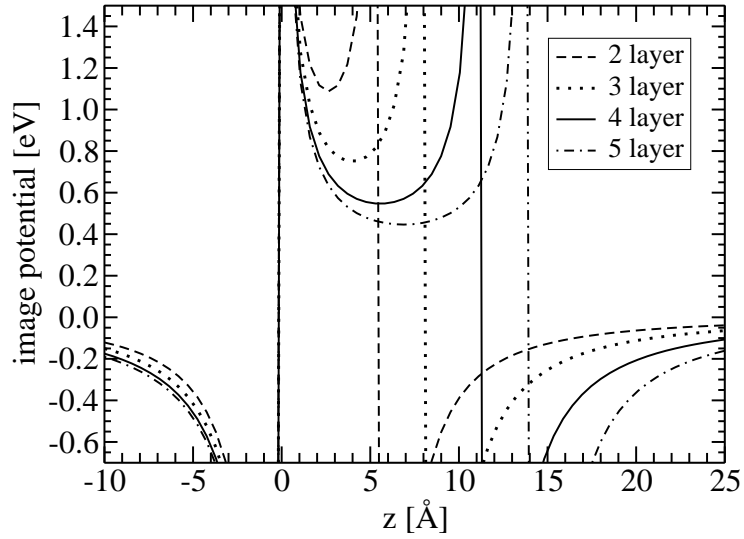


Figure 5.3: Image potential in the dielectric slab models for freestanding NaCl slabs of varying thickness.

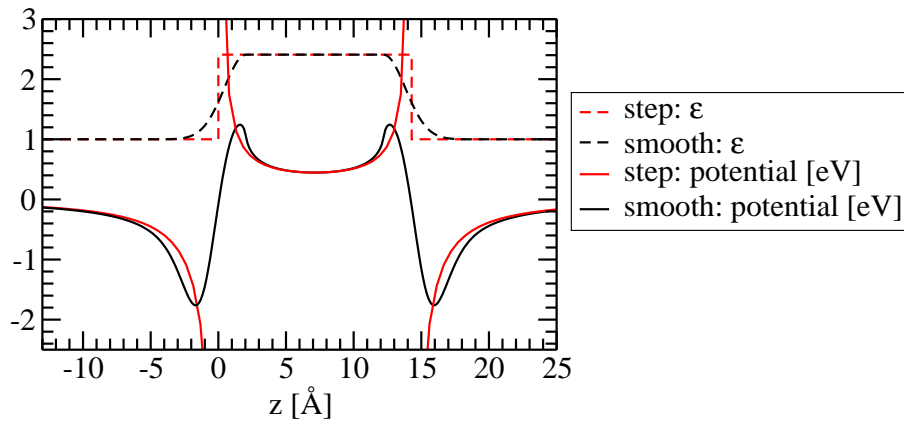


Figure 5.4: Comparison of computed image potentials for a discontinuous step profile and a smooth, Gaussian-derived profile (see text). The parameters correspond to a 5-layer NaCl slab.

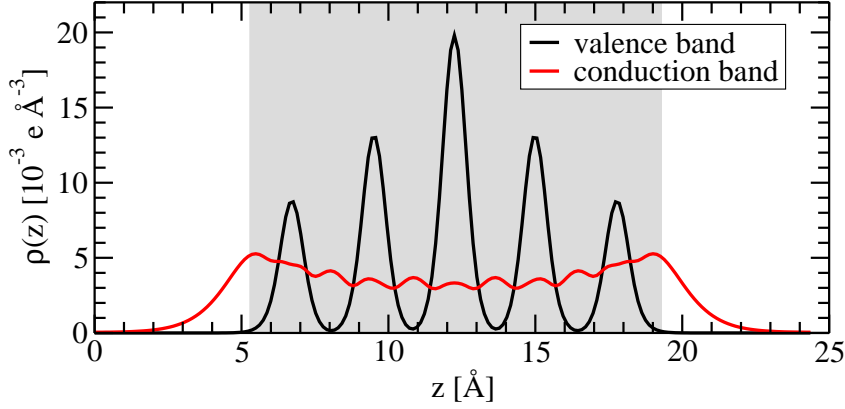


Figure 5.5: Partial density of the valence band maximum and conduction band minimum for a 5 NaCl layers. The grey region indicates the size of the dielectric model, showing that the conduction states are not confined within it.

$\varepsilon(z)$ , we can investigate the influence of the transition between the slab and the vacuum. This is demonstrated in Fig. 5.4, where we compare the image potential of the step profile with that of a smooth profile obtained by introducing a Gaussian-shaped transition region

$$\varepsilon(z) = \begin{cases} 1 + (\varepsilon - 1)e^{-\pi/4(z/t-1)^2} & z < t \\ \varepsilon & t \leq z \leq s - t \\ 1 + (\varepsilon - 1)e^{-\pi/4((z-s)/t+1)^2} & z > s - t \end{cases} . \quad (5.2)$$

The broadening of the Gaussian is controlled by the transition width  $t$  (here:  $t=2 \text{ \AA}$ ) and the prefactor ensures that the average dielectric constant agrees with the step model. We emphasise that this profile serves only as an example for a smooth profile and is not necessarily more realistic than the step profile. We also note that the concept of a local dielectric constant is not well defined. We therefore restrict ourselves to simple models consistent with the dielectric tensor. In Fig. 5.4, we compare the two resulting image potentials for a 5-layer slab. The Gaussian profile yields a continuous image potential in the transition region. The divergences of the step profile is replaced by minima and maxima. In the vicinity of these, the potential is slightly enhanced over that of the step profile. Away from the transition region, the image potentials for the two profiles agree with each other very well. We will therefore restrict our further discussions to the simpler step profiles and avoid to include the image potential in the transition region as far as possible.

Focussing on the average shift  $\Delta W$  of the potential inside the slab, we can use the finding of Sec. B.2 to estimate the change in the quasiparticle gap:

$\Delta W$  acts as an additional scissors shift. To estimate  $\Delta W$  semi-quantitatively and to avoid the divergence at the surface, we take the value at the slab centre and assume that the variations at the surface cancel out to a large degree. This is justified by the observation that at least the conduction states are not fully confined inside the model slab thickness (cf. Fig. 5.5), but extend out into the vacuum where the model image potential is negative. This approach differs from the Delerue model [137], where a sine model function confined to the slab is used to compute the expectation value of the image potential inside the slab. The Delerue model therefore predicts *larger* effects than our slab-centre method. The resulting estimates for the band gap of the NaCl slabs have been included in Fig. 5.1. Taking into account the rather crude approximations involved, they show a reasonable agreement (within about 30% for the slab-centre method or 50% for the Delerue model) with the observed increase in the quasiparticle gap.

We conclude that the change in the dielectric constant when going from NaCl to the vacuum induces an image potential inside the NaCl slabs for charged excitations. The quasiparticle gap of the slab reflects the modified screening and increases compared to the bulk value. These image-potential effects are part of the  $G_0W_0$  self-energy, but since they are due to the long-range correlation, they are absent from the DFT-LDA exchange-correlation potential. Moreover, a simple dielectric model is sufficient to reproduce the results of the full *ab initio*  $G_0W_0$  calculation to within 30%. It can thus be used to estimate the magnitude of image-potential effects in the quasiparticle energies, but it cannot replace the full calculation if a quantitative answer is required.

## 5.2 NaCl films on Ge(001)

In this section the quasiparticle spectrum will be discussed for the realistic case of NaCl films supported on Ge, which can be prepared and investigated experimentally. We therefore briefly summarise what is experimentally known about this system (Sec. 5.2.1). Then we present the atomistic model derived from the experimental results using DFT-LDA calculations (Sec. 5.2.2). The  $G_0W_0$  calculations will be discussed in Sec. 5.2.3.

### 5.2.1 Experimental situation

Supported sodium chloride films are prototypical systems for studying the properties of thin insulator films [9, 10, 29, 131, 142]. NaCl films are mostly grown on metallic substrates, e.g. Al [142] or Cu [10], but for the  $G_0W_0$

calculations we will focus on a semiconducting substrate: Ge(001). Well-ordered, crystalline NaCl films with a thickness of a few monolayers (ML) can be grown on Ge(001) with high quality [9], since the lattice mismatch between Ge (5.66 Å) and NaCl (5.63 Å) is only 0.5%. Correspondingly, the [100] and [010] directions of the substrate and the film are aligned. Films of varying thicknesses up to several nanometer can be grown by sublimation of NaCl onto a clean Ge(001) surface kept at 100 K, followed by annealing at 400 K for a few minutes. From Auger electron attenuation [9] and later STM measurements [29] it was concluded that initially a double-layer film forms on which further single layers grow layer by layer. While STM pictures for a 2ML film exhibit a NaCl  $1 \times 1$  pattern at the surface [29], LEED and SXRD experiments show a  $p(2 \times 1)$  unit cell corresponding to the typical buckled-dimer reconstruction of the Ge(001) surface [9, 29, 132]. It has therefore been concluded that the Ge dimers remain intact under the film, supported by ELS-LEED which detects the Ge surface states even when a 20 ML NaCl film is grown on top [133].

The interaction of the sodium chloride film with the Ge(001) substrate is rather weak. This leads to the “carpet” growth mode over mono-atomic steps of the substrate, where the NaCl film smoothly covers the substrate [29, 131]. From the extent of the transition region, Schwennicke *et al.* have estimated a binding energy of 0.13 eV per Ge atom [131]. The atomic structure of the interface between Ge and NaCl is not known from experiment. To explain their results from surface X-ray diffraction, Lucas *et al.* suggested that an additional sodium atom is present in the surface unit cell [132], but a detailed structural analysis was not performed. On the other hand, Ernst *et al.* have found that preadsorbed sodium atoms break the Ge dimers and modulate the structure of an adsorbed NaCl film up to about 10 ML thickness [143]. We therefore consider the presence of additional sodium atoms in the unit cell unlikely for films deposited on the clean Ge(001) surface as they do not show this modulation.

For films of two to three ML thickness, STM experiments can be performed with atomic resolution in a small energy window of 1.5 eV to 2.7 eV tip bias [29]. This energy range lies inside the gap of the NaCl overlayer, i.e. no states with dominant NaCl character contribute to the tunnelling. Nevertheless, the STM pictures must reflect the structure of the surface of the NaCl film as they show a  $1 \times 1$  square pattern with a height corrugation of 0.5 Å. Furthermore, only one type of ion is imaged. Glöckler *et al.* have concluded that the tunnelling is from Ge bulk and interface states that are “modulated” by the NaCl overlayer. They have speculated that variations in the local work function could be responsible for the modulation and that  $\text{Na}^+$  ions are imaged as the bright spots. The STM experiments have also

confirmed the carpet overgrowth of mono-atomic steps and the film thickness deduced from Auger electron spectroscopy. The apparent thickness of a single NaCl layer as measured in STM is  $2.0 \pm 0.3 \text{ \AA}$ , somewhat smaller than the geometrical thickness of  $2.8 \text{ \AA}$ , but such a difference is not unusual for insulator films [144].

From the experimental investigations, the following questions remain open:

1. The atomic structure of the interface between Ge and NaCl,
2. the observed preference for the  $p(2 \times 1)$  surface unit cell,
3. the preference for an initial double layer, and
4. the source of the STM contrast.

We will address these questions in the following using DFT-LDA *ab initio* simulations. The surfaces are simulated in the repeated slab approach. The Ge substrate is modelled by 6 layers, saturated with hydrogen at the bottom. The bottom two layers were kept fixed during the relaxations. We verified the reliability of this approach by comparing to a 12 layer Ge slab, which yields essentially identical structural parameters and adhesion energies. A  $p(2 \times 1)$  surface unit cell with the experimental lattice constant ( $5.66 \text{ \AA}$ ) was used. The separation of neighbouring slabs was  $\sim 9 \text{ \AA}$  between the outermost atoms which proved to be sufficient to avoid interactions across the vacuum. The use of asymmetric slabs in the repeated-slab approach introduces an artificial electric field across the slab when the work function of the two surfaces differ. A dipole correction was used to correct for this [78].

### 5.2.2 Atomic structure of the interface

The identification of the most stable interface is described in Sec. D.3. The resulting atomic structure for a 2ML NaCl film is shown in Fig. 5.6. We find that a negatively charged chloride ion is situated above the lower (“down”) atom of the Ge dimer, while a sodium ion bridges two “up” dimer atoms. This suggests that electrostatic interactions may be responsible for the observed geometry, since the *down* (*up*) atom is expected to carry a positive (negative) partial charge. We have therefore computed the electrostatic potential of the clean Ge surface at the geometry of the adsorbed NaCl layer (cf. Fig. 5.7). The shape of the electrostatic potential agrees qualitatively very well with the adhesion energy line scans presented in Fig. D.5. It shows a strong corrugation along the dimers, and a much flatter one in the perpendicular direction (Fig. 5.7 a). The electrostatic potential also helps to understand the atomic

relaxation in the bottom layer of the NaCl film (Fig. 5.7 b). The chloride ion above the *down* Ge atom relaxes towards this atom, in agreement with the very steep electrostatic potential in this region. Likewise, the sideward relaxation of the second chloride ion in the bottom NaCl layer is driven by the electrostatic repulsion of the occupied dangling bond state. The preference for the Cl-*down*/Na-*up* combination explains why only a  $p(2 \times 1)$  pattern is observed in LEED and SXRD. Other dimer patterns such as  $c(4 \times 2)$  would inevitably include the opposite combinations, which would be strongly disfavoured due to the electrostatic repulsion. Interestingly, also the adhesion of NaCl films on Cu(311) surfaces is effectuated by electrostatic interactions between the ions in the film and charge modulations at the substrate surface, which are in this case caused by the Smoluchowski smoothening at the corrugated Cu(311) surface [10]. This implies a surprising similarity between the NaCl/metal and NaCl/semiconductor interface in this case, which is further supported by the carpet overgrowth of mono-atomic steps for both cases.

The adsorption of the NaCl layers influences the Ge dimer, too. While the dimer is tilted by  $19^\circ$  at the clean Ge surface, this angle reduces to  $10^\circ$  when NaCl is adsorbed with Cl on top of the *down* atom<sup>2</sup>. The *down* atom relaxes upwards by  $0.24 \text{ \AA}$  and the *up* atom relaxes downwards by  $0.12 \text{ \AA}$ . The dimer bond length increases slightly by 1%. The redistribution of the electrons upon the adsorption of the NaCl overlayer was investigated by the density difference

$$\Delta\rho = \rho(\text{Ge} + \text{NaCl}) - \rho(\text{Ge}) - \rho(\text{NaCl}) , \quad (5.3)$$

shown in Fig. 5.8 across the dimer. The density is increased between the chloride ion and the *down* Ge atom as well as between the sodium ions and the *up* Ge atoms. The ionicity (charge separation) in the Ge dimer appears to be slightly increased, too. The bonding characteristics is probably best described as a weak Lewis acid/base complex, an intermediate form between an ionic and a covalent bond [94].

We have also simulated the STM experiments in the Tersoff-Hamann approximation, reported in Sec. D.4. In summary, we find strong evidence that the tunnelling is – as suggested – from the Ge states. These hybridise with the chlorine  $3p$  orbitals of the film, and the bright spots in the STM pictures must then be assigned to the Cl ions and not – as suggested in [29] to Na. We determined the average apparent heights of the NaCl layers from our simulation and find good agreement with experiment. We are however not able to reproduce the experimental corrugation laterally resolved STM pictures. We find a  $(2 \times 1)$  pattern with an apparent height difference of  $0.3 \text{ \AA}$

---

<sup>2</sup>When Na is adsorbed on top of the *up* atom, the dimer angle becomes  $15^\circ$ .



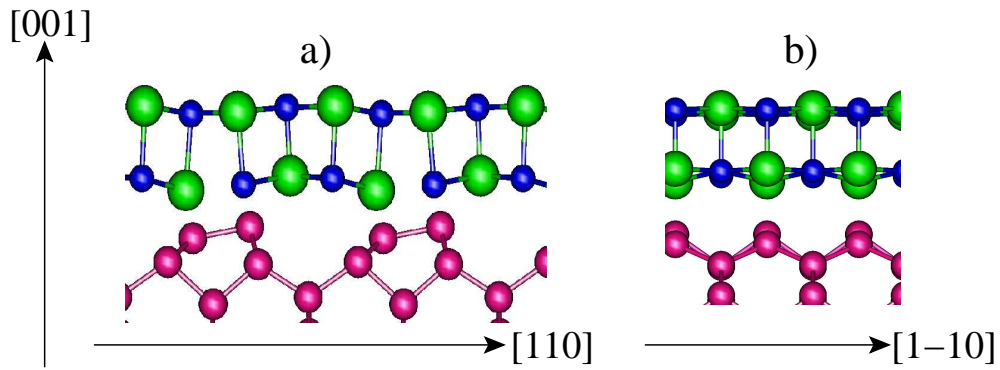


Figure 5.6: Side views of the optimised geometry for the 2 ML NaCl film on Ge(001) a) along the dimers and b) perpendicular to them. The colour coding is as follows: Ge (magenta), Na (blue), and Cl (green).

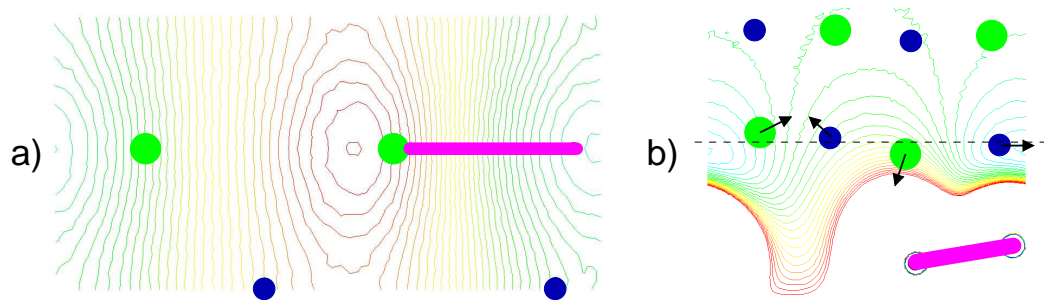


Figure 5.7: Electrostatic potential maps for the clean Ge surface (in the Ge/NaCl geometry). a) Top view at the average height (see dashed line in b) of the bottom NaCl layer (spacing 0.02 V). b) Side view through the dimer (spacing 0.05 V). The potential increases from blue-green to red. The circles indicate the atomic positions of the NaCl overlayer, the bar the Ge dimer (colours as in Fig. 5.6). The arrows indicate the direction of the electrostatic forces that drive the relaxation in the bottom layer.

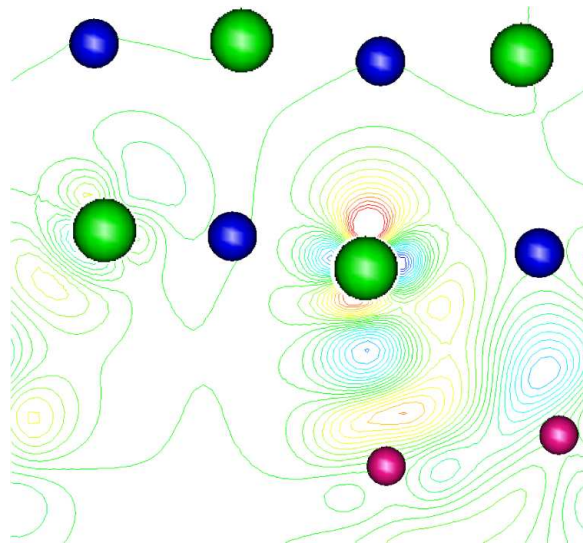


Figure 5.8: Electron density difference map across the Ge dimer. Red indicates a reduction in density, blue an increase. The spheres indicate the atom positions (colours as in Fig. 5.6).

between the bright features. Glöckler *et al.* report a  $1 \times 1$  pattern in the STM. The discrepancy might be due to limitations in the theoretical modelling of the STM (cf. Sec. D.4) or to the limited resolution in the experiment from 1996 [29]. New STM experiments might help to clarify this issue.

For thicker films we used the interface model from the 2ML film as the starting geometry for the relaxation. In general, we find that the buckling in the NaCl film visible in Fig. 5.6 a) is reduced when further layers are added. As a measure for the relative stability of the films, we show the formation energy per NaCl formula unit from bulk NaCl with the Ge lattice constant (other lattice constants would only lead to a constant offset) in Fig. 5.9. The stability of the film increases monotonously with increasing thickness. This suggests that the thermodynamically favoured structures are three-dimensional islands rather than flat films. The experimentally observed preference for flat films must therefore be attributed to kinetic limitations. This appears plausible since the thin films are annealed at only 400 K [9], well below the melting point of NaCl (1074 K, [94]). The observation that the substrate must be kept at very low temperatures during the deposition [9] supports this conclusion, too.

In summary, we have developed an atomistic model of the NaCl/Ge interface. The Ge dimers remain intact and buckled under the film as suggested from ELS-LEED. The chloride ions are found to adsorb on top of the *down*

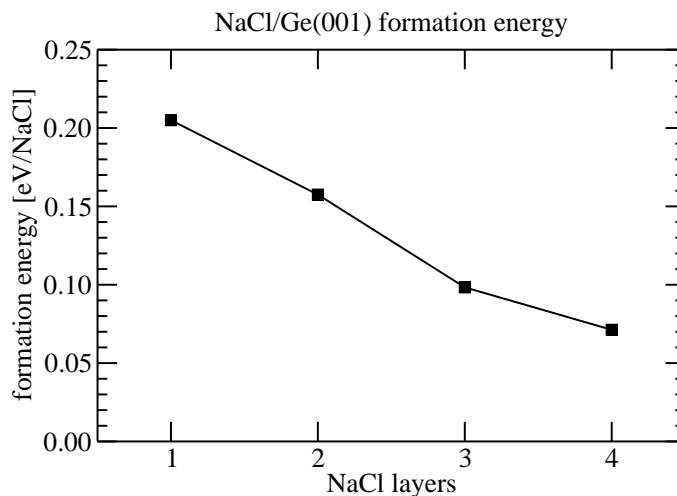


Figure 5.9: Formation energy per NaCl unit of the adsorbed NaCl film on Ge(001) as a function of the film thickness.

atom of the Ge dimer. This film geometry is stabilised by the electrostatic attraction between the ions in the film and the partially charged dimer atoms. When the overlayer is shifted perpendicular to the dimer axis, the adhesion energy changes only very little. Since the energy differences involved may be below the accuracy of the present DFT-LDA calculations the position of the film along  $[1\bar{1}0]$  cannot be determined without doubt, but also in experiment this position might not be fixed. We will focus in the following only on the most stable interface found in the DFT-LDA calculations, where one Cl atom is located on top of the *down* atom of the buckled Ge dimer.

### 5.2.3 $G_0W_0$ calculations

We now turn to the results of the  $G_0W_0$  calculations, the main topic of this chapter. Since experimental spectroscopy data for ultrathin films is not available, we will mainly focus on the physical effects and trends visible in our  $G_0W_0$  calculations. We will briefly address the comparison to the relevant UPS studies for thick films later. We note that the  $G_0W_0$  calculations are numerically converged to only 0.1 eV to 0.2 eV in the quasiparticle energies. To improve the accuracy, corrections must be included to account for

- the under-convergence in the band cutoff (Sec. F.3.3),
- the finite  $\mathbf{k}$ -point sampling (Sec. 3.3.2), and
- the finite-vacuum effect (Sec. 3.3.3).

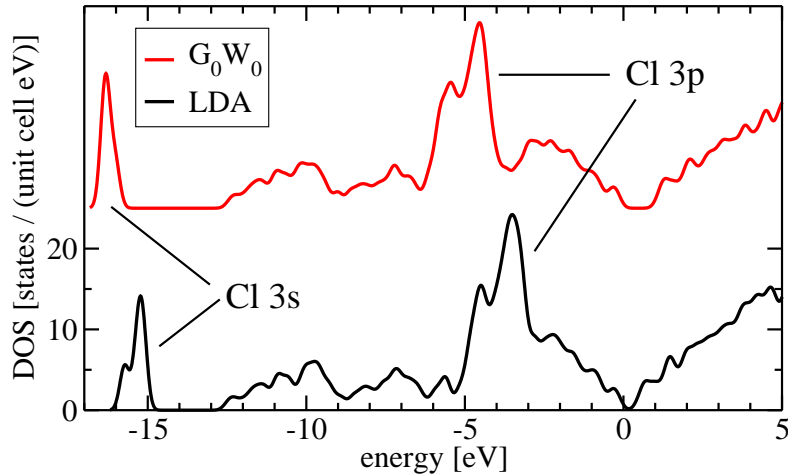


Figure 5.10: DOS for the supported 2ML NaCl film on Ge(001) at the DFT-LDA level and the  $G_0W_0$  level. The slab valence band maximum defines the energy zero.

However, all three corrections mainly affect the absolute band position, and have almost the same value for all NaCl thicknesses considered. For simplicity, we will not include these corrections when we discuss the effects within the NaCl valence bands. We estimate that the relevant differences in the  $G_0W_0$  corrections within the Cl 3p and 3s bands are obtained with an accuracy of 0.05 eV–0.1 eV.

We have also computed the  $G_0W_0$  quasiparticle band structure for the Ge bulk and the clean Ge surface. The quasiparticle gap corrections computed for bulk Ge ( $\sim 0.7$  eV,  $\mathbf{k}$ -dependent) agree reasonably with previous calculations [23]. For the surface, we obtain a good agreement with the results of Rohlffing [61] within the valence band. The correction to the quasiparticle gap is larger by  $\sim 0.2$  eV when we include the finite-vacuum correction. We attribute this discrepancy to the more accurate treatment of long-range screening in our calculations (cf. Sec. 3).

In Fig. 5.10 we present the total DOS for two NaCl layers on a 6 layer Ge substrate. The DOS is shown for the DFT-LDA and for DFT-LDA with  $G_0W_0$  corrections, denoted as  $G_0W_0$  DOS in the following. Two features can be attributed to the NaCl-derived states: the Cl 3s peak around -16 eV and the Cl 3p states at around -5 eV. The most important change when comparing the LDA and  $G_0W_0$  DOS is the shift of the NaCl bands relative to the Ge states. Such a shift is not surprising since the  $G_0W_0$  corrections in bulk NaCl are much larger (3.3 eV) than for the Ge bulk (0.7 eV). However, also the shape of the NaCl-derived features changes when going

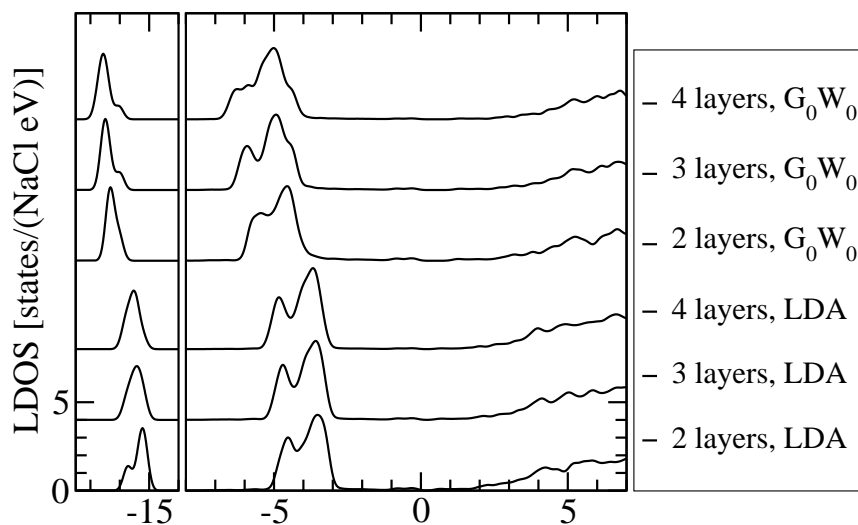


Figure 5.11: LDOS projected on the NaCl film region for supported NaCl films on Ge(001) at the DFT-LDA level and the  $G_0W_0$  level. The slab valence band maximum defines the energy zero.

from DFT-LDA to the  $G_0W_0$  DOS, indicating that the  $G_0W_0$  shifts for NaCl are not uniform. As we will show below, they are caused by the position-dependence of the self-energy that results from the dielectric response of the substrate. To compare NaCl films of varying thickness, we will use the region-projected LDOS described in Sec. C.2. The normalised DOS for 2–4 ML films is shown in Fig. 5.11. While the DOS in the Cl-3 $p$  region is essentially thickness-independent at the level of DFT-LDA, the  $G_0W_0$  corrections introduce a visible thickness dependence in the shape and the width of the NaCl bands. This reemphasises our earlier observation that the bulk corrections are not transferable to thin films. Rather, excited states undergo additional thickness-dependent and substrate-specific changes that are neither visible from a ground-state perspective nor predictable from bulk  $G_0W_0$  calculations for the separate fragments only.

Our first step to approach the unusual  $G_0W_0$  shifts of the supported films will be to prove that the self-energy corrections are position-dependent. However, most of the bands are more or less delocalised over the film, which makes it difficult to directly determine the local self-energy contributions. To extract position-dependent shifts, we employ the atomic orbital projection technique described in Sec. D.1 to determine the localisation of the state. Several states are dominated by the orbitals of only one atom. When the  $G_0W_0$  corrections of the bands are plotted against this projection, we obtain a linear dependence once the contribution from a single atom exceeds

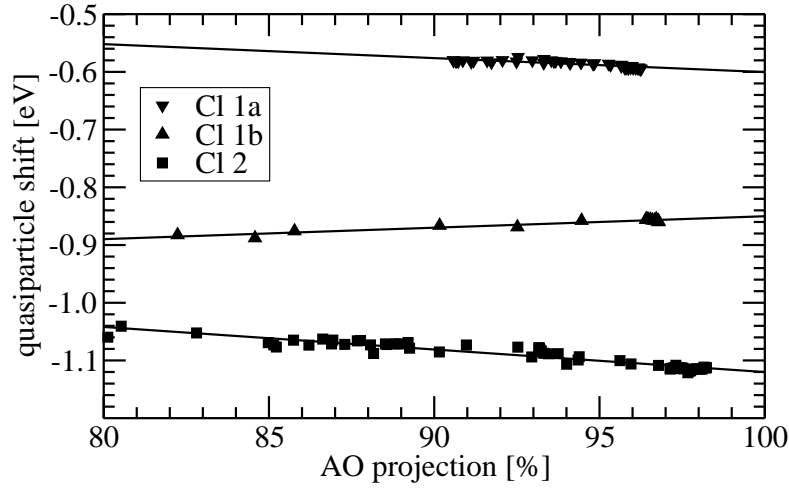


Figure 5.12: Extraction of orbital-dependent  $G_0W_0$  shifts for the example of Cl 3s orbitals of 2ML NaCl/Ge. For the atom labelling, see Fig. 5.13 a).

ML	Cl 3s					Cl 3p				
	1a	1b	2	3	4	1a	1b	2	3	4
2	-0.60	-0.85	-1.12			-0.79	-1.10	-1.40		
3	-0.54	-0.84	-1.01	-1.06		-0.70	-1.08	-1.29	-1.47	
4	-0.53	-0.83	-0.99	-1.06	-1.16	-0.76	-1.04	-1.28	-1.33	-1.52

Table 5.1: Extrapolated  $G_0W_0$  shifts (relative to the Ge vbm) for chlorine orbitals in supported NaCl films (see text). The position of the Cl atoms is given in layers, where 1 is closest to the Ge. The first layer contains Cl ions in different position, on top of the Ge dimer (1a) and between two Ge dimers (1b), cf. Fig. 5.13 a).

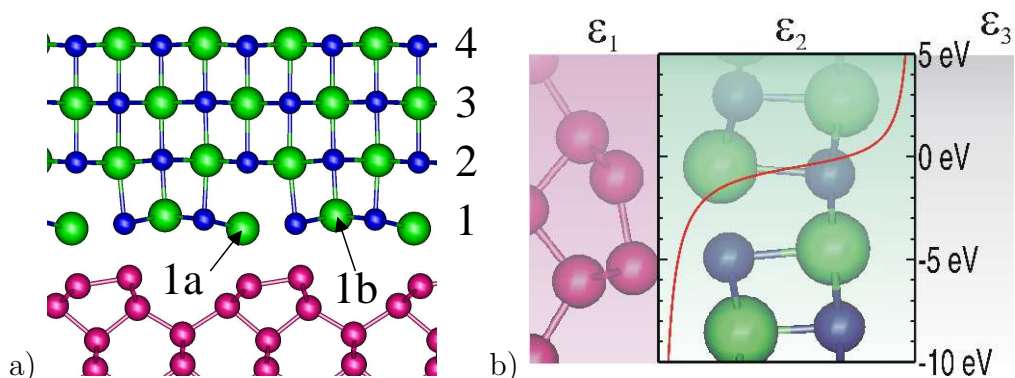


Figure 5.13: a) Cl positions in the supported 4 layer NaCl film for the position dependence of the  $G_0W_0$  shifts. b) The dielectric model for a 2ML slab (rotated by  $90^\circ$ ): the Ge ( $\epsilon_1$ ), NaCl ( $\epsilon_2$ ), and vacuum ( $\epsilon_3$ ) regions are shaded. The red line shows the image potential.

$\sim 80\%$  (cf. Fig. 5.12). These orbital-dependent  $G_0W_0$  corrections are linearly extrapolated to 100%; the results are listed in Tab. 5.1. The extrapolated shifts in the first layer differ considerably between the two nonequivalent positions 1a (Cl on the dimer) and 1b (Cl between the dimers, cf. Fig. 5.13 a), while the atoms in higher layers behave very similarly. We do therefore not differentiate between them.

The most important trend is that the quasiparticle energies are shifted upward in the vicinity of the Ge substrate – a direct result of the image potential that forms at the interface of the NaCl film ( $\epsilon \approx 2.4$ ) and the Ge substrate ( $\epsilon \approx 14$ ), cf. Fig. 5.13 b). A similar effect, but opposite in sign results from the dielectric step at the surface (vacuum  $\epsilon = 1$ ). We again use a dielectric slab model to compute the image potential within the NaCl film. For this, the Ge substrate and the NaCl films are represented by homogeneous slabs with dielectric constants as stated above. The thickness of the NaCl film was assumed to be  $2.8 \text{ \AA}$  per layer. In Fig. 5.14, we compare the extrapolated shifts from Tab. 5.1 to the image-potential effect computed from the dielectric slab model for the films of 2–4 layers thickness by plotting the following model self-energy for the occupied NaCl states

$$\Sigma(z) = \Delta\Sigma - \frac{1}{2}V(z) \quad (5.4)$$

where we include the correct prefactor for the image potential (cf. Sec. B.2).  $\Delta\Sigma$  accounts for the different corrections to the Ge and NaCl states. To a first approximation,  $\Delta\Sigma$  for the Cl  $3p$  states is estimated from the bulk

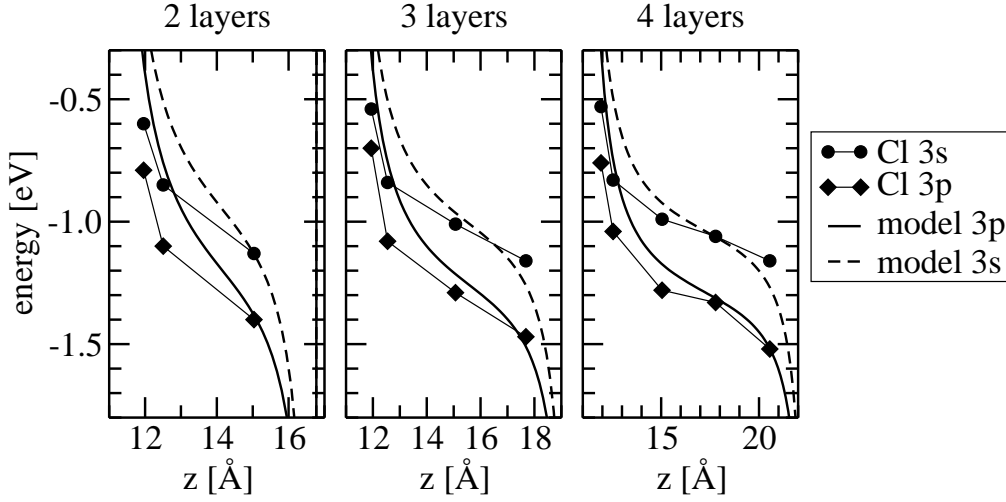


Figure 5.14: Comparison of the extrapolated  $G_0W_0$  orbital shifts to the image potential of the dielectric slab model (cf. text).

corrections of the quasiparticle gap  $\Delta^{\text{gap}}$

$$\Delta\Sigma = \frac{1}{2}\Delta^{\text{gap}}(\text{Ge}) - \frac{1}{2}\Delta^{\text{gap}}(\text{NaCl}) = -1.3 \text{ eV} . \quad (5.5)$$

For the 3s states, we assume  $\Delta\Sigma = -1.05 \text{ eV}$  to account for the different shifts of 3s and 3p states. We find a reasonable, qualitative agreement, highlighting that mainly the dielectric response effects are responsible for the position dependence of the  $G_0W_0$  self-energy.

We now turn to the question of how delocalised states are affected by this self-energy. We emphasise again that the position-dependence was extracted from the  $G_0W_0$  shift of only partially localised states by extrapolating to the “pure” orbitals. We now go the opposite way. The  $G_0W_0$  correction to the LDA energies is obtained in the perturbative approach as

$$\Delta\epsilon_{n\mathbf{k}} = \langle \psi_{n\mathbf{k}}^{\text{LDA}} | \Sigma - V_{\text{xc}} | \psi_{n\mathbf{k}}^{\text{LDA}} \rangle , \quad (5.6)$$

i.e. we freeze the composition of the quasiparticle state at the level of the LDA. Consequently, the  $G_0W_0$  corrections will reflect the position-dependence of the self-energy according to the local weight of the state in the various layers. Most of the states in the Cl 3p valence region are delocalised over the NaCl film and the  $G_0W_0$  shifts lie between -0.7 and -1.4 eV, the shifts for the bottom and top layer, respectively. Applying the  $G_0W_0$  corrections then leads – in addition to the shift with respect to the Ge vbm – to a broadening of the DOS in the Cl 3p region because the corrections vary between the



orbital	2	3	4
$4\sigma$	-2.99	-3.09	-3.14
$\pi_1$	-2.31	-2.36	-2.41
$\pi_2$	-2.18	-2.24	-2.33
$5\sigma$	-1.81	-1.90	-1.96
$\pi_1^*$	+3.05	+3.20	+3.29
$\pi_2^*$	+3.52	+3.58	+3.73

Table 5.2: Quasiparticle shifts (with respect to Ge vbm) for  $p$ -derived orbitals of CO adsorbed on Ge-supported NaCl films with 2–4 layers.

different states (cf. Fig. 5.11). The situation for the Cl 3s states is different: we observe a sharpening of the peak. At the DFT-LDA level, it consists of distinct, but close-lying states associated with the different Cl position. The shift in the peak positions is caused by a gradual increase in the electrostatic potential throughout the film and causes the apparent broadening. This is counteracted by the image potential contained in the  $G_0W_0$  self-energy which has the opposite sign. After applying the  $G_0W_0$  corrections, the peak positions are closer and we find a sharp peak with a small shoulder  $\sim 0.5$  eV higher in energy, which can be attributed to the Cl atom in the bottom layer of the film situated above the dimer. This shift is dominated by the image potential, whereas the electrostatic potential alone would lead to a downward shift (visible for example in the LDA DOS for 2ML in Fig. 5.11). We observe a similar shoulder at the top of the Cl 3p orbitals. The broadening-induced increase in the band width and the additional shoulders should be detectable by photoelectron spectroscopy for ultrathin NaCl films on Ge, but such measurements are unfortunately not available, yet.

The quasiparticle spectrum contains the contribution of the energy dispersion (which, for NaCl, is accurately described within DFT-LDA) and of the image potential. The influence of the latter can be identified best for localised states with a small energy dispersion. The conduction band of the NaCl films consists of highly dispersive, delocalised states that additionally hybridise with the Ge conduction band. We have therefore not succeeded in extracting local-orbital shifts for the unoccupied states. The lowest conduction band of the NaCl film is in resonance with the Ge substrate with a width of  $\sim 0.5$  eV, which also removes the independence of the band energies to the film thickness that we observed for the freestanding films. It is therefore impossible to determine the film's band gap with a sufficient accuracy to detect dielectric response effects.

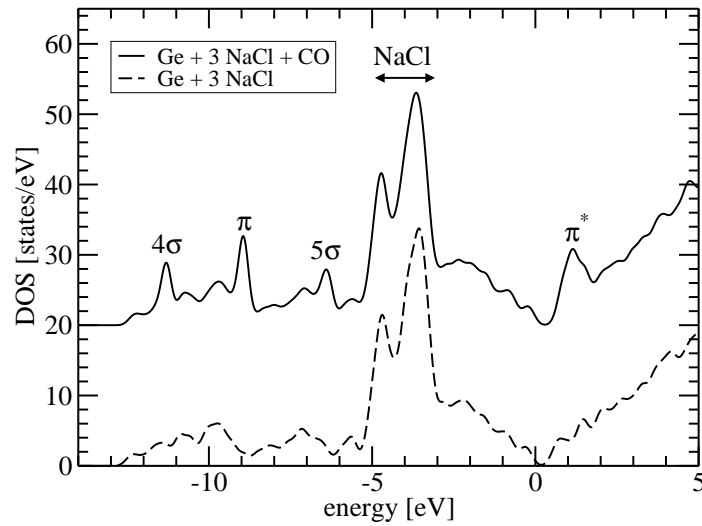


Figure 5.15: LDA DOS for CO adsorbed on a 3 ML NaCl film on Ge. The CO orbitals give rise to sharp peaks.

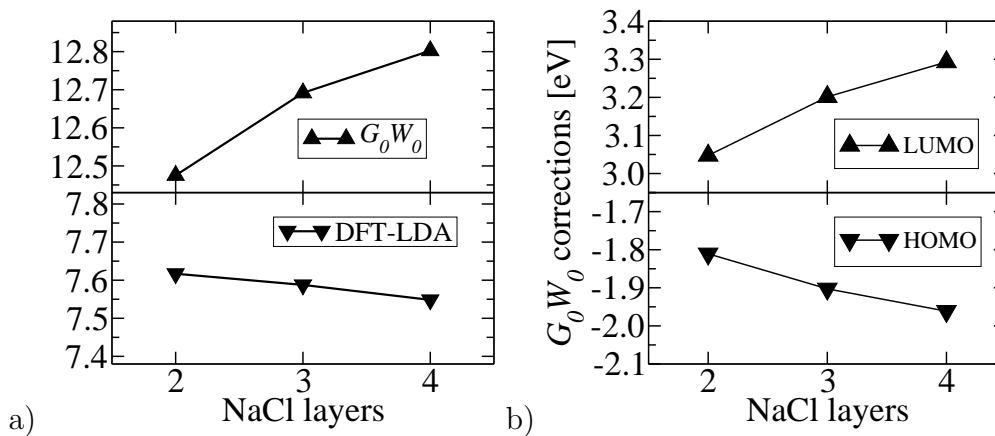


Figure 5.16: a) Molecular gap of CO on NaCl/Ge(001) films at the DFT-LDA and  $G_0W_0$  levels of theory as a function of the film thickness. b)  $G_0W_0$  corrections for the CO highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals.

In order to solve this delocalisation problem for unoccupied states and to demonstrate that not only the film, but also adsorbates on it are affected by the dielectric response of the substrate, we have studied the adsorption of CO on the NaCl films. We note that the CO is the “fruit fly” of surface physics to probe the adsorption and related properties for metal as well as insulator surfaces. We have placed a single CO molecule per  $p(2\times 1)$  unit cell on top of a Na atom with the carbon end pointing towards the surface, which represents the most stable adsorption mode. We have chosen this molecule because it only weakly adsorbs on NaCl and does not introduce a significant perturbation of the NaCl film. In addition, it is a closed shell molecule for which we can perform a  $G_0W_0$  calculation. The CO valence orbitals then serve as a local probe to the image potential. The CO bands are clearly visible in the density of states (cf. Fig. 5.15) and have almost no dispersion in the bandstructure. We have therefore determined the  $G_0W_0$  shifts for the  $\Gamma$ -point only. Already at DFT-LDA, we observe a change of the orbital energies with respect to the Ge vbm when the thickness of the NaCl film is varied. This can mainly be attributed to the change in the work function associated with the thickness change since all CO orbitals are shifted by the same amount.<sup>3</sup> We will therefore concentrate again on the  $G_0W_0$  shifts, shown for films with 2–4 layers in Tab. 5.2. The image-potential effect in the quasiparticle energies amounts to 0.05 eV to 0.15 eV per layer, which is at the limit of the numerical accuracy. We therefore observe a significant numerical noise that prohibits a meaningful quantitative analysis. If a metal were used as a substrate, the image-potential effects would be of course larger. Nevertheless, the qualitative trend is clearly visible. With increasing thickness of the NaCl layer, the occupied CO levels are lowered and the unoccupied levels are raised, thereby increasing the quasiparticle gap (see Fig. 5.16). We thus obtain the same qualitative behaviour for CO on NaCl/Ge as found for benzene on graphite [31]. However, by varying the thickness of the NaCl film the distance of the molecular adsorbate from the Ge substrate can be changed which is impossible for the case of benzene/graphite. Very recently, this effect has been observed in scanning tunneling spectroscopy (STS) for pentacene molecules on NaCl/Cu(111) [6]. For this system, STS allows to determine the molecular HOMO/LUMO gap. For NaCl films of 1, 2, and 3 ML thickness, Repp *et al.* found a pentacene gap of 3.3, 4.1, and 4.4 eV, respectively. This is a considerable reduction compared to the gas phase value of 5.3 eV.

Unfortunately, there is almost no experimental data available for comparison for the NaCl/Ge(001) system, because UPS experiments have focused on the surface properties of thick NaCl films [134]. Possible differences in the UP

---

<sup>3</sup>With the exception of the  $5\sigma$  orbital that is involved in the bonding.

	correction method	correction [eV]		
		Ge	NaCl	NaCl - Ge
finite vacuum	dielectric model	+0.08	+0.12	+0.04
finite Ge slab	dielectric model	-0.14	-0.12	+0.02
<b>k</b> -sampling	extrapolation			-0.06
band cutoff	extrapolation	-0.36	-0.43	-0.07
bulk/slab alignment	electrostatic potential	-0.13		-0.13
sum				-0.20

Table 5.3: Corrections for systematic errors in the numerical  $G_0W_0$  calculations to the absolute band positions.

spectra for thin films have not been discussed. The absolute position of the upper NaCl bands with respect to the Ge vbm was reported to be -4.2 eV in Ref. [134], which is in good agreement with our  $G_0W_0$  calculations. For this comparison, we include a correction of -0.2 eV for the systematic modelling errors that we have neglected so far. These are listed in Tab. 5.3 to demonstrate that only few of them significantly affect the relative alignment of the Ge and NaCl states. We note that these effects are largely independent of each other; the corrections should therefore be additive. The most important ones result from low convergence parameters in the numerical  $G_0W_0$  calculations that can be extrapolated very reliably. The reliability was tested for each correction by increasing the corresponding parameter. Using low parameters in the numerical calculation and correcting for them *a posteriori* drastically reduces the computational effort without affecting the accuracy of the final result.

Although the thickness of the NaCl films on the Ge substrate can be well controlled and has been varied for other techniques, e.g. ELS-LEED [133], thickness-dependent photoemission data have not been published for this system. For other thin insulator systems, however, such studies have been performed, e.g. LiCl/Cu(001) [145], MgO/Ag(001) [21], and Al<sub>2</sub>O<sub>3</sub>/Ru(0001) [140, 146]. In these studies, a variation of the valence band position in UPS [21, 145] or the band gap measured from EELS [140, 146] has been observed upon varying the thickness. It was speculated that image-potential are responsible for this [21, 140, 145]. However, applying our dielectric model to these systems we find only poor agreement for the cases where a quantitative analysis was performed [21, 140]. Compared to the dielectric model, the shifts observed in experiment decay too slowly with increasing thickness and the band width changes comparatively little in contrast to the trend observed in our  $G_0W_0$  calculations. We therefore suggest that other effects, e.g. the

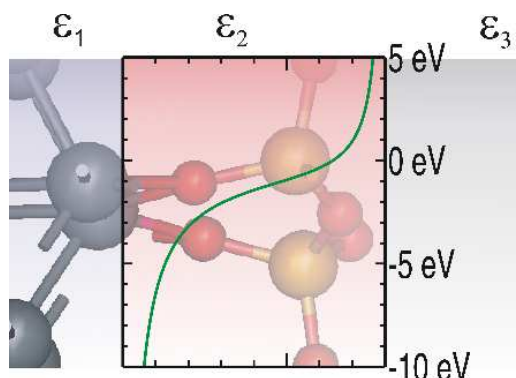


Figure 5.17: The dielectric model for the silica film on Mo(112), rotated by  $90^\circ$ : the Mo ( $\epsilon_1$ ), SiO<sub>2</sub> ( $\epsilon_2$ ), and vacuum ( $\epsilon_3$ ) regions are shaded. The green line shows the image potential.

presence of defects, must be taken into consideration to explain these experiments. On the other hand, NaCl on Ge(001) might be a good model system to experimentally verify the image-potential effects because of the high quality of the films achievable and the possibility to compare to accurate  $G_0W_0$  calculations.

In summary, dielectric response effects are important for a quantitative understanding of the electronic structure of supported insulator films. In contrast to free-standing films, the systematic shifts in the absolute band positions are small because the image effects from the surface and the interface to the substrate oppose each other. Instead, an increase of the band width and a broadening of the peaks in the DOS are observed. The effect of the image potential is most noticeable for (partially) localised states close to the Ge substrate. The Cl atom above the dimer provides such a case, introducing small shoulders at the top of the Cl 3s and Cl 3p bands that should be detectable in the photoelectron spectra for ultrathin films.

### 5.3 Image effects in oxide films

We will now return to the oxide films of the previous Chapter and briefly discuss the implications of our findings in this Chapter for these films. Since they are grown on metallic substrates, the dielectric response effects as demonstrated here for the NaCl films on Ge should be even more pronounced due to the larger difference in the respective dielectric constants.

We indeed find experimental evidence that dielectric response effects play an important role in the photoemission spectra of the silica film on Mo(112).

	linkage	inter-row	on-row
$\Delta V$	-0.2 eV	-0.1 eV	$\pm 0.0$ eV
$\Delta W$	-3.2 eV	-0.7 eV	-0.2 eV
total	+1.4 eV	+0.3 eV	+0.1 eV

Table 5.4: Electrostatic and image potential shifts  $\Delta V$  and  $\Delta W$  at the position of the oxygen atoms for SiO<sub>2</sub> on Mo(112). The electrostatic potential at the on-row atom defines the energy zero. The total effect includes the correct prefactor ( $-\frac{1}{2}$ ) for the image potential.

In the crystalline film the O 1s core levels of the surface and linkage oxygens as measured by XPS were found to differ by 1.3 eV [13, 105, 108]. This difference was reproduced within DFT [108] by a modified-projector PAW technique that conceptually bases on Slater transition state theory [147].<sup>4</sup> We approach the difference in the core levels in a different way. The ground-state effects are mainly contained in the electrostatic potential at the position of the nucleus, whereas the dielectric response for the core hole is obtained from our dielectric model (cf. Fig. 5.17). Combining the ground-state electrostatic potential shifts with those arising from the core-hole induced polarisation, we are able to reproduce the experimental difference reasonably well (cf. Tab. 5.4). We find that mainly the dielectric response of the Mo(112) substrate is responsible for this difference, while the electrostatic contribution is opposite and small (0.2 eV). We take this as evidence that the dielectric response plays an important role for charged excited states in metal-supported oxide films.

We have also used a model self-energy similar to Eq. 5.4 to compute an image-potential corrected valence DOS for the silica film on Mo(112). Similar to the NaCl film on Ge(001), we find that the height and width of the peak is modified in the upper valence region, but the relative peak positions vary by only 0.1–0.2 eV, because the delocalised states undergo average self-energy shifts. This justifies the use of the uncorrected DOS for discussing the experimental peaks in Sec. 4.3.3. Since the precise peak positions of the model-corrected spectrum sensitively depend on the chosen position of the metal/film interface, we refrain from discussing it in further detail. Strong shifts are found for the localised O 2s states that were unfortunately not probed by the UPS experiments since the photon energy of the He<sup>I</sup> light (21.2 eV) used for these experiments is not sufficient to excite the O 2s electrons (binding energy  $\approx 22$  eV). The good agreement achievable for the O

<sup>4</sup>Slater transition state theory takes into account (static) polarisation effects by calculating the self-consistent potential in the presence of an extra charge.

1s core level shifts with our simple electrostatic + image potential approach demonstrates that the dielectric response effects are very important for a quantitative understanding of the photoemission spectra.

We note that image-potential effects for this film were discussed earlier in connection with Si 2p XP spectra. During the growth of the silica films a drift in the Si 2p binding energy was observed in XPS [104, 105]. The shifts were attributed to image effects and explained with a dielectric 3-layer model [105] assuming that the silica film grows layer-wise. However, this is incompatible with the atomic structure of the final film that was not known at that time. Since the final film is only 1 ML thick, an intermediate growth with several different layers below the silicon height in the crystalline film seems unlikely. We therefore believe that the drift in the Si 2p binding energy is not an image-potential effect.

The dielectric response is not only relevant for electron spectroscopy, but generally plays a role for charged states in a supported thin film, at its surface, or for adsorbates. This has recently been demonstrated theoretically for Au atoms on a MgO film on a Mo(100) substrate by Pacchioni *et al.* [121] using DFT. While dielectric response effects for charged excitations are not contained in the DFT band structure because the effective potential is that of the neutral ground-state, the total energy reflects them when the extra charge is explicitly considered or if the ground-state involves a charge transfer.<sup>5</sup> The ground-state surface properties of the MgO film do not differ strongly between a free-standing and a supported film and are converged at  $\sim 3$  ML in both cases. However, the adsorption of gold atoms differs drastically between free-standing and supported films. In this particular case, the presence of the metal substrate alters the charge state for adsorbed Au atoms, which become  $\text{Au}^-$ . Pacchioni *et al.* have found a significant thickness dependence of the adsorption energy which they attributed to the difference in the image potential [121].<sup>6</sup> While the metal-induced change of the adsorbate's charge is probably an extreme case for the influence of the metal substrate, the dielectric response of the substrate certainly affects the stability of charged adsorbates [121] or charged defects at the surface [91]. This has dramatic consequences for the use of supported oxide films as model supports for heterogeneous catalysts. The image potential induced

---

<sup>5</sup>This is also the reason why Slater transition state theory accounts well for the polarisation effects. However, the factor 1/2 arising from the adiabatic switching on of the induced potential is put in by hand.

<sup>6</sup>However, the magnitude of the effect in the calculation may have been enhanced by the use of a finite metal slab. Since the electron for  $\text{Au}^-$  is taken from the metal, the metal slab will acquire a finite excess charge density which adds an electrostatic contribution to the interaction energy.

by the substrate directly modifies the energetics of all charged states at the surface of the oxide film and of adsorbates on it. Moreover, this influence is invisible in the absence of charged states, i.e. neutral spectroscopies like optical absorption or EELS will not show this effect.

Likewise, charge carrier processes in oxide films used in electronic devices may be influenced by image potential effects. Since the films have two identical (or at least dielectrically similar) surfaces, the situation is comparable to free-standing films, but with opposite sign. In electronic device applications, the oxide films are embedded in a material with a higher dielectric constant. The image potential therefore reduces the band gap, thereby lowering the barriers for electron tunnelling. From the NaCl films, which have a similar electronic dielectric constant as the oxides, we estimate that such effects become relevant (i.e. exceed 0.1 eV) at thicknesses below  $\sim 10$  Å, depending on the dielectric constant of the oxide and of the surrounding semiconductor or metal. For a specific combination of dielectric constants, simple dielectric slab models can provide an upper estimate of the effect.

## 5.4 Summary

In this Chapter we have presented  $G_0W_0$  quasiparticle calculation for free-standing and supported thin insulator films, using NaCl as a prototypical example. We have shown that the surface polarisation in free-standing films and the dielectric response of a substrate introduce important changes beyond the shifts in the bulk materials. Thereby, free-standing and supported films behave quite unlike: In free-standing films the quasiparticle gap increases. When, on the other hand, the film is supported on a substrate with a higher dielectric constant the self-energy becomes position-dependent, leading to non-uniform  $G_0W_0$  corrections within the valence and conduction bands of the LDA. These effects can be qualitatively understood with simple dielectric models, but a full  $G_0W_0$  calculation for the supported system is necessary if quantitative results are required. We have then presented evidence that image effects are also visible in the metal-supported oxide films. Finally we have discussed other properties besides the quasiparticle spectrum that may be affected by the dielectric response effects.



# Chapter 6

## Conclusions

### 6.1 Summary

In Sec. 3 we have reviewed the repeated-slab approach for  $G_0W_0$  calculations, identified certain difficulties associated with it and presented appropriate solutions:

- For a numerically stable scheme, the anisotropy in the dielectric screening must be taken into account for the treatment of the  $1/k^2$  singularity of the screened interaction (Sec. 3.2.2). A robust scheme for the  $GW$  space-time method has been developed.
- An efficient image-charge method has been developed for the calculation of the screened interaction in two-dimensional model systems (homogeneous in two spatial directions with an arbitrary dielectric profile in the third, cf. Sec. B.1). These models proved to be instrumental for the understanding of the screening effects in two-dimensional systems.
- The convergence behaviour of the quasiparticle energies with respect to the  $\mathbf{k}$ -point sampling parallel to the surface is slow (Sec. 3.3.2). This can be attributed to the transition between the local and the average dielectric screening in non-homogeneous systems. For the repeated-slab approach, small vacuum/slab ratios are numerically advantageous. A functional form has been derived from a dielectric model that allows to extrapolate the convergence curves to the limit of infinite  $\mathbf{k}$ -point sampling, thereby reducing the computational cost for accurate  $G_0W_0$  calculations considerably.
- The polarisation of the neighbouring slabs in the repeated-slab approach introduces a slow convergence with respect to the slab separation (vacuum thickness). Dielectric models quantitatively describe this

effect and are used to correct for the finite vacuum thickness *a posteriori* (Sec. 3.3.3). Further modifications to the  $G_0W_0$  scheme are then not necessary.

In Sec. 4, we have analysed ground-state effects for thin oxide films in the context of heterogeneous model catalysts.

- The thin silica film on Mo(112) exhibits the same bonding connectivity as the reconstructed  $\alpha$ -quartz surface and other silica thin films or surfaces (Sec. 4.3.1). It can thus be seen as a surface form of silica.
- For oxide films of one or two monolayers thickness, the substrate exerts a significant influence on the atomic structure. The resulting structures can be understood in terms of the same principles that govern the ionic bonding in bulk materials. For this, substrate constraints such as the substrate-defined lattice constants for commensurate films as well as the linkage to the substrate surface must be taken into account.
- The electronic structure of supported thin films is modified by the substrate. This is mainly due to the substrate-induced structural variations, whereas the direct influence e.g. via hybridisation is small (Sec. 4.3.2).
- The thickness of the oxide films significantly influences the atomic and electronic structure of ultrathin films (Sec. 4.4). Once bulk-like layers are present in the centre of the film, the properties quickly converge with increasing thickness.
- Band gaps and band widths have been found to be insensitive to the substrate or the thickness for the investigated cases. For the valence bands this is attributed to the localised nature of these states, whereas the conduction band minimum is pinned by the hybridisation with a surface resonance.

Returning to the fundamental hypothesis of the surface science approach “do ultrathin oxide films resemble oxide surfaces” we can now say: Monolayer films can be regarded as special forms of the oxide materials which in general are not representative for the bulk materials and their surfaces.

Self-energy effects for ultrathin NaCl films have been analysed in Sec. 5 in the context of photoelectron spectroscopy.

- The dominant changes beyond the scissor-like shifts observed in the bulk material can be attributed to dielectric response effects. However, simple models overestimate the magnitude of the effects compared to the full  $G_0W_0$  calculations.

- In free-standing films, the surface polarisation leads to an increase of the quasiparticle gap. This introduces a thickness-dependence of the quasiparticle gap not observed at the level of DFT-LDA (Sec. 5.1).
- In supported films, the image potential is asymmetric (Sec. 5.2.3). The image potentials induced by the change in the dielectric constant at the surface and the interface oppose each other. This weakens the thickness-dependence of the peak positions compared to the free-standing films.
- For delocalised states such as the Cl  $3p$  states, the varying self-energy corrections lead to a broadening of the peaks compared to the DFT-LDA DOS. Localised states undergo position-dependent shifts that reflect the local image potential.
- The  $G_0W_0$  calculations agree with available experimental data for the position of the NaCl bands with respect to the Ge valence band maximum.
- As a first prototypical molecule adsorbed on Ge-supported NaCl films, carbon monoxide was studied. The CO levels undergo shifts that depend on the NaCl film thickness. This prediction agrees with the trend observed in STS experiments for pentacene adsorbates for NaCl films on copper.
- The dielectric model has been used to estimate the magnitude of image-charge effects for metal-supported oxide films. We find that they are significant and can shift localised states by more than 1 eV. For silica/Mo(112), the differences in the O  $1s$  binding energies observed in XPS are dominated by these final-state effects. Likewise, we expect the energetics of defects and charged adsorbates to be strongly modified by the metal substrate.

Moreover, an atomistic model of the NaCl/Ge(001) interface structure was obtained from DFT-LDA calculations (Sec. 5.2.2).

- The adhesion of NaCl films is effectuated by the electrostatic attraction of the ions in the film and the charge modulation in the buckled dimers of the Ge(001) surface. The adhesion mechanism shows a surprising similarity to that found for NaCl/Cu(311).
- The NaCl film induces a strong preference for a  $p(2\times 1)$  ordering of the Ge dimers. Other patterns that are observed for the clean Ge surface such as  $c(4\times 2)$  are energetically unfavourable.

- Flat NaCl films on Ge(001) are only metastable; thermodynamically, three-dimensional islands are favoured.
- In STM experiments with atomic resolution, the Cl atoms at the surface appear as bright spots. The theoretical simulated apparent layer heights agree reasonably with experiment.

The  $G_0W_0$  calculations for repeated-slab systems required several technical improvements of the `gwst` code used (see App. E):

- A new, highly efficient algorithm for computing the contribution of the non-local pseudopotential to the matrix elements of the position operator has been developed.
- A more efficient implementation for the computation of the Green's function has been designed and implemented, which is about six times faster than the original one.
- A block algorithm for the inversion of Hermitean matrices in packed storage form has been devised and implemented. The resulting algorithm is about two times faster than the standard iterative LAPACK algorithm.
- A new storage scheme for the two-point functions reduces the disk space requirements by 30–40%.

## 6.2 Outlook

The  $G_0W_0$  scheme for slab systems presented in this thesis, combining accurate numerical calculations with extrapolation and corrections procedures, provides an efficient approach for computing the quasiparticle band structure of slab systems. Using the current implementation of the  $GW$  space-time method, slab calculations can now be performed routinely. The current developments for metallic systems [32] are conceptually independent of our improvements and will soon extend the applicability also to the metal-supported insulator films. This opens new possibilities for analysing the photoemission spectra of these experimentally well-characterised systems and may help to clarify the role of dielectric response effects for delocalised states.

A further perspective for improving the efficiency of the  $G_0W_0$  calculations is to incorporate the dielectric slab models directly into the treatment of the screened interaction. As stated in Sec. 3.2.2, any model function that

is known in real and reciprocal space may be used for treating the long-range part. Appropriate dielectric profiles could be derived consistently from the inverse dielectric matrix employing e.g. Rohlfing's averaging procedure [61]. The reciprocal-space representation can be obtained from the two-dimensional Ewald summation and a numerical FFT for the  $z$ -directions.

Our substrate-influence and thickness-dependence study indicates that monolayer films are special forms of the oxide materials. In order to improve the comparability to the bulk materials, the thickness of the films must be increased. However, this may not be sufficient alone. A second key issue is to produce interfaces compatible with the bulk structures. We emphasise that the growth of flat crystalline films does in general not indicate a good match between the substrate and the bulk oxide; on the contrary a preference for a specific thickness points towards unique structures differing from the bulk. We therefore suggest that the currently employed substrates are not the ideal starting point in the quest for thicker well-defined films. Repp *et al.* suggested that the electrostatic adhesion mechanism for insulator films on corrugated vicinal metal surfaces observed for the NaCl film on Cu(311) may offer a general strategy for the growth of flat crystalline films [6]. This may provide an alternative to the currently prevailing chemical-bond interfaces for silica and alumina.

Before we finish we will comment on the perspectives of the NaCl film on Ge(001). Since this is a computationally feasible system for theoretical investigations, an experimental verification of the predictions in Sec. 5.2.3 would be highly welcome. According to our calculations, the dielectric response effects should introduce noticeable changes in the band structure for the delocalised thin film states. In particular, angle-resolved UPS measurements for films of a few monolayer thickness could provide the data for a direct comparison between experiment and our  $G_0W_0$  results to establish how accurately dielectric response effects are described in the theory. Likewise, new STM experiments could help to clarify the current discrepancy between theory ( $2 \times 1$ , as in LEED) and experiment ( $1 \times 1$ ). Furthermore, the planned extension of the  $GW$  space-time method to metallic system will soon allow to study also the copper-supported NaCl films for which a wealth of data is already available.

The study of molecules on surfaces, ranging from biomolecules over molecular switches to small metal clusters, is an active area of research. In particular the use of ultrathin insulating spacer layers allows to reduce the direct coupling of the molecular levels to the metal states of the substrate. The coupling to the insulating spacer layer is much weaker, in particular for molecular levels in the band gap of the insulator film. We expect STS or PES to provide new insights into the electronic structure of the molecules

---

and its coupling to the substrate in the future. While standard DFT describes the ground-state properties such as the adsorption geometries and energies reasonably well, it fails for the excited states. The  $G_0W_0$  calculations offer an accurate and appropriate framework to describe these new experiments theoretically and to analyse the decisive physical effects in detail. While simple image-charge models are sufficient to qualitatively explain the relevant effects for these systems, the  $G_0W_0$  are required for a quantitative comparison. We have demonstrated that such calculations for composite molecule/insulator/semiconductor systems are feasible with present-day computer codes and resources. This opens the way for a fruitful collaboration of experiment and theory in the future that has been so successful in other areas in the past.

# Appendix A

## Abbreviations

AO	atomic orbital
AES	Auger emission spectroscopy
AFM	atomic force microscopy
BIS	bremstrahlung isochromat spectroscopy
BLAS	basic linear algebra subroutines
COHSEX	Coulomb-hole / screened exchange
CPU	central processing unit
DFT	density functional theory
DOS	density of states
EXX	exact exchange
FET	field effect transistor
FFT	fast Fourier transform
GGA	generalised gradient approximation
GW	is <i>no</i> abbreviation, but stands for the product of Green's function $G$ and screened interaction $W$
GWA	GW approximation
ELS-LEED	energy-loss spectroscopy of low-energy electron diffraction
HEG	homogeneous electron gas
HREELS	high resolution electron energy loss spectroscopy
IPES	inverse photoemission spectroscopy
IR	infrared
IRAS	infrared reflection absorption spectroscopy
KS	Kohn-Sham

---

LAPW	linear augmented plane waves
LCAO	linear combination of atomic orbitals
LDOS	local density of states
LDA	local-density approximation
LEED	low-energy electron diffraction
LEIS	low-energy ion scattering
MBPT	many-body perturbation theory
MIES	metastable impact electron spectroscopy
ML	monolayer
OEP	optimised effective potential
PES	photoemission spectroscopy
PRE	parallel repetition error
RPA	random phase approximation
SCF	self-consistent field
SXRD	surface X-ray diffraction
STM	scanning tunnelling microscopy
UPS	ultraviolet photoelectron spectroscopy
UV-VIS	ultraviolet & visible [light]
XPS	X-ray photoelectron spectroscopy
XRD	X-ray diffraction



# Appendix B

## Dielectric models

### B.1 Image-charge method for dielectric layer models

In this section a simple scheme is presented to compute the screened interaction in laterally homogeneous model systems with a layered structure (cf. Fig. B.1a). Each layer  $z$  has the same thickness  $L$  and a layer-specific dielectric constant  $\varepsilon_z$ . To simplify the notation, we work with reduced units in the following, i.e. lengths are measured in units of  $L$ , charges in units of the unit charge  $Q$ , and potentials in units of  $Q/L$ . Our coordinate system is chosen such that the layers are centred around integer  $L$ , and the interfaces are at half integer  $L$ . We note that any given dielectric profile  $\varepsilon(z)$  can be approximated by such a layer model when the profile is discretised into individual layers of a sufficiently small thickness. For the model calculations in this work, we usually use  $L=1$  bohr. At the boundary between two layers we assume sharp interfaces so that the dielectric constant jumps from  $\varepsilon_z$  to  $\varepsilon_{z+1}$ .

The screened interaction  $W(\mathbf{r}, \mathbf{r}')$  is obtained as the potential  $V(\mathbf{r})$  when a unit charge is placed at  $\mathbf{r}'$ . We compute  $V(\mathbf{r})$  by the method of image charges [90]. As an introductory remark, let us first consider the textbook situation of two semi-infinite dielectric media  $\Omega_A$  and  $\Omega_B$  with dielectric constants  $\varepsilon_A$  and  $\varepsilon_B$ , sketched in Fig. B.1b. For a charge  $q$  at  $\mathbf{r}'$  in  $\Omega_A$ , the potential  $V(\mathbf{r})$  is given by

$$\mathbf{r} \in \Omega_A: \quad V(\mathbf{r}) = \frac{1}{\varepsilon_A} \left( \frac{q}{|\mathbf{r} - \mathbf{r}'|} + \frac{q''}{|\mathbf{r} - \mathbf{r}''|} \right), \quad (\text{B.1})$$

$$\mathbf{r} \in \Omega_B: \quad V(\mathbf{r}) = \frac{1}{\varepsilon_B} \frac{q'}{|\mathbf{r} - \mathbf{r}'|}, \quad (\text{B.2})$$

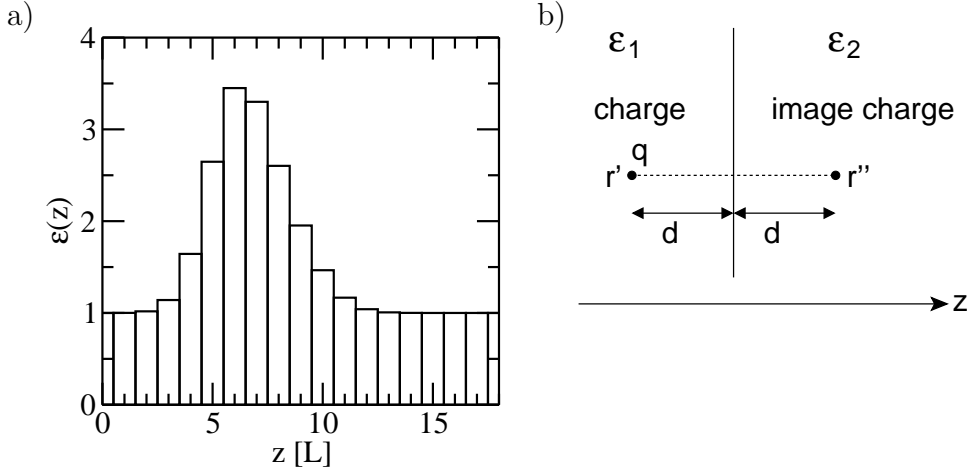


Figure B.1: a) Dielectric layer model. b) Charge and image charge at a dielectric interface.

where  $q'$  and  $q''$  are image charges. They are determined from the continuity equations of the electric field and the electric displacement at the interface, yielding

$$q' = \frac{2\varepsilon_B}{\varepsilon_A + \varepsilon_B} q \quad (\text{B.3})$$

$$q'' = \frac{\varepsilon_A - \varepsilon_B}{\varepsilon_A + \varepsilon_B} q. \quad (\text{B.4})$$

$r''$  is obtained by reflecting  $r'$  at the interface, and we will therefore denote  $q''$  as “reflected charge”, whereas the effect of  $q$  is propagated into  $\Omega_B$  by the “propagated charge”  $q'$ . The image charges  $q'$  and  $q''$  are no physically observable charges, but only mathematical constructs to simplify the computation of the potential. Most importantly, the image charges for the potential on one side of the interface are always located on the other side of the interface, whereas the original charge is on the same side. This criterion is useful in multi-layer systems to identify the interfaces for higher-order image charges.

In order to develop a computational scheme for a multi-layer system, a proper book-keeping is crucial to keep track of the various image charges. For reasons that will become clear below, we denote an image charge that contributes to the potential in layer  $z$  and is located at  $z + \sigma d$  by  $q(z, d, \sigma)$ , where  $d \geq 0$  is the distance from the layer and  $\sigma = \pm 1$ . To show that the image charges can be determined iteratively we will now derive the iteration for  $d \rightarrow d + 1$ . Consider a charge  $q(z, d, \sigma)$  relevant for the potential in layer  $z$ . Due to the interface at  $z - \frac{1}{2}\sigma$  two additional image charges appear. The

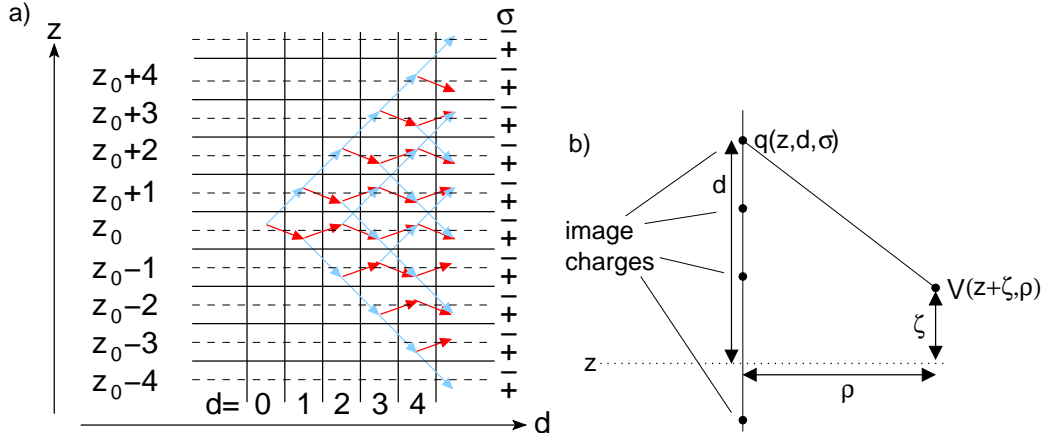


Figure B.2: a) Iterative determination of image charges (see text). b) Computation of the potential from the image charges.

reflected image charge, located at

$$\underbrace{z - \frac{\sigma}{2}}_{\text{interface}} - \underbrace{\left( (z + \sigma d) - \left( z - \frac{\sigma}{2} \right) \right)}_{\text{distance from interface}} = z - \sigma(d + 1) \quad (\text{B.5})$$

describes the potential in layer  $z$  and is given by (cf. Eq. B.4)

$$q^{\text{rf}}(z, d + 1, -\sigma) = \frac{\varepsilon_z - \varepsilon_{z-\sigma}}{\varepsilon_z + \varepsilon_{z-\sigma}} q(z, d, \sigma). \quad (\text{B.6})$$

The propagated image charge remains at the position  $z + \sigma d$  and describes the potential in layer  $z - \sigma$ . Using our book-keeping notation and Eq. B.3 it can be written as

$$q^{\text{pr}}(z - \sigma, d + 1, \sigma) = \frac{2\varepsilon_{z-\sigma}}{\varepsilon_z + \varepsilon_{z-\sigma}} q(z, d, \sigma). \quad (\text{B.7})$$

Obviously, the distance parameter is increased by 1 for each interface taken into account. The image charges for the distance  $d + 1$  can thus be computed iteratively from those at distance  $d$  and will in general combine a reflected and a propagated contribution  $q = q^{\text{rf}} + q^{\text{pr}}$ .

The dielectric discontinuity at the interfaces between two layers introduces a divergence of the potential close to the interface. However, this is an artefact of assuming homogeneous layers with sharp interfaces. To avoid it, we restrict the position of the original charge to the center of each slab; the image charges are then located at the center of a layer, too. The iterations are then started by setting

$$q(z_0, 0, \pm 1) = 1. \quad (\text{B.8})$$

The flow of the iterations is schematically depicted in Fig. B.2. Each rectangular box in the scheme corresponds to one image charge  $q(z, d, \sigma)$  and the red (light blue) arrows indicate the reflected (propagated) contributions to the image charges of the next generation. For clarity, only the flow for  $q(z_0, 0, -1)$  is shown. In practice, the iterations are stopped at some  $d_{\max}$  which thereby becomes a convergence parameter. In the iteration scheme in Fig. B.2a this corresponds to stop going to the right. In addition, we truncate the system and neglect image charges that fall outside, which corresponds to ignoring charges at the bottom or the top in Fig. B.2a. This truncation becomes a second convergence parameter. The convergence for both parameters was tested by doubling the parameter until the changes became negligible.

The screened potential depends on the layer  $z$  and the lateral distance  $\rho$  from the vertically aligned image charges. By summing the Coulomb potential of all the image charges relevant for this layer (cf. Fig. B.2b), we obtain

$$V(z + \zeta, \rho) = \frac{1}{\varepsilon_z} \sum_{d=0}^{d_{\max}} \sum_{\sigma} \frac{q(z, d, \sigma)}{\sqrt{(\zeta - \sigma d)^2 + \rho^2}}, \quad (\text{B.9})$$

where  $\zeta$  denotes the vertical position within the layer ( $|\zeta| < \frac{1}{2}$ ). We usually restrict the calculation to  $\zeta = 0$ . When Eq. B.9 is evaluated successively during the iterations, there is no need to store the image charges for all  $d$ , which makes the implementation very memory-efficient.

## B.2 Connection to $G_0W_0$ : the static COHSEX approximation

In this section, we will show how the screened interaction calculated from dielectric models can be used to estimate changes in the quasiparticle energies. In the end, we will arrive at a very simple scissors operator. We will then show that the same result can be obtained by considering the image-potential energy of the charged  $N \pm 1$ -electron systems that result from an electronic excitation.

The  $G_0W_0$  self-energy can be decomposed into physically meaningful entities. One such possibility is to decompose it into a screened exchange and a Coulomb-hole part. To this end, we start from the expression for the self-energy in the frequency domain

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' G_0(\mathbf{r}, \mathbf{r}', \omega + \omega') W_0(\mathbf{r}, \mathbf{r}', \omega') e^{i0^+\omega'}. \quad (\text{B.10})$$

The residual theorem states that only the poles of  $G_0$  and  $W_0$  contribute to the integral. Closing the integration contour above the real axis includes only the poles of the Green's function for the occupied states [23], which leads to the screened exchange self energy

$$\Sigma_{\text{sx}}(\mathbf{r}, \mathbf{r}', \omega) = - \sum_n^{\text{occ}} \phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}') W(\mathbf{r}, \mathbf{r}', \omega - \epsilon_n). \quad (\text{B.11})$$

The poles of the screened interaction are given by the plasmon energies  $\pm(\omega_p - i0^+)$ , i.e.

$$W(\mathbf{r}, \mathbf{r}', \omega_p) = v(\mathbf{r}, \mathbf{r}') + \sum_p \frac{2\omega_p \chi_p(\mathbf{r}) \chi_p^*(\mathbf{r}')}{\omega^2 - (\omega_p - i0^+)^2}, \quad (\text{B.12})$$

where  $\chi_p$  denotes the corresponding plasmon functions. When we expand also  $G_0$  in its spectral representation, we arrive at the Coulomb-hole self-energy

$$\Sigma_{\text{coh}}(\mathbf{r}, \mathbf{r}', \omega) = \sum_p \sum_n \frac{\phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}')}{\omega - \omega_p - \epsilon_n} \chi_p(\mathbf{r}) \chi_p^*(\mathbf{r}'). \quad (\text{B.13})$$

Assuming a static interaction (or more precisely,  $\omega - \epsilon_n \ll \omega_p$  for the frequency range of interest) leads to the static COHSEX approximation. Comparison of the equations for  $\Sigma_{\text{coh}}$  and  $\Sigma_{\text{sx}}$  to Eq. B.12 for  $\omega = 0$  and exploiting the identity

$$\sum_n \phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \quad (\text{B.14})$$

we obtain

$$\Sigma = \Sigma_{\text{coh}} + \Sigma_{\text{sx}}, \quad (\text{B.15})$$

$$\Sigma_{\text{coh}}(\mathbf{r}, \mathbf{r}') = \frac{1}{2} (W(\mathbf{r}, \mathbf{r}') - v(\mathbf{r} - \mathbf{r}')) \delta(\mathbf{r} - \mathbf{r}'), \quad (\text{B.16})$$

$$\Sigma_{\text{sx}}(\mathbf{r}, \mathbf{r}') = - \sum_n^{\text{occ}} \phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}') W(\mathbf{r}, \mathbf{r}'). \quad (\text{B.17})$$

A particularly useful expression can be derived from the static COHSEX approach when it is applied to long-range screening effects in separated subsystems. For each subsystem, the presence of the other subsystems does not alter the Green's function. However, the polarisation of the other subsystems adds a contribution  $\Delta W^{\text{p}}$  to the screened interaction which varies only smoothly, i.e., is essentially constant over the subsystem as shown in Section 3.3.3. The additional self-energy then becomes

$$\Delta \Sigma^{\text{p}}(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \sum_n^{\text{unocc}} \phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}') \Delta W^{\text{p}} - \frac{1}{2} \sum_n^{\text{occ}} \phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}') \Delta W^{\text{p}}. \quad (\text{B.18})$$

This is a symmetric scissors operator that opens the quasiparticle gap by  $\Delta W^p$ .

We arrive at the same result by considering the change in the total energy  $E_{N\pm 1,s}$  after the electronic excitation. The induced image-potential in the charged final state shifts the energy by  $\frac{1}{2}\Delta W$  for both positively charged (hole) states or negatively charged (electron) states. The hole energy is (cf. 2.34)

$$\epsilon_s = E_{N,0} - (E_{N-1,s} + \frac{1}{2}\Delta W) < E^{\text{Fermi}} \quad (\text{B.19})$$

and corresponds to occupied states in the initial system, which are thus shifted by  $-\frac{1}{2}\Delta W$ . For the electron energy (cf. 2.35)

$$\epsilon_s = (E_{N+1,s} + \frac{1}{2}\Delta W) - E_{N,0} > E^{\text{Fermi}} \quad (\text{B.20})$$

the opposite is true: the final state effect shifts the unoccupied band energy by  $+\frac{1}{2}\Delta W$ .

# Appendix C

## Ultrathin oxide films

### C.1 Alternative terminations for the $\alpha$ -quartz (0001)

In order to see if the siloxane surface is the most stable surface termination for silica slabs at all thicknesses, we have also investigated unreconstructed quartz slabs with three possible terminations: a Si-terminated surface containing two-fold coordinated silicon atoms at the surface, a  $O_1$ -terminated where a single oxygen atom is added to the Si-termination, and an  $O_2$ -termination with two oxygen atoms per Si surface atom. Only the  $O_1$ -terminated surface yields stoichiometric slabs, whereas the Si-termination ( $O_2$ -termination) has a silicon (oxygen) excess. The  $O_1$ -terminated surface exhibits an almost planar configuration at the surface silicon atom with a relatively short Si-O bond (1.50 Å). An analysis of the electronic wavefunctions reveals that no strong double-bond character can be detected and that it is better described as a  $Si^+-O^-$  entity where the short bond results from a strong Coulomb attraction. We note here that the the  $Si_4O_8$  film with an  $O_1$ -termination spontaneously decomposed into two reconstructed  $Si_2O_4$  layers during a standard relaxation. Only when some of the central atoms were kept frozen, the structure analogous to the  $Si_3O_6$  and  $Si_5O_{10}$  films could be obtained. This indicates that such under-coordinated Si-O species are highly reactive and may play an important role in the restructuring processes in real materials. Also at the  $O_2$ -terminated surface, we observe a reaction: the oxygen atoms dimerise. The resulting structure can be understood as a peroxide ion coordinated side-on to the silicon centre (cf. Fig. C.1). This assignment is supported by the O-O bond length of 1.60 Å characteristic for oxygen single bonds, and by the electronic structure. Similar dimerisations are observed when oxygen atoms are adsorbed on oxide surfaces [148].

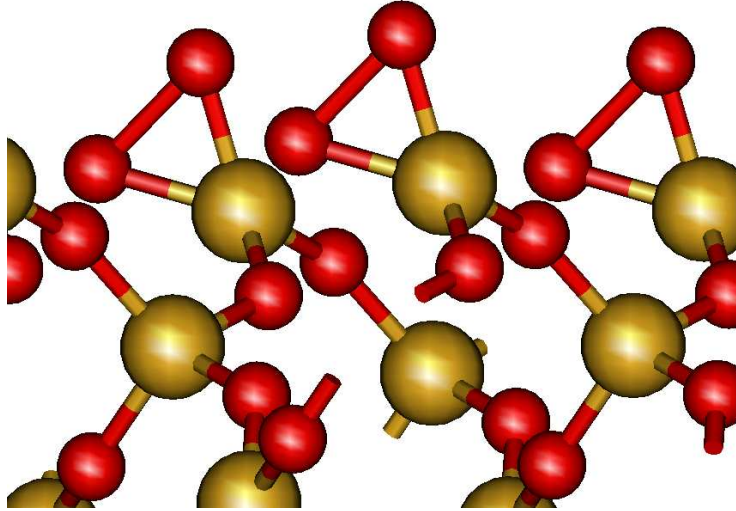


Figure C.1: The relaxed  $O_2$ -terminated surface structure with a direct O-O peroxide bond.

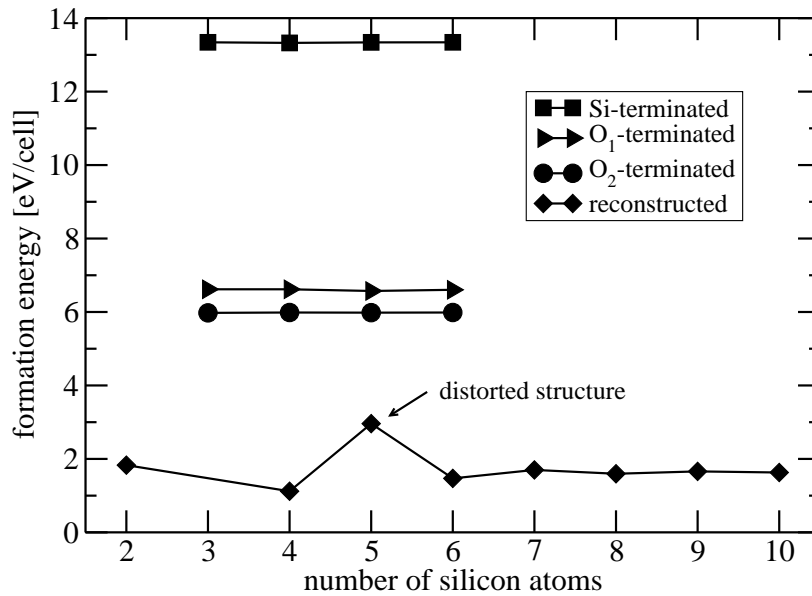


Figure C.2: Formation energies (DFT-LDA) of thin quartz-like slabs with various terminations as well as the reconstructed silica slabs as a function of thickness.



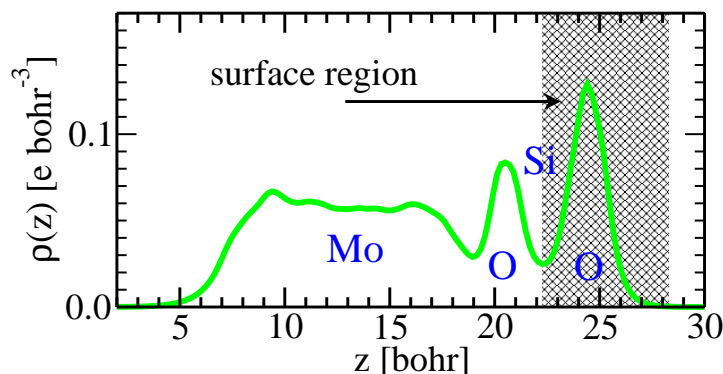
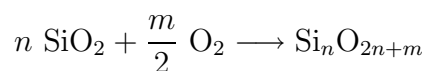


Figure C.3: Definition of the surface region for the  $\text{Mo}_{10}\text{Si}_2\text{O}_5$  system.

In Fig. C.2 the formation energies of the slabs from bulk  $\alpha$ -quartz and molecular oxygen<sup>1</sup> according to



are plotted as a function of  $n$ .  $m$  is the surface oxygen excess, ranging from -2 for the Si-terminated slab to +2 for the  $\text{O}_2$ -terminated one. The non-reconstructed structures show no visible dependence of the formation energy on the slab thickness. In other words the formation energy is determined by the surface only and reaches a constant value already for very small slabs. The formation energy of the reconstructed slabs on the other hand shows more oscillations. For slabs up to  $n = 6$ , this reflects the variations in the slab structure since a quartz-like central part is present only from  $n = 6$  on. However, the oscillations continue even for thicker slabs. This can be traced back to the misfit of the siloxane surface structure to the quartz substrate. The resulting strain is accommodated in the substrate and decays only slowly with increasing depth. This can also be monitored by the Si-O-Si angle – the most sensitive parameter to structural deformations – which approaches its bulk value of  $140^\circ$  only slowly with increasing depth. The reconstructed films are more stable than the other terminations, which is not surprising since all dangling bonds are saturated after the reconstruction.

## C.2 Surface-projected density of states

For a meaningful comparison between different substrates, the DOS must be projected onto the surface region. Here, we will describe how we define the

<sup>1</sup>Spin-polarized  $\text{O}_2$  with the theoretical bond of  $1.209 \text{ \AA}$ . To employ the plane-wave code, the molecule is placed in a large simulation box.

surface region and perform the spatial projection. The z-resolved electron density for the (pseudo)valence states

$$\rho(z) = \frac{1}{A} \int dx dy \rho(x, y, z) , \quad (\text{C.1})$$

where the integral is taken over the surface unit cell (with area  $A$ ), exhibits minima and maxima corresponding to the ionic layers (cf. Fig. C.3). Since the valence electronic structure of the oxides is dominated by oxygen-derived states, we can use the minima to divide the system into individual oxygen layers. The last oxygen layer then defines the surface region  $\Omega_{\text{srf}}$ . The surface local density of states (LDOS) is obtained as

$$LDOS(E) = \sum_n \delta(E - \epsilon_n) \int_{\Omega_{\text{srf}}} d^3\mathbf{r} |\psi_n(\mathbf{r})|^2 , \quad (\text{C.2})$$

i.e. the partial density integrated over  $\Omega_{\text{srf}}$  is used to weight the peaks in the DOS.

# Appendix D

## NaCl films

### D.1 Atomic orbital projections

In order to assess the character of a wave function  $\psi_{n\mathbf{k}}$  in a crystal, it is often instructive to decompose it into atomic contributions. While this is very natural when atom-centred local orbitals are used as basis set to expand the wavefunction, plane waves are not associated with particular atoms. It is therefore necessary to project the wavefunctions onto an atomic basis set [149]. Here, we employ the atomic pseudo-wavefunctions that are used to define the pseudopotential projectors for this purpose. We denote the atomic orbitals by  $\chi$  and employ Greek letters  $\mu, \nu, \rho$  for the indices. From these orbitals, Bloch states  $\chi_{\mu\mathbf{k}}$  are formed.

However, atomic orbital basis sets are in general non-orthogonal, i.e. the overlap matrix

$$S_{\mu\nu}(\mathbf{k}) = \langle \chi_{\mu\mathbf{k}} | \chi_{\nu\mathbf{k}} \rangle \quad (\text{D.1})$$

is not diagonal. In the following we will employ a notation in analogy to the covariant and contravariant coordinates in non-orthogonal coordinate systems. The inverse overlap matrix is then written  $S^{\mu\nu}(\mathbf{k})$ , and is defined by

$$\sum_{\rho} S_{\mu\rho}(\mathbf{k}) S^{\rho\nu}(\mathbf{k}) = \delta_{\mu\nu} . \quad (\text{D.2})$$

The projections

$$c_{n\mu}(\mathbf{k}) = \langle \chi_{\mu\mathbf{k}} | \psi_{n\mathbf{k}} \rangle \quad (\text{D.3})$$

then differ from the expansion coefficients  $c_n^{\mu}(\mathbf{k})$  that generate the crystal wavefunctions via

$$|\psi_{n\mathbf{k}}\rangle = \sum_{\mu} c_n^{\mu}(\mathbf{k}) |\chi_{\mu\mathbf{k}}\rangle . \quad (\text{D.4})$$

The expansion coefficients are connected to the projections by the inverse overlap matrix  $S^{\mu\nu}$

$$c_n^\mu(\mathbf{k}) = \sum_\nu S^{\mu\nu}(\mathbf{k}) c_{\nu i}(\mathbf{k}) . \quad (\text{D.5})$$

We note that the norm of the wave functions is given in the atomic orbital basis by

$$\langle \psi_{n\mathbf{k}} | \psi_{n\mathbf{k}} \rangle = \sum_\mu c_i^\mu(\mathbf{k}) c_{\mu i}(\mathbf{k}) . \quad (\text{D.6})$$

We then define the projection onto a subset of the atomic orbital basis (in general orbitals associated with one specific atom or one class of atoms) by restricting the sum in Equation D.6 to this subset.

The advantage of this technique lies in the fact that it is closely related to the Mulliken population analysis for atomic orbital basis sets. For example, summing the projections over occupied states

$$q_\mu = \int d^3\mathbf{k} \sum_i^{\text{occ}} c_i^\mu(\mathbf{k}) c_{\mu i}(\mathbf{k}) \quad (\text{D.7})$$

recovers the Mulliken gross populations  $q_\mu$ , which may give a qualitative picture of the charge redistribution in the solid.

In general, the atomic orbital basis set is not complete. Even including all atomic orbitals in the sum does not recover the full (plane-wave) norm. This spill-over amounts to typically 0.5–2% for occupied states but may become larger for unoccupied states. Since we use the projection onto atomic orbitals to assess the character of single bands, we do not reorthonormalise the projected bands as suggested in [149] for the population analysis because this would mix different bands. The spill-over is largely due to the minimal, non-adjusted basis set that results from the atomic pseudo-wavefunctions. We refrain from adapting the atomic orbitals to minimise the spill-over because we are interested in the qualitative picture and not in numbers.

## D.2 NaCl: the character of the bulk conduction band

While the valence bands of bulk NaCl are easily understood as being derived from the chlorine  $3p$  orbitals, the character of the lowest conduction band has proved to be more difficult to assess and has been a matter of discussion since the earliest investigations by Slater and Shockley [150]. Assigning an atomic character to a delocalised wavefunction is not unique, and different methods yield different results for the conduction band in NaCl. A simple

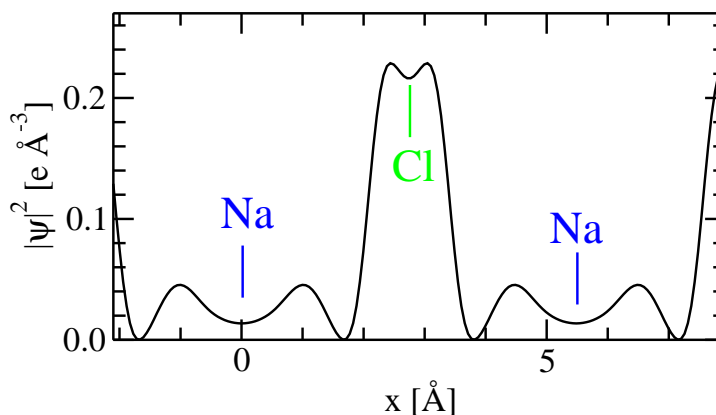


Figure D.1: Partial density profile of the conduction band minimum along the (001) direction, suggesting a dominant Cl character. The position of the Na and Cl atoms is indicated.

linear combination of atomic orbitals (LCAO) picture suggests that the conduction bands should be mainly composed of the sodium states. Indeed, the AO projection technique with the (minimal) atomic pseudo-orbital basis yields 70–93% Na character for the lowest conduction band. The maximum Na character is obtained for the conduction band minimum. However, the partial density computed from this state is clearly centred on the chlorine atoms. In Fig. D.1 we show the partial density  $|\psi|^2$  of the conduction band minimum along the [001] direction through the Na and Cl atoms. The high density close to the Cl nucleus reveals a dominant Cl character for this state. Further evidence for a considerable Cl character has been provided by de-Boer and de Groot [151, 152]: they considered a hypothetical  $\text{Cl}^-$  fcc lattice with a homogeneous background charge and found that it reproduces the band structure of NaCl very well. Moreover, they showed that the lowest conduction state disappears in a muffin-tin sphere calculation when the Cl 4s orbital is removed from the basis set. They concluded from this evidence that the lowest conduction state has mainly Cl 4s character.

Another way to approach this problem is to monitor how the bands develop when the ions are brought together from infinite separation, i.e. by changing the lattice constant from very large values to its equilibrium value [150]. At large separations, the ions do not interact and the bands reflect the ionic levels.<sup>1</sup> When the ions approach each other, the Madelung potential will lower the anion states (which are surrounded by the positively charged

<sup>1</sup>We note that in this case, the charge transfer from Na to Cl is even endothermic, i.e. the ground state are Na and Cl atoms rather than  $\text{Na}^+$  and  $\text{Cl}^-$ . We will however assume a charge-transfer state at all separations.

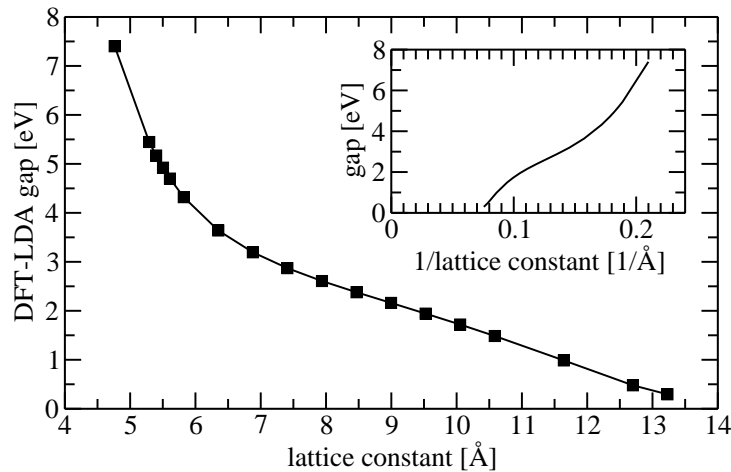


Figure D.2: Dependence of the DFT-LDA gap on the lattice constant for bulk NaCl. The inset shows the same data as function of the inverse lattice constant.

cations) and destabilise the cation states. A second effect is the broadening of the ionic levels into bands when the wavefunctions of the ions overlap. It is then possible that a broadened Cl 4s band drops below the corresponding Na 3s band at sufficiently small separations, but is this really the case for NaCl?

The change in the band gap should be a sensitive test for this question. With decreasing lattice constant, the sodium states will be shifted upward with respect to the Cl 3p states due to the increase in the Madelung potential. A gap for a sodium-derived conduction band should thus increase. On the other hand, the Cl 4s band broadening would increase and lead to a reduction in the band gap if Cl 4s states were dominant in the conduction band. We have therefore computed the band gap over a large range of lattice constants, cf. Fig. D.2. We find that the band gap increases with decreasing ion separation.<sup>2</sup> This indicates an important role of the Na orbitals in determining the band gap.

However, the influence of the Cl ions for the electronic band structure at the equilibrium lattice constant cannot be neglected. In agreement with de-Boer *et al.*, we find that a Cl<sup>-</sup> fcc lattice with a neutralising background alone reproduces the conduction band and even the band gap of bulk NaCl, while a corresponding Na<sup>+</sup> lattice shows a free-electron like band structure. Even when the Cl<sup>-</sup> ions are modelled by negative point charges for the Na<sup>+</sup>-only

<sup>2</sup>At even smaller lattice constants below 4.2 Å, the band gap reduces again since a further band drops below the dispersive band in question.

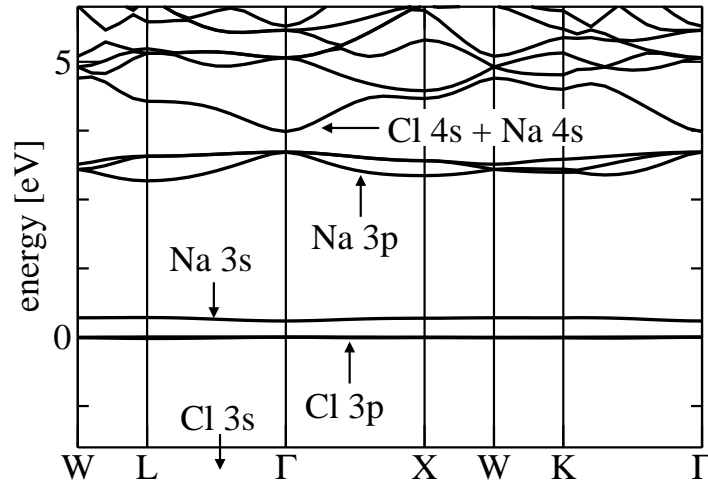


Figure D.3: DFT-LDA band structure of bulk NaCl at a lattice constant of  $13.2 \text{ \AA}$  (25 bohr).

calculation, strong deviations from the NaCl band structure are observed. The Cl ions are absolutely crucial for the observed dispersion of the conduction band. At large lattice constants, however, the Na-derived bands separate from the dispersive band continuum (cf. Fig. D.3) and cross the Cl  $3p$  level at  $\sim 14 \text{ \AA}$ . Surprisingly, we find at a lattice constant of  $13.2 \text{ \AA}$  (25 bohr) that the unoccupied Na  $4s$  states hybridise with the chlorine states and produce a dispersion reminiscent of bulk NaCl at the equilibrium lattice constant. None of these aspects can be reproduced with a  $\text{Cl}^-$  lattice alone. When the lattice constant is varied between the theoretical equilibrium ( $5.49 \text{ \AA}$ ) and  $13.2 \text{ \AA}$ , the lowest conduction band smoothly transforms into the localised Na  $3s$  band. We thus conclude that the energetic position of the lowest conduction band is coupled to the Na  $3s$  state, but that the Na  $3s$  state hybridises with a Cl scattering state (that may be denoted as Cl  $4s$ ). This hybridisation is responsible for the dispersion of the conduction band. That the band structure of NaCl around the equilibrium lattice constant can be reproduced with fcc  $\text{Cl}^-$  alone can be explained by the fact that the scattering behaviour of the sodium core is similar to that of the vacuum in the relevant energy range. It does hence not mean that the Na would be irrelevant for the conduction states, but that a jellium background is a sufficient approximation for the sodium core in this very case.

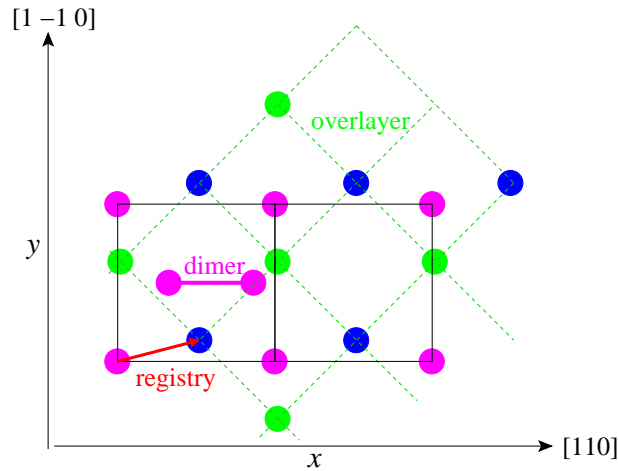


Figure D.4: Schematic representation of the NaCl overlayer on the Ge  $p(2 \times 1)$  dimer surface. Pink: Ge, Blue: Na, Green: Cl.

### D.3 Structure of the NaCl/Ge(001) interface

To investigate the interface between a sodium chloride overlayer and the Ge(001)  $p(2 \times 1)$  surface, a 2 ML NaCl film was put on top of the buckled-dimer Ge(001) surface with various lateral shifts ("registry"). Since there is no experimental evidence that the film thickness influences the interface, the interface structure was not investigated for other thicknesses. To get an overview over the potential energy surface, the lateral position of the top NaCl layer was kept fixed, while the vertical position of this layer was relaxed. The bottom NaCl and the top four Ge layers were relaxed, too. For the lateral position of the top layer ("registry", cf. Fig. D.4), three linescans along the dimers (x-direction) were performed at  $\Delta y=0$  (Cl above the dimer), 0.25 (low-symmetry), and 0.5 (Na above the dimer).<sup>3</sup> Higher y-offsets need not be considered due to the mirror symmetry of the NaCl overlayer and the substrate, respectively. Likewise, the x-offsets above 0.5 (in relative coordinates of the  $2 \times 1$  unit cell) have been obtained by symmetry considerations since the (unrelaxed) overlayer has a  $1 \times 1$  unit cell. We do not observe a dimer flip (down-up  $\leftrightarrow$  up-down) during the relaxation.

The linescans are shown in Fig. D.5. We display the adhesion energy of the 2 ML NaCl adlayer as a function of the registry; negative values correspond to binding of the overlayer. It can be clearly seen that the displacement along the dimer plays a very important role. All three linescans have a similar shape and agree for the position of the minima ( $x=0.1$ ) and maxima

<sup>3</sup>All coordinates are relative with respect to the  $2 \times 1$  unit cell.



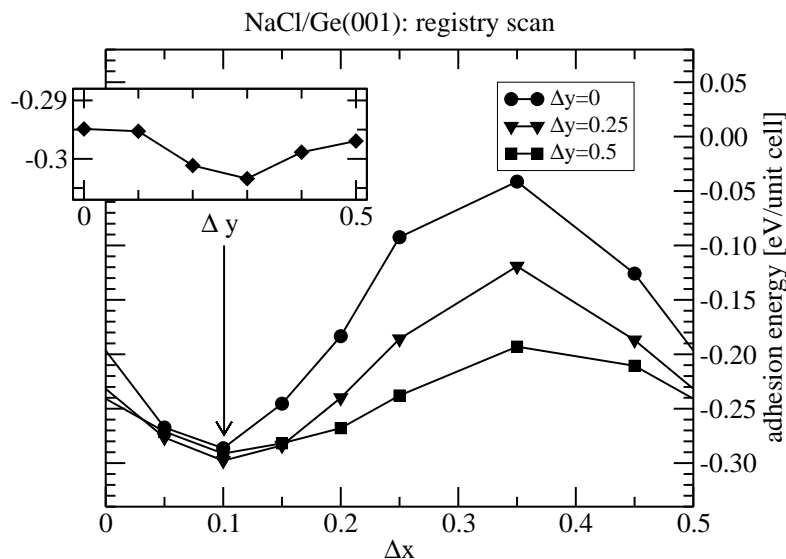


Figure D.5: Linescans for the NaCl registry above the Ge(001)  $2 \times 1$  surface along the dimers ( $\Delta x$ ) and in the perpendicular direction ( $\Delta y$ ) close to the minimum.

$\Delta y$	$\Delta x$	E
0	0.099	-0.310 eV
0.25	0.100	-0.306 eV
0.5	0.100	-0.301 eV

Table D.1: Minimum search for three displacements  $\Delta y$  perpendicular to the dimers ( $\Delta x$  is kept fixed).

( $x=0.35$ ). The height of the maximum depends on the displacement along  $[1\bar{1}0]$  perpendicular to the dimer axis, whereas the minima are all very close in energy. Therefore, a linescan along  $[1\bar{1}0]$  was performed for  $\Delta x = 0.1$ , which is shown in the inset in Fig. D.5. This linescan demonstrates that the potential energy surface is very flat along this direction. In order to find the minimum, a  $\Delta x$  relaxation with tighter convergence parameters<sup>4</sup> was started from three points  $\Delta y=0, 0.25$ , and  $0.5$ . The result shown in Table D.1 reveals that the minimum is at  $\Delta y=0$ , but is indeed very shallow in the  $y$ -direction. The adhesion energy at the minimum of  $0.31$  eV per unit cell agrees reasonably with the experimental estimate of  $0.26$  eV [131].

<sup>4</sup>The tighter convergence and the relaxation of the top layer NaCl along  $x$  are responsible for the small offset of  $5$  meV for  $\Delta y = 0.5$  between the inset of Fig. D.5 and Tab. D.1.

## D.4 STM simulations

According to Tersoff-Hamann approximation [153], the tunnelling current in STM is proportional to the LDOS at the Fermi energy at the position of the tip. To account for the experimental tip bias of 1.5 – 2.7 eV, we integrate the LDOS between the bias  $U$  and the Fermi energy. This partial density

$$\rho^{\text{STM}}(\mathbf{r}) = \int_U^0 dE \sum_{n\mathbf{k}} |\psi_{n\mathbf{k}}(\mathbf{r})|^2 \delta(E - \epsilon_{n\mathbf{k}}) \quad (\text{D.8})$$

should then reflect the STM tunnelling currents.<sup>5</sup> We use  $U = -2$  eV in the following. A variation of the energy range gives no indication that the STM pictures would change outside the experimental range. The experimental observation that only a small energy range allows for STM pictures cannot be explained from our results. An inspection of  $\rho^{\text{STM}}$  (not shown) at the surface of the NaCl films reveals that the bright spots in the STM must be assigned to the chloride ions. In conflict with the experimental  $1 \times 1$  pattern, the theoretical two-layer model invariably produces a  $2 \times 1$  STM pattern with a chloride corrugation of  $\sim 0.3$  Å. This result is also independent of the shift of the overlayer perpendicular to the dimers. Positioning the sodium atom above the top dimer atom and the chlorine atom between the dimers removes the geometrical corrugation of  $0.2$  Å at the surface and changes the total energy by only 10 meV, which may be below the absolute accuracy of our DFT-LDA approach. Nevertheless, the partial density shows a corrugation of  $\sim 0.2$  Å. We therefore do not believe that this discrepancy between the STM simulations and the experiment indicates an error in the atomistic model. Rather, deficiencies in the STM simulation, e.g. the use of the Tersoff-Hamann approximation or the neglect of atomic relaxation due to the tip-induced electric field may be responsible. The current approximations can hence not account for the observed STM pictures in full detail. However, we find strong evidence that the chlorine atoms appear as bright spots, since the Ge states hybridise only with these over the full valence energy range.

On the other hand, we can reproduce the apparent heights of the NaCl layers in the STM experiments. For this, we average  $\rho^{\text{STM}}$  laterally and plot it on a logarithmic scale (cf. Fig. D.6 bottom). Above the surface the density decays exponentially. The (local) decay constant  $\alpha(z)$  can be obtained via

$$\alpha(z) = \frac{\frac{d}{dz} \rho^{\text{STM}}(z)}{\rho^{\text{STM}}(z)} = \frac{d}{dz} \ln \rho^{\text{STM}}(z) \quad (\text{D.9})$$

---

<sup>5</sup>It is possible to introduce an energy-dependent weighting factor in the above integral, containing additional parameters. However, by varying  $U$  we observe that the main characteristics of the  $\rho^{\text{STM}}$  are not sensitive to the energy-dependence and we estimate that an energy-dependent weighting would not introduce qualitative differences.

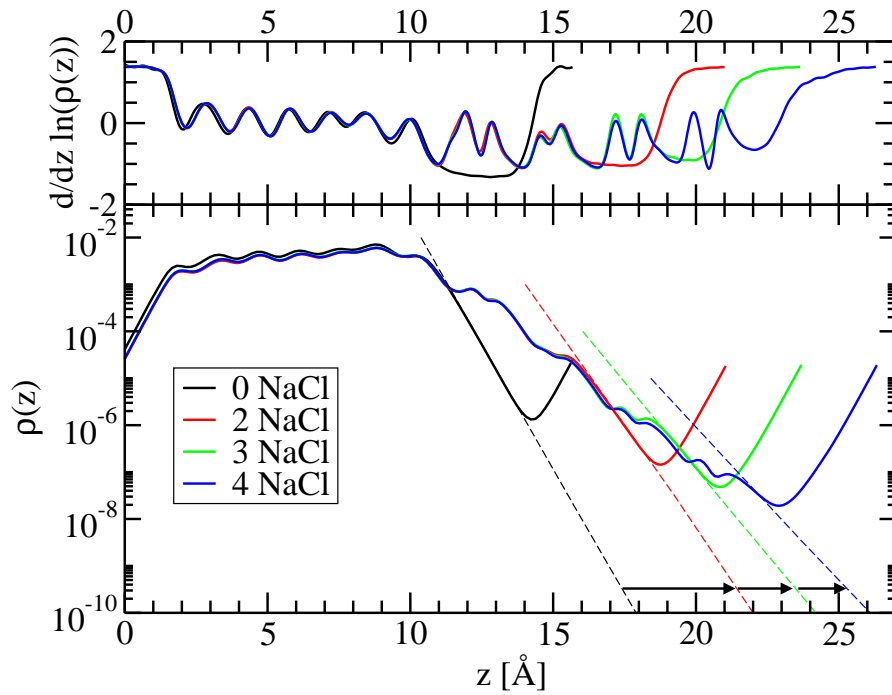


Figure D.6: Laterally averaged partial densities for the STM simulations for a 6 layer Ge slab with a NaCl of 2,3, and 4 layer thickness and without a NaCl layer (0 NaCl). The top figure shows the local decay constants. The dashed lines indicate the exponential decay, and the distance between the curves correspond to the apparent heights in STM. The density increase is an artifact due to the periodicity in the simulations.

and is shown in the top graph of Fig. D.6. When NaCl layers are adsorbed on the Ge substrate, the average density curves are shifted towards higher  $z$  as indicated by the black arrows in Fig. D.6. When the STM is used in constant current mode, an equidensity surface is scanned. We can therefore identify the shifts in the density decay curves (dashed lines) with increasing NaCl coverage with the experimentally determined apparent heights<sup>6</sup>. We obtain  $4.1 \pm 0.4 \text{ \AA}$  for the first double layer and  $2.2 \pm 0.4 \text{ \AA}$  for the next single layer, in reasonable agreement with the experimental values of  $3.8 \pm 0.3 \text{ \AA}$  and  $2.0 \pm 0.3 \text{ \AA}$  [133]. A closer inspection of the decay curves reveals an interesting explanation for the discrepancy between the geometrical height (i.e. the inter-layer separation,  $2.8 \text{ \AA}$ ) and the heights determined in STM ( $2.0 \text{ \AA}$ ). Around the chlorine atoms, the density curves exhibit plateau-like regions which can be attributed to the hybridisation of the tails of the Ge states with the chlorine  $3p$  states. Between the chloride layers, the density decays again exponentially with a constant similar to the vacuum (cf. Fig. D.6, top). The apparent height in STM is therefore mainly determined by the width of the hybridisation-induced plateaus, which is of course not directly linked to the chloride layer separation. This also implies that STM may not be able to detect vertical displacements of the ions within the film unless they affect the hybridisation width.

---

<sup>6</sup>Since the slope of the curves changes a little with film thickness (cf. Fig. D.6 top) the results depend on the density value chosen. The results are given for  $10^{-10} \text{ e/bohr}^3$  and change by  $\sim 0.2 \text{ \AA}$  per order order of magnitude.

# Appendix E

## Computer code developments

The *GW* space-time method as published in previous papers [73–75] has been implemented in the `gwst` code by Rex Godby and coworkers. When the Fortran code was made available to us, it became clear that

- the lack of program structure and its documentation would make extensions and modifications difficult,
- the code had been substantially modified with little or no documentations of the changes,
- major parts of the code had a suboptimal performance,
- the computation for large systems would be limited by the disk space requirements.<sup>1</sup>

In other words, the slab systems of this work were beyond the capabilities of this original version of the code. Here, we will explain how block algorithms were used to speed up the computation of the Green’s function and the inversion of dielectric matrices, and how the disk space consumption could be reduced by 30-50%.

Before addressing the details, we note a few things about the technical background. On modern computer systems, algebraic operations are often more efficient when performed on blocks. The reason behind this is that the basic computational steps are no longer the bottleneck of the computation, but rather the data transfer between the main memory and the central processing unit (CPU). Block algorithms exploit the fact that small, but considerably faster memory chips (level caches) are used to buffer the data transfer.

---

<sup>1</sup>The storage requirements for the non-local operators of large systems exceeds the typical main memory sizes by an order of magnitude, taking into account that the code is not parallelised and is therefore run on work stations with  $\sim 8$ GB main memory.

When all data for a calculation fits into this level cache, dramatic speedups are observed (often by a factor two or three), in particular for matrix-matrix operations. Unfortunately, most compilers are not able to transform standard high-level code (“plain Fortran”) into optimally blocked machine code, and usually specialised libraries (basic linear algebra subroutines = BLAS) must be used. BLAS routines are classified into level 1 (vector operations), level 2 (matrix-vector), and level 3 (matrix-matrix). The speedup compared to conventional computer code increases with the BLAS level since the higher levels profit more from data reuse.

A second point concerns the memory management. Since the full size of the two-point functions exceeds by far the main memory of the work stations used for the calculation<sup>2</sup>, only parts of it (“slices”) can be kept in the main memory. Currently unused data is stored on the hard disk, which is typically 10–100 times larger than the main memory. A slice contains all points for one or more active indices for only one point of the inactive indices. Let's take a simple example with only two indices A and B with  $N_A$  and  $N_B$  many points, respectively. For transformations on A (e.g. a Fourier transformation), we use  $N_A$ -sized A-slices for one particular value of B. After the transformation has been performed, the slice is written into a “scratchfile” on the hard disk. One can then reuse the memory to perform the same transformation for the next B until the full  $N_A \times N_B$  matrix has been transformed and written to the scratchfile. The data is then read back in a different order for the transformation of the previously inactive index B, i.e. we work with  $N_B$ -sized B-slices for one particular value of A. The limiting factor for the size of the two-point functions is then the available disk space. Moreover, the read/write operations take a significant amount of time and it is important to minimise these operations by carefully balancing the order of the transformations and the “layout”, i.e. the index ordering, of the scratchfile.

## E.1 Green's function

The construction of the Green's function is one of the time-critical steps in the  $GW$  space-time approach. In practice, the Green's function is computed in mixed space, i.e.  $G_{\mathbf{k}}(\mathbf{r}, \mathbf{r}', i\tau)$ . As explained above, this is done piecewise. In one step,  $G$  is computed for one specific  $\mathbf{r}$  and  $\mathbf{k}$ -point, but all possible  $\mathbf{r}'$

---

<sup>2</sup>The `gwst` code is a serial code, and therefore, parallel computers (shared memory model) or computer clusters (distributed memory model) which have much larger memories could not be used.

and  $\tau$ , by summing over bands

$$G_{\mathbf{k}}(\mathbf{r}, \mathbf{r}', i\tau) = \sum_n \psi_{n\mathbf{k}}^*(\mathbf{r}) \psi_{n\mathbf{k}}(\mathbf{r}') f_{n\mathbf{k}}(\tau) \quad (\text{E.1})$$

where  $\psi$  are the wavefunctions (which are already stored in a large block) and  $f_{n\mathbf{k}}(\tau)$  are the precomputed frequency factors  $\pm \exp(\pm \epsilon_{n\mathbf{k}} \tau)$ . In the original "plain Fortran" code, this operation was performed in a triple loop over  $\mathbf{r}'$ ,  $\mathbf{n}$ , and  $\tau$ , where the performance bottleneck is the memory access.

However, the band summation can be reinterpreted as a matrix-matrix multiplication. By precomputing an auxiliary  $N_{\text{bands}} \times N_\tau$  matrix

$$\eta(n, \tau) = \psi_{n\mathbf{k}}^*(\mathbf{r}) f_{n\mathbf{k}}(\tau), \quad (\text{E.2})$$

the summation over bands can be performed as a highly efficient matrix-matrix multiplication, schematically written as

$$G(N_{\mathbf{r}'} \times N_\tau) = \psi(N_{\mathbf{r}'} \times N_{\text{bands}}) \times \eta(N_{\text{bands}} \times N_\tau), \quad (\text{E.3})$$

which has been implemented as a BLAS level 3 call (`zgemm`). The resulting code is approximately two to three times faster than the original "plain Fortran" code.

For large  $N_{\text{bands}}$ , an additional trick allows to reduce the computational effort even more. The real-space representation of the wavefunctions  $\psi_{n\mathbf{k}}(\mathbf{r})$  is computed from their Fourier representation

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}} u_{n\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \quad (\text{E.4})$$

using Fast Fourier Transforms (FFTs). Since a spherical cutoff is used in reciprocal space, the number of  $\mathbf{G}$ -vectors is considerably smaller than the number of real-space points (typically  $N_{\mathbf{G}} \approx (0.3 - 0.5)N_{\mathbf{r}}$ ). As the Fourier transform for  $\mathbf{r}'$  commutes with the band summation, it is advantageous to perform the sum over bands in reciprocal space for only  $N_{\mathbf{G}}$  many points

$$G_{\mathbf{k}}(\mathbf{r}, \mathbf{G}', i\tau) = \sum_n \eta(n, \tau) u_{n\mathbf{k}}(\mathbf{G}'), \quad (\text{E.5})$$

and perform the Fourier transformation from  $\mathbf{G}'$  to  $\mathbf{r}'$  afterwards

$$G_{\mathbf{k}}(\mathbf{r}, \mathbf{r}', i\tau) = e^{i\mathbf{k}\cdot\mathbf{r}'} \sum_{\mathbf{G}'} G_{\mathbf{k}}(\mathbf{r}, \mathbf{G}', i\tau) e^{i\mathbf{G}'\cdot\mathbf{r}'} . \quad (\text{E.6})$$

Although this involves additional computational work for the Fourier transformation, the smaller matrix sizes for the expensive band summation reduce the overall computational cost significantly even for moderate  $N_{\text{bands}}$ . In practice, those  $\psi_{n\mathbf{k}}(\mathbf{r})$  that are required for computing  $\eta(n, \tau)$  are stored on disk and loaded on demand. This reciprocal-space summation is particularly helpful for high band cutoffs because in this case the *GW* calculation is dominated by the computation of the Green's function.

## E.2 Block inversion of Hermitean packed matrices

The inversion of the dielectric matrices is another computationally demanding step in a  $GW$  calculation that we were able to speed up considerably by developing and implementing a block algorithm. The standard Hermitean packed-matrix inversion in the linear algebra package originally employed (LAPACK) uses an iterative scheme which achieves only 20-40% of the peak performance of the CPU. In the block algorithm developed here, the computation is reformulated in matrix-matrix operations that make full use of the performance of the underlying BLAS library.

We will demonstrate the block algorithm in the following for 'L'-packed matrices that store the lower triangle of the matrix column-wise. The corresponding algorithm for 'U'-packed matrices, where the upper triangle is stored, works analogously. Schematically, the blocked inversion (with a block size  $N_b$ ) of a Hermitean  $N \times N$  matrix can be written as

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{a} & \mathbf{b}^\dagger \\ \mathbf{b} & \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0}^\dagger \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \quad (\text{E.7})$$

where  $\mathbf{A}$  is the  $N_b \times N_b$  top left submatrix of the original matrix,  $\mathbf{B}$  the  $N_r \times N_b$  bottom left submatrix, and  $\mathbf{C}$  the  $N_r \times N_r$  bottom right submatrix.  $N_r = N - N_b$  is the rest size of the matrix when the first  $N_b$  columns are separated. Small letters denote the corresponding submatrices of the inverse which are to be computed, and  $\mathbf{1}$  and  $\mathbf{0}$  on the right hand side are properly sized unit and zero matrices, respectively.

Equation (E.7) defines a set of four coupled matrix equations, which can be solved for the submatrices  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  of the inverse matrix:

$$\mathbf{c} = (\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^\dagger)^{-1}, \quad (\text{E.8})$$

$$\mathbf{b} = -\mathbf{c}\mathbf{B}\mathbf{A}^{-1}, \quad (\text{E.9})$$

$$\mathbf{a} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}^\dagger\mathbf{b}. \quad (\text{E.10})$$

The matrix inversion for the (small) matrix  $A$  is performed by a standard LAPACK routine. For the inversion in Eq. E.8, the blocked inversion routine is called recursively until the matrix is small enough for a standard LAPACK inversion. The size of the matrix is reduced by  $N_b$  for each level of recursion.

In practice, the result of the inversion is stored on the memory location of the original matrix. Moreover the intermediate steps should not require large amounts of additional memory. The scheme presented here requires only a  $N \times N_b$  work space, and uses also the input/output matrix memory



for intermediate results. The memory blocks, denoted by Greek letters, are listed in the following table:

name	size	storage	where
$\alpha$	$N_b \times N_b$	Hermitean packed	input/output
$\beta$	$N_r \times N_b$	general	input/output
$\gamma$	$N_r \times N_r$	Hermitean packed	input/output
$\delta$	$N_b \times N_b$	Hermitean unpacked	work space
$\epsilon$	$N_r \times N_b$	general	work space

In the beginning,  $\mathbf{A}$  and  $\mathbf{B}$  are stored unseparated in  $\alpha$  and  $\beta$ , indicated as  $\mathbf{A}/\mathbf{B}$ , and the same is true for  $\mathbf{a}$  and  $\mathbf{b}$  at the end.

The computational steps are listed in Table E.1. The mathematical operation to take is listed in the left column, the status of the various storage locations after the operation in the centre and the subroutine employed in the rightmost column. `hpgesub` either extracts (o) the bottom left submatrix from a Hermitean packed matrix, or stores it back (i). Likewise, `hphesub` performs the same operations for the top left square block (which is Hermitean). `hpinv` is the name of the blocked inversion routine that is called recursively to invert the  $N_r \times N_r$  matrix  $\gamma$ . If the size of the matrix is smaller than  $N_b$ , the standard LAPACK iterative inversion (`hptrf` + `hptri`) is performed instead; and likewise for the inversion of the Hermitean unpacked matrix  $\delta$ . `hemm` and `her2k` are standard BLAS matrix-matrix multiplication and rank-2 updates for Hermitean unpacked matrices. The corresponding subroutines for Hermitean packed matrices are missing from the standard BLAS and have been implemented using a block algorithm and available standard BLAS matrix-matrix routines (`hpmm_b` and `hpr2k_b`).

This blocked inversion algorithm has been implemented in the `gwst` code for the inversion of the dielectric matrix, one of the time-critical steps notably in large *GW* calculations. As expected from the performance of the underlying BLAS library, the blocked algorithm is faster by a factor 1.5–2.5 compared to the standard LAPACK routines because it uses matrix-matrix operations throughout.

The modifications described here (and others that are not described here) have greatly improved the computational efficiency of the `gwst` implementation. This is demonstrated in Fig. E.1. We observe a reduction in the overall run-time by a factor 3–5. The most important reduction of the computational time arises from the Green’s function, the previously dominant part in the calculation. After the improvements, the computational effort is typically equally distributed between 1) the computation of the Green’s function, 2) the Fourier transformation between the real-space/imaginary time and reciprocal-space/imaginary frequency representations, 3) the inversion

	$\alpha$	$\beta$	result $\gamma$	$\delta$	$\epsilon$	
		<b>A/B</b>	<b>C</b>	–	–	
$\delta \leftarrow \mathbf{A}$		<b>A/B</b>	<b>C</b>	<b>A</b>	–	hphesub(o)
$\epsilon \leftarrow \mathbf{B}$		<b>A/B</b>	<b>C</b>	<b>A</b>	<b>B</b>	hpgesub(o)
$\delta \leftarrow \delta^{-1}$		<b>A/B</b>	<b>C</b>	$\mathbf{A}^{-1}$	<b>B</b>	hetrf+hetri
$\alpha \leftarrow \delta$	$\mathbf{A}^{-1}$	–	<b>C</b>	$\mathbf{A}^{-1}$	<b>B</b>	hphesub(i)
$\beta \leftarrow \epsilon\delta$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	<b>C</b>	$\mathbf{A}^{-1}$	<b>B</b>	hemm
$\gamma \leftarrow \gamma - \epsilon\beta^\dagger$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	$\mathbf{C} - \mathbf{BA}^{-1}\mathbf{B}^\dagger$	–	<b>B</b>	hpr2k.b <sup>1</sup>
$\gamma \leftarrow \gamma^{-1}$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	<b>c</b>	–	–	hpinv <sup>2</sup> (Eq. E.8)
$\epsilon \leftarrow -\gamma\beta$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	<b>c</b>	–	<b>b</b>	hpmm.b <sup>1</sup> (Eq. E.9)
$\delta \leftarrow \alpha$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	<b>c</b>	$\mathbf{A}^{-1}$	<b>b</b>	hphesub(o)
$\delta \leftarrow \delta - \beta^\dagger\epsilon$	$\mathbf{A}^{-1}$	$\mathbf{BA}^{-1}$	<b>c</b>	<b>a</b>	<b>b</b>	her2k (Eq. E.10)
$(\alpha\beta) \leftarrow \delta$		<b>a/–</b>	<b>c</b>	<b>a</b>	<b>b</b>	hphesub(i)
$(\alpha\beta) \leftarrow \epsilon$		<b>a/b</b>	<b>c</b>	<b>a</b>	<b>b</b>	hpgesub(i)

<sup>1</sup> using  $\delta$  as work space      <sup>2</sup> using  $\epsilon$  as work space

Table E.1: Computational steps and memory usage for the blocked inversion of Hermitean packed matrices. See text.

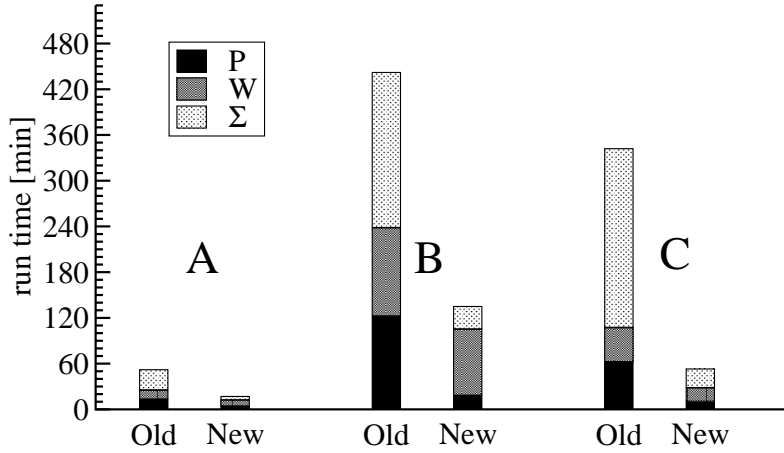


Figure E.1: Comparison of the original implementation of the space-time method and the improved version for three test cases: A)  $\text{Si}_4\text{H}_4$  slab, 1 band-structure point. B)  $\text{Na}_2\text{Cl}_2$  slab, 1 band structure point, band cutoff 5 Hartree. C) NaCl bulk, 8 band structure points, band cutoff 10 Hartree. The calculation can be separated into three distinct parts: computation of the polarisability in mixed space (P), computation of the screened interaction in mixed-space (W), computation of the self-energy and its matrix elements ( $\Sigma$ ). In P and  $\Sigma$ , the Green's function is computed for which a typical speed-up by a factor of 6 could be achieved.

of the dielectric matrices, and 4) the computation of the self-energy matrix elements.

### E.3 Disk storage

The size of the two-point functions in the  $GW$  formalism is in general too large to keep all the necessary data in the main memory. Most of the data is therefore kept on disk and only the currently required data is loaded into the main memory. After the optimisations described above (and many others not described here) the disk usage has proved to be the limiting factor for the calculations.

In the most disk-space efficient `grecomp-mode`<sup>3</sup> of the `gwst` program, an  $(\mathbf{r}', \mathbf{R}, i\tau)$ -slice of the Green's function for one special point  $\mathbf{r}_s$  is computed from the wavefunctions whenever it is needed to compute the corresponding slice of the polarisability  $P$  or the self-energy  $\Sigma$ . The storage of the polarisability/the screened interaction is then the critical point. The key to reducing the disk space requirements is again the observation that the reciprocal space representation is 2–3 times smaller than the real-space representation.

The successive transformation of  $P$  from real space and imaginary time to reciprocal space and imaginary frequency is listed in Tab. E.2 together with the scratchfiles used. The backward transformation of  $W^{\text{sr}}$  is completely analogous. In the original implementation (highlighted in red), the main scratchfile 'cordata' contains the data after the first transformational step  $\mathbf{R} \rightarrow \mathbf{k}^+$ . We have introduced an alternative path (highlighted in blue) where the code proceeds until the Fourier transformation  $\mathbf{r}' \rightarrow \mathbf{G}'$  before the data is stored on hard disk, thereby changing the content and the size of the scratchfile. Comparing the scratchfile size for the two alternatives listed in Tab. E.2, we find a common factor  $N_{\mathbf{r}}N_{\tau}$ , and differing factors  $N_{\mathbf{r}_s}N_{\mathbf{k}^+}$  for the original layout and  $N_{\mathbf{G}}N_{\mathbf{k}_s}$  for the new one. Which of the two is smaller is critically influenced by the symmetry reduction for  $\mathbf{r}$  and  $\mathbf{k}$ , respectively, which depends on the relative number of high-symmetry points. For high symmetries, the symmetry reduction is usually more effective for  $\mathbf{r}$  than for  $\mathbf{k}$ , which makes the original layout the optimal choice for highly symmetric bulk systems. However, when the number of symmetries is low as in our slab systems, the reciprocal-space reduction of  $N_{\mathbf{G}}/N_{\mathbf{r}}$  favours the new layout and reduces the scratchfile size by 30–50% compared to the original layout. Only these reductions in disk space have made it possible to thoroughly check the convergence parameters beyond the values that were finally found to be

<sup>3</sup>`grecomp` stands for Green's function recomputed. The alternative is to store the Green's function on hard disk, too, thereby doubling the disk space requirements.

	representation		scratchfile	scratchfile size
	$P(\mathbf{R}, \mathbf{r}_s, \mathbf{r}', i\tau)$			
FFT	$\downarrow$			
	$P(\mathbf{k}^+, \mathbf{r}_s, \mathbf{r}', i\tau)$	$\iff$	cordata	$N_{\mathbf{k}^+} \times N_{\mathbf{r}_s} \times N_{\mathbf{r}} \times N_{\tau}$
rotate	$\downarrow$		$\downarrow$	
	$P(\mathbf{k}_s, \mathbf{r}_s^*, \mathbf{r}', i\tau)$			
FFT	$\downarrow$			
	$F(\mathbf{k}_s, \mathbf{r}_s^*, \mathbf{G}', i\tau)$	$\iff$	cordata	$N_{\mathbf{k}_s} \times N_{\mathbf{r}} \times N_{\mathbf{G}} \times N_{\tau}$
reorder	$\downarrow$		$\uparrow$	
	$P(\mathbf{k}_s, \mathbf{r}, \mathbf{G}', i\tau)$	$\iff^b$	semifft	$N_{\mathbf{r}} \times N_{\mathbf{G}}$
FFT	$\downarrow$			
	$P(\mathbf{k}_s, \mathbf{G}, \mathbf{G}', i\tau)$			
map	$\downarrow$			
	$P(\mathbf{k}_s, [\mathbf{G}\mathbf{G}'], i\tau)$	$\iff^a$	file_tw	$N_{\mathbf{G}\mathbf{G}'} \times \max(N_{\tau}, N_{\omega})$
FT $\tau \rightarrow \omega$	$\downarrow$			
	$P(\mathbf{k}_s, [\mathbf{G}\mathbf{G}'], i\omega)$	$\iff^a$	file_tw	$N_{\mathbf{G}\mathbf{G}'} \times \max(N_{\tau}, N_{\omega})$

<sup>a</sup> use of this scratchfile can be switched off ('twbio' option).

<sup>b</sup> improvement: not used when sufficient memory is available.

Indices:

- $\mathbf{R}$  real space lattice vector
- $\mathbf{k}^+$  Brillouin zone vector, reduced by time reversal symmetry
- $\mathbf{k}_s$  special (symmetry-reduced) Brillouin zone vector
- $\mathbf{r}$  real space vector (in unit cell)
- $\mathbf{r}_s$  special (symmetry-reduced) vector (in unit cell)
- $\mathbf{r}_s^*$  star of a special (symmetry-reduced) vector (in unit cell)
- $\mathbf{G}$  reciprocal space lattice vector
- $[\mathbf{G}\mathbf{G}']$  packed form of reciprocal space lattice vector matrix

Table E.2: Use of scratchfiles in the transformations from real space and imaginary time to reciprocal space and imaginary frequency. The red and blue colours refer to the old and new layout.

sufficient. In particular the plane-wave cutoff is a critical parameter, since the size of the two-point functions scales cubically with the cutoff.

## E.4 Efficient computation of the matrix elements of $\mathbf{r}$

Head and wings of the dielectric matrix at the  $\Gamma$ -point cannot be computed from the numerical polarisability matrix and the Coulomb potential directly. The small- $\mathbf{k}$  asymptotic behaviour of the polarisability ( $k^2$  for the head,  $k$  for the wings) cancels with the corresponding singularity of the Coulomb potential and its square root, respectively. Therefore, these elements are obtained directly from the Kohn-Sham wavefunctions using a  $\mathbf{k} \cdot \mathbf{p}$  perturbation approach [59, 86, 154]. For this purpose, the matrix elements of the position operator  $\mathbf{r}$  are required, which are in practice calculated via the commutator of  $\mathbf{r}$  with the Kohn-Sham Hamiltonian  $h^{\text{KS}}$

$$\langle \psi_{c\mathbf{q}} | \mathbf{r} | \psi_{v\mathbf{q}} \rangle = \frac{\langle \psi_{c\mathbf{q}} | [h^{\text{KS}}, \mathbf{r}] | \psi_{v\mathbf{q}} \rangle}{\epsilon_{c\mathbf{q}} - \epsilon_{v\mathbf{q}}} . \quad (\text{E.11})$$

The Kohn-Sham pseudopotential Hamiltonian consists of three parts: the effective potential, the kinetic energy and the non-local pseudopotential. The effective potential commutes with  $\mathbf{r}$ . The contribution from the kinetic-energy operator  $\frac{1}{2}\mathbf{p}^2$  is trivial to compute. Exploiting the commutator identity  $[AB, C] = A[B, C] + [A, C]B$  with  $A = B = \mathbf{p}$  and  $C = \mathbf{r}$  yields

$$[\frac{1}{2}\mathbf{p}^2, \mathbf{r}] = \frac{1}{2} (\mathbf{p}[\mathbf{p}, \mathbf{r}] + [\mathbf{p}, \mathbf{r}]\mathbf{p}) = -i\mathbf{p} , \quad (\text{E.12})$$

where we have made use of the fundamental commutator  $[\mathbf{p}, \mathbf{r}] = -i$ . This is readily implemented for a plane-wave basis. The contribution from the non-local pseudopotential  $V_{\text{nl}}$  is more cumbersome and has often been neglected in earlier calculations. We will show in the following that it can be computed efficiently in a separable expression.

In its separable Kleinman-Bylander form [46], the non-local pseudopotential operator is written in the Dirac notation as

$$V_{\text{nl}} = \sum_{\mu} |\chi_{\mu}\rangle E_{\mu} \langle \chi_{\mu}| , \quad (\text{E.13})$$

where  $\mu$  is a composed index  $\{\mathbf{R}_{\mu}, n_{\mu}, l_{\mu}, m_{\mu}\}$  that runs over all pseudopotential projectors while  $\chi_{\mu}$  is in general given in a radial basis around a certain atomic position  $\mathbf{R}_{\mu}$ , i.e.,

$$\chi_{\mu}(\mathbf{r}) = f_{n_{\mu}l_{\mu}}(|\mathbf{r} - \mathbf{R}_{\mu}|) Y_{l_{\mu}m_{\mu}}(\Omega_{\mathbf{r}-\mathbf{R}_{\mu}}) . \quad (\text{E.14})$$

$\mu$  can additionally run over chemical species, which does not alter the following derivation, except that  $f_{nl}$  additionally depends on the species. We will now show that the matrix elements can be written in separable form, which reduces the scaling to be linear in the number of plane waves instead of quadratic as demonstrated in a previous approach [88].

To this end consider the commutator of  $\mathbf{r}$  with a single projector:

$$\begin{aligned}
& (|\chi_\mu\rangle E_\mu \langle \chi_\mu | \mathbf{r}) - (\mathbf{r} | \chi_\mu\rangle E_\mu \langle \chi_\mu |) \\
&= E_\mu \left[ (|\chi_\mu\rangle \langle \chi_\mu | \mathbf{r}) - |\chi_\mu\rangle \mathbf{R}_\mu \langle \chi_\mu | + |\chi_\mu\rangle \mathbf{R}_\mu \langle \chi_\mu | - (\mathbf{r} | \chi_\mu\rangle \langle \chi_\mu |) \right] \\
&= E_\mu \left[ |\chi_\mu\rangle \langle \chi_\mu | (\mathbf{r} - \mathbf{R}_\mu) - (\mathbf{r} - \mathbf{R}_\mu) |\chi_\mu\rangle \langle \chi_\mu | \right] \tag{E.15}
\end{aligned}$$

Now we make use of the fact that  $\mathbf{r} - \mathbf{R}_\mu$  can be expressed in the same radial basis as  $\chi_\mu$

$$[\mathbf{r} - \mathbf{R}_\mu]_\alpha = |\mathbf{r} - \mathbf{R}_\mu| \sum_{m=-1}^1 c_{\alpha m} Y_{1m}(\Omega_{\mathbf{r}-\mathbf{R}_\mu}), \tag{E.16}$$

where  $\alpha \in \{x, y, z\}$  are the spatial directions, and  $c_{\alpha m}$  yield the spatial components of the spherical harmonics for  $l = 1$ :

$c_{\alpha m}$	$\alpha = x$	$\alpha = y$	$\alpha = z$
$m = -1$	$\frac{1}{\sqrt{2}}$	$\frac{i}{\sqrt{2}}$	0
$m = 0$	0	0	1
$m = 1$	$-\frac{1}{\sqrt{2}}$	$\frac{i}{\sqrt{2}}$	0

We can then write the product in the radial basis, too,

$$\begin{aligned}
|\chi_\mu^\alpha\rangle &:= [\mathbf{r} - \mathbf{R}_\mu]_\alpha |\chi_\mu\rangle \\
&= |\mathbf{r} - \mathbf{R}_\mu| \sum_{m=-1}^1 c_{\alpha m} Y_{1m}(\Omega_{\mathbf{r}-\mathbf{R}_\mu}) f_{n_\mu l_\mu}(|\mathbf{r} - \mathbf{R}_\mu|) Y_{l_\mu m_\mu}(\Omega_{\mathbf{r}-\mathbf{R}_\mu}) \\
&= \sum_{L=l_\mu \pm 1} \sum_{M=m_\mu - 1}^{m_\mu + 1} c_{\alpha M, m_\mu}^{L, l_\mu} f_{n_\mu l_\mu}^r(|\mathbf{r} - \mathbf{R}_\mu|) Y_{LM}(\Omega_{\mathbf{r}-\mathbf{R}_\mu}) \tag{E.17}
\end{aligned}$$

with

$$f_{nl}^r(\rho) = \rho f_{nl}(\rho), \tag{E.18}$$

$$c_{\alpha M, m}^{L, l} = \sum_{m'} c_{\alpha m'} (l m 1 m' | L M). \tag{E.19}$$

where  $(l m 1 m' | L M)$  is a Clebsch–Gordan coefficient. It is convenient to expand  $\chi_\mu^\alpha$  in a plane-wave basis similar to what is done for  $\chi_\mu$ . When the radial functions  $f_{nl}$  are given on a radial grid [43],  $f_{nl}^r$  is trivial to compute and

the same routines that are used to compute  $\chi_\mu(\mathbf{G}+\mathbf{k})$  in the DFT calculation can be employed for the summands in  $\chi_\mu^\alpha(\mathbf{G}+\mathbf{k})$ . It must be emphasised that the sums over  $L$  and  $M$  contain only a very small number of non-zero terms (at most six).

The final formula is thus again a separable expression

$$[V_{\text{nl}}, r_\alpha] = \sum_\mu E_\mu \left( |\chi_\mu\rangle\langle\chi_\mu^\alpha| - |\chi_\mu^\alpha\rangle\langle\chi_\mu| \right). \quad (\text{E.20})$$

The computational effort to set up a full  $N_v \times N_c$  matrix for all three directions requires  $(1 + 3)N_{\mathbf{G}}N_\mu(N_v + N_c)$  operations to calculate the  $\langle\chi_\mu^\alpha|\psi_{v/c}\rangle$  projections and  $6N_\mu N_v N_c$  operations to build up the 3 matrices from the projections in Equation (E.20). The scaling is thus linear in the number of  $\mathbf{G}$ -vectors  $N_{\mathbf{G}}$  and not quadratic [88].

Only this efficient algorithm for the matrix elements of the position operator made the accurate computation of the dielectric tensor for the slab systems of interest feasible.

# Appendix F

## Computational parameters

### F.1 Pseudopotentials

The pseudopotentials used in this work were generated with the `fhi98PP` program [43]. The parameters were varied from the default parameters in most cases to improve the accuracy and efficiency of the pseudopotentials. The optimised parameters are listed in the following table.

	occupation			$r_c^a$			$l_{loc}$	remarks
	s	p	d	s	p	d		
Al	2	1	-	1.05(H)	(H)	1.4(H)	p	
Si	2	2	-	(H)	(H)	(H)	p	
Hf	2	6	2	0.75(H)	1.60(T)	0.9(H)	s	5-shell, 6s empty
O	1.7	4.8	-	1.35(T)	1.7(T)	(T)	d	
Ge	2	2	-	1.2(H)	(H)	(H)	p	
Na	1	0	-	2.7(T)	1.6(H)	2.5(T)	s	$\rho_{cut}^b=1.8$
Cl	2	5	-	1.05(H)	1.1(H)	1.8(T)	p	
Mo	1	0	5	1.6(H)	(H)	2.44(T)	s	valence: 4d,5s,5p

<sup>a</sup> All radii in bohr. (T)=Troullier-Martins, (H)=Hamann. If no radius is specified, the `fhi98PP` default was used.

<sup>b</sup> density cut-off radius for non-linear core corrections.

### F.2 DFT-LDA calculations

For DFT-LDA plane-wave calculations, there are two important convergence parameters: the plane-wave cutoff (that mainly depends on the pseudopotentials used) and the Monkhorst-Pack  $\mathbf{k}$ -point folding which depends on the material investigated. These convergence parameters have been tested for each system and the parameters employed for the SCF calculations are listed



in the following table:

system	$E_{\text{cut}}$ [Ry]	$\mathbf{k}$ -point grid
SiO <sub>2</sub> bulk	60	$3 \times 3 \times 3$
SiO <sub>2</sub> slabs	60	$3 \times 3 \times 1$
Mo bulk	60	$22 \times 22 \times 22$
SiO <sub>2</sub> /Mo	60	$6 \times 6 \times 1$
Al <sub>2</sub> O <sub>3</sub> bulk	60	$3 \times 3 \times 3$
Al <sub>2</sub> O <sub>3</sub> slabs	60	$3 \times 3 \times 1$
HfO <sub>2</sub> bulk	60	$4 \times 4 \times 4$
HfO <sub>2</sub> slabs	60	$4 \times 4 \times 1$
NaCl bulk	40	$3 \times 3 \times 3$
NaCl slabs	40	$4 \times 4 \times 1$
Ge bulk	20	$4 \times 4 \times 4$
NaCl/Ge	40	$6 \times 3 \times 1$

## F.3 $G_0W_0$ calculations

### F.3.1 Time-frequency grids

The time-frequency grids are characterised by two parameters: the maximum time or frequency for the numerical part and the number of Gauss-Legendre points used per half-axis [75]. We have tested the grids for each of the bulk systems investigated in this work, i.e. Ge, SiO<sub>2</sub>, NaCl, Al<sub>2</sub>O<sub>3</sub>, and HfO<sub>2</sub>. In all cases, a maximum of 6 atomic units and 15 Gauss-Legendre points proved to be sufficient to achieve an accuracy below 0.05 eV.

### F.3.2 System-dependent convergence parameters

The following table summarises further convergence parameters for the  $G_0W_0$  calculations. They were tested and give results within 0.05 eV for each parameter, except for NaCl/Ge where the achieved accuracy is 0.05–0.1 eV.

We will briefly comment on the various parameters. The plane-wave (pw) cutoff determines the real-space resolution of the two-point functions and is the most critical parameter for the computational effort and the required memory and disk space. The  $G_0W_0$  calculation in the space-time method scales cubically with the pw cutoff (in reciprocal space, the computational effort in some parts even scales with a power 4.5). The quasiparticle energies usually vary non-monotonously with the pw cutoff, the absolute accuracy can therefore only be estimated. Usually, the bare exchange part of the self-energy requires a higher pw cutoff than the correlation part, which can be

parameter	NaCl bulk	NaCl slabs	NaCl/Ge
pw cutoff $\Sigma_x$ [Hartree]	14	14	10
pw cutoff $\Sigma_c$ [Hartree]	14	10	7
<b>k</b> -sampling	$4 \times 4 \times 4$	–	$6 \times 3$
<b>k</b> -sampling head/wings	$4 \times 4 \times 4$	$4 \times 4$	$10 \times 10$
bcut head/wings [Hartree]	2	2	1

Table F.1: Convergence parameters for the  $GW$  calculations (see text).

used to drastically reduce the computational effort when the two parts are determined independently.

Head and wings of the dielectric matrix at the  $\Gamma$ -point are computed via perturbation theory directly from the Kohn-Sham wavefunctions [86]. Since the convergence behaviour for the band cutoff and the **k**-point sampling differs between head, wings, and body of the dielectric matrix [86, 154], we determine the parameters for head and wings independently. Typically, we converge the dielectric tensor at the smallest frequency to 0.3–1% and assume that this is sufficient for the wings, too. Earlier test calculations in our group indicate that variations in the dielectric tensor of up to 10% introduce errors below 0.1 eV for semiconductor systems [56]. We note that head and wings include the contributions of the non-local pseudopotentials [86] for which we have developed a highly efficient algorithm, see Sec. E.4. This is essential to obtain a consistent dielectric tensor for the anisotropy treatment, cf. Sec. 3.2.2.

The importance of the **k**-point sampling in slab systems has been discussed in Sec. 3.3.2 and been explicitly studied for the free-standing NaCl films. For NaCl/Ge, the extrapolation technique was used only for the 2ML case to determine the correction term given in Sec. 5.2.3, otherwise the  $6 \times 3 \times 1$  sampling as listed in the table was used.

### F.3.3 Band cutoff

The convergence of NaCl with respect to the band cutoff *bcut* is shown in Fig. F.1 for the bulk and a slab, respectively. As for most systems, the convergence shows a  $1/bcut$  behaviour and the data is plotted correspondingly. Fig. F.1 illustrates that NaCl requires a band cutoff of about 10 Hartree for a convergence to within 0.1 eV of the extrapolated value. This is much higher than for semiconductors, where 2–4 Hartree prove to be sufficient for this accuracy, see e.g. [23, 56, 74]. However, an absolute convergence is usually not required. Using a simple linear extrapolation yields highly reliable results in comparison with extended convergence tests for bulk systems. Nevertheless,

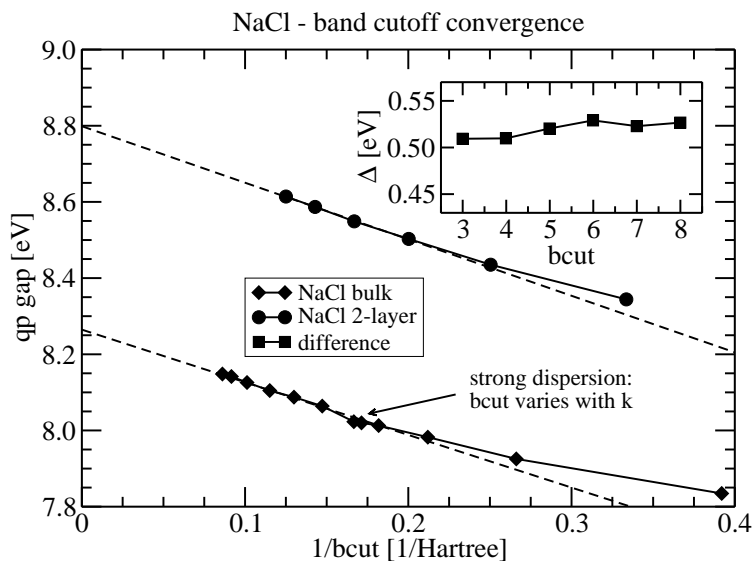


Figure F.1: Convergence of the quasiparticle gap for NaCl bulk and a 2-layer slab. The inset shows the difference  $\Delta$  between the two curves and demonstrates that it is practically independent of  $bcut$ .

the data in Fig. F.1 shows also that a careful inspection of the results is necessary in some cases. In practice, the band summations for the Green's function are not truncated at a certain energy, but at a certain band index. The highest band energy of all  $\mathbf{k}$ -points then defines the nominal band cutoff which was used in the plot. In NaCl, there are highly dispersive bands around 5-6 Hartree which vary in energy by  $\sim 0.5$  Hartree. Correspondingly, the effective band cutoff depends strongly on the  $\mathbf{k}$ -point and differs considerably from the nominal band cutoff. It must be emphasized that this is a rather rare situation typical for small high-symmetry systems and does not affect the converged result.

Furthermore, we note that the underconvergence of the quasiparticle gap appears to be largely independent of the long-range order. It can be clearly seen that convergence curves for the bulk and slab are essentially parallel. This becomes obvious when considering the difference between the two curves, shown in the inset in Fig. F.1. The variations are below 0.025 eV and result mainly from the variations in the bulk. We can therefore extract the underconvergence with respect to the band cutoff from the bulk and transfer it to the slab systems, thus reducing the computational effort drastically. For the free-standing slab systems, a band cutoff of 5 Hartree was used and the correction to obtain the extrapolated value amounts to 0.28 eV.

Since the bare exchange self-energy and the DFT exchange-correlation

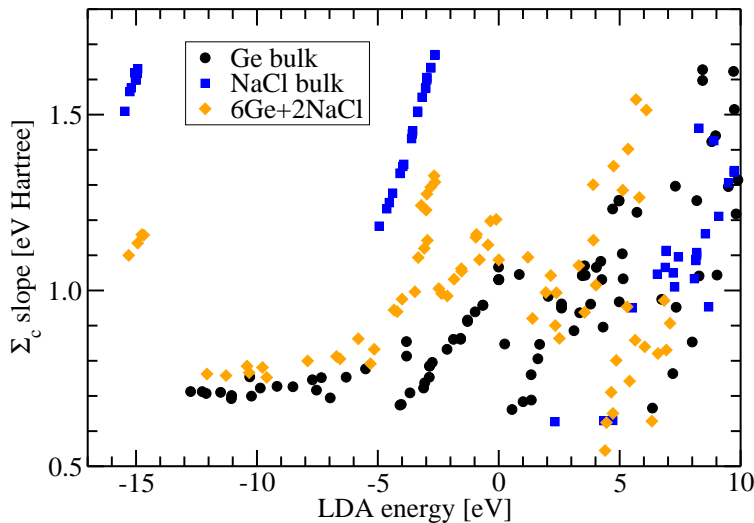


Figure F.2: Comparison of the  $\Sigma_c$  slope parameter for three systems in comparison: bulk NaCl, bulk Ge, and a Ge-supported NaCl film system.

potential are independent of the band summation, the variations exclusively affect the matrix elements of the correlation self-energy. For the supported films, we directly assume a  $1/bcut$  behaviour

$$\Sigma_c(bcut) = \Sigma_c(\infty) + \frac{A}{bcut}. \quad (\text{F.1})$$

Using a few cutoff energies, we can then determine the slope parameter  $A$  for each band from a linear regression. We note that the  $bcut$ -dependence of the gap results from the difference of the slope parameter for the band edge states, but does not depend on the absolute value. Indeed, comparing the slope parameters between different systems reveals that the low band cutoffs in the semiconductors must be attributed to the similarity of the slope parameters for the different bands (cf. F.2). The absolute values have a similar magnitude as for the ionic systems, about 0.5–2 eV·Hartree. For the calculations presented in Sec. 5.2.3, we then used a band cutoff of 3 Hartree.

# Bibliography

- [1] S. A. Chambers, Surf. Sci. Rep. **39**, 105 (2000).
- [2] D. Goodman, J. Vac. Sci. Technol. A **14**, 1526 (1996).
- [3] J. Libuda and H.-J. Freund, Surf. Sci. Rep. **57**, 157 (2005).
- [4] J. Repp, G. Meyer, F. E. Olsson, and M. Persson, Science **305**, 493 (2004).
- [5] J. Repp, G. Meyer, S. Paavilainen, F. E. Olsson, and M. Persson, Science **312**, 1196 (2006).
- [6] J. Repp and G. Meyer, Appl. Phys. A **85**, 399 (2006).
- [7] H.-J. Freund, M. Bäumer, J. Libuda, T. Risse, G. Rupprechter, and S. Shaikhutdinov, J. Catal. **216**, 223 (2003).
- [8] D. Goodman, J. Catal. **216**, 213 (2002).
- [9] S. Fölsch, U. Barjenbruch, and M. Henzler, Thin Solid Films **172**, 123 (1989).
- [10] J. Repp, S. Fölsch, G. Meyer, and K.-H. Rieder, Phys. Rev. Lett. **86**, 252 (2001).
- [11] R. Franchy, Surf. Sci. Rep. **38**, 195 (2000).
- [12] H.-J. Freund, Surf. Sci. **500**, 271 (2002).
- [13] J. Weissenrieder, S. Kaya, J.-L. Lu, H.-J. Gao, H.-J. F. S. Shaikhutdinov, M. Sierka, T. Todorova, and J. Sauer, Phys. Rev. Lett. **95**, 076103 (2005).
- [14] L. Giordano, D. Ricci, G. Pacchioni, and P. Ugliengo, Surf. Sci. **584**, 225 (2005).

- 
- [15] G. Kresse, M. Schmid, E. Napetschnig, M. Shishkin, L. Köhler, and P. Varga, *Science* **308**, 1440 (2005).
- [16] M. S. Chen, A. K. Santra, and D. W. Goodman, *Phys. Rev. B* **69**, 155404 (2004).
- [17] S. Wendt, E. Ozensoy, T. Wei, M. Frerichs, Y. Cai, M. S. Chen, and D. W. Goodman, *Phys. Rev. B* **72**, 115409 (2005).
- [18] M. Kulawik, N. Nilius, and H.-J. Freund, *Phys. Rev. Lett.* **96**, 036103 (2006).
- [19] M. C. Wu, J. S. Corneille, C. A. Estrada, J. W. He, and D. W. Goodman, *Chem. Phys. Lett.* **182**, 472 (1991).
- [20] S. Altieri, L. H. Tjeng, F. C. Voogt, T. Hibma, and G. A. Sawatzky, *Phys. Rev. B* **59**, R2517 (1999).
- [21] M. Kiguchi, T. Goto, K. Saiki, T. Sasaki, Y. Iwasawa, and A. Koma, *Surf. Sci.* **512**, 97 (2002).
- [22] G. D. Wilk, R. M. Wallace, and J. M. Anthony, *Appl. Phys. Rev.* **89**, 5243 (2001).
- [23] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, *Solid State Phys.: Advances in Research and Applications* **54**, 1 (2000).
- [24] P. Eggert, *Theoretische Untersuchung von Vielteilcheneffekten auf Silizium-Halbleiteroberflächen*, PhD thesis, Freie Universität Berlin, 2005.
- [25] S. Ismail-Beigi, *Phys. Rev. B* **73**, 233103 (2006).
- [26] C. D. Spataru, S. Ismail-Beigi, L. X. Benedict, and S. G. Louie, *Appl. Phys. A* **78**, 1129 (2004).
- [27] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, *Phys. Rev. B* **73**, 205119 (2006).
- [28] J. Repp, G. Meyer, S. M. Stojkovic, A. Gourdon, and C. Joachim, *Phys. Rev. Lett.* **94**, 026803 (2005).
- [29] K. Glöckler, M. Solokowski, A. Soukopp, and E. Umbach, *Phys. Rev. B* **54**, 7705 (1996).

- 
- [30] P. Rinke, K. Delaney, P. García-González, and R. W. Godby, *Phys. Rev. A* **70**, 063201 (2004).
- [31] J. B. Neaton, M. S. Hybertsen, and S. G. Louie, *Phys. Rev. Lett.* **97**, 216405 (2006).
- [32] M. Verstraete and R. W. Godby, unpublished.
- [33] M. Born and R. Oppenheimer, *Ann. Phys.* **84**, 457 (1927).
- [34] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [35] M. Levy, *Proc. Natl. Acad. USA* **76**, 6062 (1979).
- [36] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [37] W. Kohn and K. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [38] W. J. Carr, *Phys. Rev.* **122**, 1437 (1961).
- [39] M. G.-M. und K. A. Brückner, *Phys. Rev.* **106**, 364 (1957).
- [40] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [41] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Review of Modern Physics* **64**, 1045 (1992).
- [42] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, *Comp. Phys. Comm.* **107**, 187 (1997).
- [43] M. Fuchs and M. Scheffler, *Comput. Phys. Commun.* **119**, 67 (1999).
- [44] S. Boeck and J. Neugebauer, editors, *SFHI<sub>ng</sub>X manual*, 2003.
- [45] F. Bloch, *Z. Physik* **52**, 555 (1928).
- [46] L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982).
- [47] D. R. Hamann, *Phys. Rev. Lett.* **42**, 662 (1979).
- [48] N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).
- [49] D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
- [50] P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
- [51] H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **13**, 5188 (1976).

- 
- [52] J. B. Pendry, Surf. Sci. **57**, 679 (1976).
- [53] W. Bardyszewski and L. Hedin, Physica Scripta **32**, 439 (1985), and references therein.
- [54] M. Hase, M. Kitajima, A. M. Constantinescu, and H. Petek, Nature **426**, 51 (2003).
- [55] J. F. Janak, Phys. Rev. B. **18**, 7165 (1978).
- [56] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).
- [57] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).
- [58] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).
- [59] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
- [60] L. Hedin, Phys. Rev. **139**, A796 (1965).
- [61] M. Rohlfing, *Quasiteilchen-Bandstrukturen von Halbleitern und Halbleiter-Oberflächen*, PhD thesis, Universität Münster, 1996.
- [62] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
- [63] B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).
- [64] K. Delaney, P. Garcia-Gonzalez, A. Rubio, P. Rinke, and R. W. Godby, Phys. Rev. Lett. **93**, 249701 (2004).
- [65] W.-D. Schöne and A. G. Eguluz, Phys. Rev. Lett. **83**, 242 (1998).
- [66] W. Luo, S. Ismael-Beigi, M. L. Cohen, and S. G. Louie, Phys. Rev. B **66**, 195215 (2002).
- [67] A. Fleszar and W. Hanke, Phys. Rev. B **71**, 045207 (2005).
- [68] M. Marsili, O. Pulci, F. Bechstedt, and R. D. Sole, Phys. Rev. B **72**, 115415 (2005).
- [69] M. van Schilfgarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
- [70] F. Bruneval, N. Vast, and L. Reining, Phys. Rev. B **74**, 045102 (2006).



- [71] I. White, R. W. Godby, M. Rieger, and R. Needs, *Phys. Rev. Lett.* **80**, 4265 (1998).
- [72] O. Pulci, F. Bechstedt, G. Onida, R. D. Sole, and L. Reining, *Phys. Rev. B* **60**, 16758 (1999).
- [73] H. N. Rojas, R. W. Godby, and R. J. Needs, *Phys. Rev. Lett.* **74**, 1827 (1995).
- [74] M. M. Rieger, L. Steinbeck, I. D. White, H. N. Rojas, and R. W. Godby, *Comput. Phys. Commun.* **117**, 211 (1999).
- [75] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. D. White, and R. W. Godby, *Comput. Phys. Commun.* **125**, 105 (2000).
- [76] V. Chis and B. Hellsing, *Phys. Rev. Lett.* **93**, 226103 (2004).
- [77] G. Makov and M. C. Payne, *Phys. Rev. B* **51**, 4014 (1995).
- [78] J. Neugebauer and M. Scheffler, *Phys. Rev. B* **46**, 16067 (1992).
- [79] M. Rohlfing, P. Krüger, and J. Pollmann, *Phys. Rev. B* **52**, 1905 (1995).
- [80] T. Kotani and M. van Schilfgaarde, *Solid State Commun.* **121**, 461 (2002).
- [81] R. Gomez-Abal, X. Li, and M. Scheffler, LAPW+lo *GW* code in development.
- [82] S. Lebègue, B. Arnaud, M. Alouani, and P. E. Bloechl, *Phys. Rev. B* **67**, 155208 (2003).
- [83] R. Hott, *Phys. Rev. B* **44**, 1057 (1991).
- [84] O. Pulci, G. Onida, and R. D. Sole, *Phys. Rev. Lett.* **81**, 5374 (1998).
- [85] B. Wenzien, G. Cappellini, and F. Bechstedt, *Phys. Rev. B* **51**, 14701 (1994).
- [86] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler, *Comp. Phys. Commun.* **176**, 1 (2007).
- [87] R. M. Pick, M. H. Cohen, and R. M. Martin, *Phys. Rev. B* **1**, 910 (1970).

- [88] V. Olevano, M. Palumbo, G. Onida, and R. D. Sole, Phys. Rev. B **60**, 14224 (1999).
- [89] V. I. Lebedev and D. N. Laikov, Doklady Mathematics **59**, 477 (1999), Fortran code is distributed through CCL (<http://www.ccl.net/>).
- [90] J. D. Jackson, *Classical electrodynamics*, p. 767, Wiley, 1975.
- [91] L. Giordano, J. Goniakowski, and G. Pacchioni, Phys. Rev. B **67**, 045410 (2003).
- [92] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
- [93] R. T. Poole, J. G. Jenkin, J. Liesegang, and R. C. G. Leckey, Phys. Rev. B **11**, 5179 (1975).
- [94] A. F. Holleman and N. Wiberg, *Lehrbuch der anorganischen Chemie*, de Gruyter, Berlin - New York, 101st edition, 1995.
- [95] T. Demuth, Y. Jeanvoine, J. Hafner, and J. G. Ángyán, J. Phys.: Condens. Matt. **11**, 3833 (1999).
- [96] U. Bardi, A. Atrei, and G. Roviida, Surf. Sci. **268**, 87 (1992).
- [97] P. Gassmann, R. Franchy, and H. Ibach, Surf. Sci. **319**, 95 (1994).
- [98] R. M. Jaeger, H. Kuhlenbeck, H.-J. Freund, M. Wuttig, W. Hoffmann, R. Franchy, and H. Ibach, Surf. Sci. **259**, 235 (1991).
- [99] J. Aarik, A. Aidla, H. Mändar, T. Uustare, K. Kukli, and M. Schuisky, Appl. Surf. Sci. **173**, 15 (2001).
- [100] S. Van Elshocht *et al.*, J. Electrochem. Soc. **151**, F228 (2004).
- [101] J.-W. He, X. Xu, J. S. Corneille, and D. W. Goodman, Surf. Sci. **279**, 119 (1992).
- [102] X. Xu and D. W. Goodman, Appl. Phys. Lett. **61**, 774 (1992).
- [103] X. Xu and D. W. Goodman, Surf. Sci. **282**, 323 (1993).
- [104] T. Schroeder, M. Adelt, B. Richter, M. Naschitzki, M. Bäumer, and H.-J. Freund, Surf. Rev. Lett. **7**, 7 (2000).
- [105] T. Schroeder, A. Hammoudeh, M. Pykavy, N. Magg, M. Adelt, M. Bäumer, and H.-J. Freund, Solid-State Electronics **45**, 1471 (2001).

- [106] T. Schroeder, J. B. Giorgi, M. Bäumer, and H.-J. Freund, Phys. Rev. B **66**, 165422 (2002).
- [107] D. Ricci and G. Pacchioni, Phys. Rev. B **69**, 161307(R) (2004).
- [108] T. K. Todorova, M. Sierka, J. Sauer, S. Kaya, J. Weissenrieder, J.-L. Lu, H.-J. Gao, S. Shaikhutdinov, and H.-J. Freund, Phys. Rev. B **73**, 165414 (2006).
- [109] G.-M. Rignanese, A. D. Vita, J.-C. Charlier, X. Gonze, and R. Car, Phys. Rev. B **61**, 13250 (2000).
- [110] J. Bernhardt, J. Schardt, U. Starke, and K. Heinz, Appl. Phys. Lett. **74**, 1084 (1999).
- [111] G. Sposito, *The surface chemistry of soils*, Oxford Univ. Press, 1984.
- [112] F. Liebau, *Structural Chemistry of Silicates*, Springer, Berlin · Heidelberg, 1985.
- [113] P. Käckell, B. Wenzien, and F. Bechstedt, Phys. Rev. B **50**, 10761 (1994), the optimised bulk lattice constants are 3.032 Å (4H-SiC) and 3.033 Å (6H-SiC), respectively.
- [114] Y. Kim and D. Goodman, Langmuir **19**, 354 (2003).
- [115] A. Stierle, F. Renner, R. Streitl, H. Dosch, W. Drube, and B. C. Cowie, Science **303**, 1652 (2004).
- [116] S. Andersson, P. A. Brühweiler, A. Sandell, M. Frank, J. Libuda, A. Giertz, B. Brena, A. J. Maxwell, M. Bäumer, H.-J. Freund, and N. Mårtensson, Surf. Sci. **442**, L964 (1999).
- [117] G. Ceballos, Z. Song, J. I. Pascal, H.-P. Rust, H. Conrad, M. Bäumer, and H.-J. Freund, Chem. Phys. Lett. **359**, 41 (2002).
- [118] Y. Harada, S. Masuda, and H. Ozaki, Chem. Rev. **97**, 1897 (1997).
- [119] E. K. Chang, M. Rohlfing, and S. G. Louie, Phys. Rev. Lett. **85**, 2613 (2000).
- [120] L. Ramos, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **69**, 085102 (2004).
- [121] G. Pacchioni, L. Giordano, and M. Baistrocchi, Phys. Rev. Lett. **94**, 226104 (2005).

- 
- [122] D. Ceresoli, M. Bernasconi, S. Iarlori, M. Parinello, and E. Tosatti, *Phys. Rev. Lett.* **84**, 3887 (2000).
- [123] M. Milun, P. Pervan, and D. P. Woodruff, *Rep. Prog. Phys* **65**, 99 (2002), and references therein.
- [124] D. Yu, M. Scheffler, and M. Persson, *Phys. Rev. B* **74**, 113401 (2006).
- [125] X.-G. Wang, A. Chaka, and M. Scheffler, *Phys. Rev. Lett.* **84**, 3650 (2000).
- [126] G. Pacchioni, T. Minerva, and P. S. Bagus, *Surf. Sci.* **275**, 450 (1992).
- [127] R. Nada, A. C. Hess, and C. Pisani, *Surf. Sci.* **336**, 353 (1995).
- [128] A. D'Ercole, A. M. Ferrari, and C. Pisani, *J. Chem. Phys* **115**, 509 (2001).
- [129] C. Noguera, *Surf. Rev. Lett.* **8**, 121 (2001).
- [130] S. Wendt, M. Frerichs, T. Wei, M. Chen, V. Kempster, and D. Goodman, *Surf. Sci.* **565**, 107 (2004).
- [131] C. Schwennicke, J. Schimmelpfennig, and H. Pfnür, *Surf. Sci.* **293**, 57 (1993).
- [132] C. Lucas, G. C. L. Wong, C. S. Dower, F. J. Lamelas, and P. H. Fuoss, *Surf. Sci.* **286**, 46 (1993).
- [133] V. Zielasek, T. Hildebrandt, and M. Henzler, *Phys. Rev. B* **69**, 205313 (2004).
- [134] U. Barjenbruch, S. Fölsch, and M. Henzler, *Surf. Sci.* **211/212**, 749 (1989).
- [135] U. Malaske, C. Tegenkamp, M. Henzler, and H. Pfnür, *Surf. Sci.* **408**, 237 (1998).
- [136] V. Zielasek, T. Hildebrandt, and M. Henzler, *Phys. Rev. B* **62**, 2912 (2000).
- [137] C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. Lett.* **90**, 076803 (2003).
- [138] R. Browning, M. Sobolewski, and C. Helms, *Phys. Rev. B* **38**, 13407 (1988).

- [139] A. Pasquarello, M. Hybertsen, and R. Car, *Phys. Rev. B* **53**, 10942 (1996).
- [140] Y. Murata, K. Nagata, H. Fujimoto, T. Sakurai, M. Okada, and Y. Ebe, *J. Phys. Soc. Jap.* **70**, 793 (2001).
- [141] J. Vogt and H. Weiss, *Surf. Sci.* **491**, 155 (2001).
- [142] W. Hebenstreit, J. Redinger, Z. Horozova, M. Schmid, R. Podlucky, and P. Varga, *Surf. Sci.* **424**, L321 (1999).
- [143] W. Ernst, M. Eichmann, H. Pfnür, K.-L. Jonas, V. von Oeynhausen, and K. H. Meiwes-Broer, *Appl. Phys. Lett.* **80**, 1595 (2002).
- [144] F. E. Olsson, M. Persson, J. Repp, and G. Meyer, *Phys. Rev. B* **71**, 075419.
- [145] M. Kiguchi, H. Inoue, K. Saiki, T. Sasaki, Y. Iwasawa, and A. Koma, *Surf. Sci.* **522**, 84 (2003).
- [146] K. Nagata, C. Yamada, T. Takahashi, and Y. Murata, *J. Phys.: Condens. Matter* **15**, 8165 (2003).
- [147] J. C. Slater, *Adv. Quantum Chem.* **6**, 1 (1972).
- [148] G. Geneste, J. Morillo, and F. Finocchi, *J. Chem. Phys.* **122**, 174707 (2005).
- [149] D. Sanchez-Portal, E. Artacho, and J. M. Soler, *Solid State Commun.* **95**, 685 (1995).
- [150] J. C. Slater and W. Shockley, *Phys. Rev.* **50**, 705 (1936).
- [151] P. de Boer and R. A. de Groot, *Phys. Lett. A* **256**, 227 (1999).
- [152] P. de Boer and R. A. de Groot, *Eur. Phys. J. B* **4**, 25 (1998).
- [153] J. Tersoff and D. R. Hamann, *Phys. Rev. B* **31**, 805 (1985).
- [154] S. Baroni and R. Resta, *Phys. Rev. B* **33**, 7017 (1986).



## Danksagung

Zuallerst und besonders möchte ich mich bei Prof. Dr. Matthias Scheffler für die Überlassung dieses interessanten Themas und die Möglichkeit bedanken, in seiner Abteilung am Fritz-Haber-Institut zu promovieren. Darüber hinaus habe ich in zahlreichen Diskussionen von seiner großen Erfahrung und Weitsicht profitiert.

Ein ganz besonderer und herzlicher Dank gilt meinem Betreuer Dr. Patrick Rinke, ohne dessen stete Unterstützung in den großen und kleinen Problemen diese Doktorarbeit nicht möglich gewesen wäre. Dich als "Chef" zu haben war ein großer Glücksfall. Danke vor allem auch für die Geduld in den letzten Tagen vor der Abgabe...

Besonders danken möchte ich auch meinem Mitstreiter in der Vakuum-Problematik für wiederholte Schichtsysteme, Philipp Eggert, für die produktive und gute Zusammenarbeit. In diesem Zusammenhang wäre auch Arno Schindlmayr zu nennen, dem wir wesentliche theoretische Ansätze verdanken.

Meine wissenschaftliche Bildung während der Promotionszeit ist von verschiedenen Seiten wesentlich befördert worden, insbesondere durch die International Max-Planck Research School, das Nanoquanta Network of Excellence und die Hardy Groß'schen Gruppenseminare.

Ich habe mit vielen Kollegen am Fritz-Haber-Institut zahllose anregende Diskussionen über wissenschaftliche und weniger wissenschaftliche Themen geführt, die einen wesentlichen Reiz für diese Zeit ausmachten. Namentlich nennen möchte ich hier nur wenige, denen ich besonders viel verdanke, nämlich Patrick Rinke, Sixten Boeck, Philipp Eggert, Thomas Hammerschmidt, Martin Fuchs und Volker Blum. Gleichfalls danke ich meinen "office mates" Martin, Serdar, Jie und Ville für die schöne Zeit.

Mit meinen Mitdoktoranden aus der IMPRS habe ich viele Seminare und vor allem Kaffee-Pausen mit interessanten und ermutigenden Gesprächen verbracht. Besonderer Dank an Tanya Todorova und Sarp Kaya für die stets aktuellen Nachrichten von der moly-Front.

Dem SFHIngX-Entwickler-Team, allen voran Sixten Boeck, verdanke ich nicht nur das coolste DFT-plane wave-Programm, sondern auch fast alle Kenntnisse über numerisch effiziente Programmierung.

Bei Anne Kubin bedanke ich mich fürs Korrekturlesen. Ein Augenpaar mehr findet doch immer noch wieder Flüchtigkeitsfehler.

Det finns ett liv utanför institutet (fast jag nästan inte trodde på det på senaste tiden). Tack till svenskursen för å hålla mig vid livet.

Eine der größten Hürden für diese Doktorarbeit waren die immer wiederkehrenden Frustphasen, die ich ohne die moralische Unterstützung und den übergroßen Optimismus meiner beiden Familien, den Germännern und den

Freysoldts, wohl kaum überstanden hätte. Danke.

Meiner Frau Tonia habe ich zuguterletzt mehr zu danken als ich aufzählen vermag. Danke für deine Geduld, deine Unterstützung und alles andere.



## Abstract

Insulator surfaces and thin films play an important role in a variety of technological applications such as electronic devices or heterogeneous catalysts. Aiming at an atomistic understanding of the complex materials found in real applications, ultrathin epitaxial films are frequently studied as simplified model systems. Unlike their bulk and surface counterparts they are – when grown on metals or semiconductors – amenable to the arsenal of sophisticated surface science techniques currently available. In this thesis, first-principles calculations based on density-functional theory (DFT) and many-body perturbation theory in the  $GW$  approximation have been performed to clarify if these systems are indeed representative surface models or unique films instead. In other words how do ultrathin films differ from their macroscopic counterparts? The substrate-induced changes in a supported film’s atomic and electronic structure are investigated for a silica film on Mo(112) by comparing it to the closely related  $\alpha$ -quartz (0001) surface, and found to be significant. It is then demonstrated for three characteristic wide-gap oxides (silica, alumina, and hafnia) that films of 1–2 monolayer thickness generally show noticeable differences to thicker films.

The excited-state perspective is studied for NaCl on Ge(001) as a prototypical insulator/semiconductor interface by means of  $G_0W_0$  calculations. First, the reliability of the repeated-slab approach, the standard technique to describe films or surfaces with plane-wave methods, for  $G_0W_0$  is established for freestanding NaCl films by a thorough analysis of the relevant polarisation effects at different length scales. It is found that the inter-slab polarisation is non-negligible, but can easily be corrected for. The  $G_0W_0$  corrections for NaCl films differ from those of the bulk case because they include surface and interface polarisation effects for charged excited states that are absent from DFT-LDA. These dielectric response effects can be qualitatively reproduced by the classical theory of dielectric screening. For free-standing films, the main effect is a thickness-dependent change in the film’s band gap. For supported films, on the other hand, an image-potential-like  $z$ -dependence of the self-energy is observed. The implications of these findings for molecular adsorbates and metallic substrates are discussed.

## Zusammenfassung

In vielen technischen Anwendungen wie z. B. in der heterogenen Katalyse oder in elektronischen Schaltkreisen spielen isolierende Oxidfilme oder -oberflächen eine entscheidende Rolle. Darüber hinaus dienen Isolatorfilme in der Grundlagenforschung als Modelloberflächen, um experimentell mehr über fundamentale Eigenschaften oder Prozesse isolierender Oberflächen auf atomarer Ebene zu erfahren, da hinreichend dünne Filme auf leitenden Substraten (Metalle oder stark dotierte Halbleiter) im Gegensatz zu makroskopischen Proben die Verwendung hochauflösender Oberflächentechniken wie Rastertunnelmikroskopie oder Elektronenspektroskopie erlauben. Allerdings zeigen die kürzlich aufgeklärten Strukturen experimenteller Oxidfilmsysteme Filmstärken von nur 1–2 Monolagen und wecken Zweifel an der Vergleichbarkeit mit den Oberflächen der Volumenmaterialien. Um diese Fragen systematisch zu klären, wurden in der vorliegenden Arbeit *ab-initio*-Berechnungen an dünnen Filmen für drei charakteristische Oxide ( $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{HfO}_2$ ) im Rahmen der Dichtefunktionaltheorie in der lokalen Dichtenäherung (DFT-LDA) durchgeführt. Der Einfluss des Metallsubstrats wurde anhand von  $\text{Si}_2\text{O}_5/\text{Mo}(112)$ , des experimentell wichtigsten Films von Siliziumoxid, untersucht. Im Vergleich mit der strukturell verwandten Oberfläche von  $\alpha$ -Quarz zeigen sich deutliche Unterschiede in der atomaren und elektronischen Struktur der Oxidoberflächen. Hinsichtlich der Filmstärkenabhängigkeit ergeben sich für freistehende Filme aller drei Oxide deutliche Veränderungen, wenn eine kritische Dicke von 3–4 Monolagen unterschritten wird.

Der Vergleich mit experimentellen Einteilchen-Anregungsenergien erfordert Korrekturen zu den DFT-LDA-Bandstrukturen, die sich mit Vielteilchenstörungstheorie in der *GW*-Näherung berechnen lassen. Allerdings müssen im Superzellenansatz (dem Standardverfahren zur Beschreibung von Oberflächen und Filmen) langreichweitige Polarisierungseffekte berücksichtigt werden. In dieser Arbeit wurden diese Effekte und das sich daraus ergebende Konvergenzverhalten systematisch evaluiert und ein praktisches Verfahren gefunden, um *GW*-Resultate für das isolierte System zu extrapolieren. Dieses Verfahren wurde dann auf  $\text{NaCl}/\text{Ge}(001)$  als prototypisches Isolator-Halbleiter-System angewendet. Die  $G_0W_0$ -Korrekturen für Filme unterscheiden sich von denen des Volumenmaterials, da Oberflächen- und Grenzflächen-Polarisationseffekte auftreten, die in der DFT-LDA-Bandstruktur fehlen, sich aber mit dielektrischen Modellen physikalisch verstehen lassen. Bei freistehenden Filmen bewirkt die Oberflächenpolarisation eine filmstärkenabhängige Öffnung der Bandlücke. Bei aufgewachsenen Filmen beobachtet man dagegen eine bildpotential-artige  $z$ -Abhängigkeit der Selbstenergie. Die Auswirkungen solcher Effekte bei Adsorbaten und Metallsubstraten werden diskutiert.

## Veröffentlichungen

Teile dieser Arbeit wurden wie folgt veröffentlicht:

1. C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler: "Dielectric anisotropy in the GW space-time method", *Comp. Phys. Commun.* **176**, 1 (2007).
2. C. Freysoldt, P. Rinke, and M. Scheffler: "Ultrathin oxides: bulk-oxide-like model surfaces or unique films?", eingereicht bei *Phys. Rev. Lett.*

# Lebenslauf

## Persönliche Daten

Name: Christoph Freysoldt

Geburtsdatum: 19.02.1977

Familienstand: verheiratet

## Schulbildung

1983-1986	Römerschule Stuttgart
1986-1987	Albschule Stuttgart
1987-1996	Karls gymnasium Stuttgart
06/1996	Abitur (1,0)
09/1996-07/1997	Au-pair-Aufenthalt Paris

## Studium

10/1997-11/2002	Diplomstudium Chemie an der Universität Leipzig
08/2000-06/2001	Auslandsstudium in Uppsala/Schweden
11/2002	Diplom mit Auszeichnung (1,0)
seit 12/2002	Doktorarbeit am Fritz-Haber-Institut der Max-Planck-Gesellschaft
2003	Aufnahme in die International Max Planck Research School "Complex Surfaces in Materials Science"

## Auszeichnungen und Preise

07/1996	Goldmedaille bei der Internationalen Chemieolympiade
07/1996	Aufnahme in die Studienstiftung des Deutschen Volkes

Hiermit erkläre ich, die vorliegende Arbeit ohne weitere Hilfsmittel und Hilfen als die angegebenen selbständig angefertigt zu haben.

Berlin, den