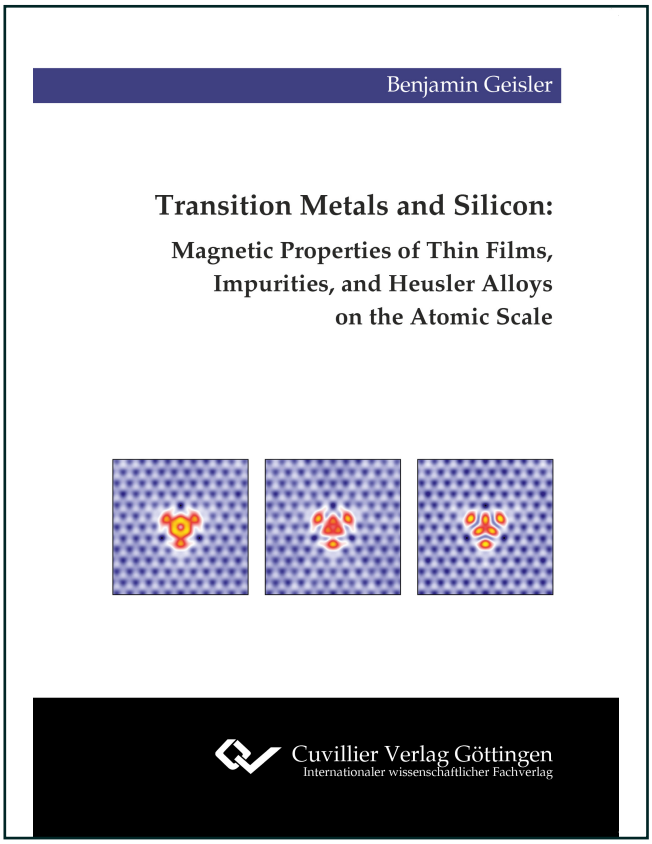
Benjamin Geisler (Autor)

Transition Metals and Silicon

Magnetic Properties of Thin Films, Impurities, and Heusler Alloys on the Atomic Scale



https://cuvillier.de/de/shop/publications/6921

Chapter 1

Introduction

Since the invention of the universal programmable computer in the 1940s, solid state physics has contributed tremendously to the development of computer hardware. Nevertheless, a quite unphysical separation has occurred: It was soon realized in practical computer engineering that electric and magnetic phenomena, which are known to depend on each other and are described in a combined theory, each have their own advantages and disadvantages. Electricity, on the one hand, is fast, but volatile. Magnetism, on the other hand, that arises in condensed matter as a macroscopic quantum effect, is slower, but persistent. Consequently, electric charges and capacitors are used nowadays for information processing and temporary data storage, while magnetism is used almost exclusively in reliable long-term storage devices.

The field of spintronics [1–3], which emerged some years ago, attempts to recombine electricity and magnetism, which both originate from electrons: While the former is a result of the electron charge, the latter is mostly a consequence of the electron spin. It should therefore be possible to exploit the electron spin as a carrier of information in addition to its charge. Fundamental and continuously investigated topics in the field of spintronics are the injection, transport, manipulation, storage, and detection of spin-polarized carriers [3–5]. The vision is to devise a new electronics, with new devices for information processing and data storage.

Metal-based spintronics devices already play an important role in our everyday life. The 2007 Nobel Prize in Physics was awarded to A. Fert and P. Grünberg for the discovery of the giant magnetoresistance effect, which is used in read heads of hard discs. Another device belonging to this field is the magnetoresistive random access memory (MRAM), which is based on the tunnel magnetoresistance (TMR) effect and already commercially available. In a MRAM module, the capacitive storage used in conventional dynamic random access memory (DRAM) is replaced by a form of magnetic storage that employs magnetic tunnel junctions (MTJs). A considerable advantage of MRAM is the persistence of the stored information, even if the energy supply of the device is removed. This is clearly not the case in DRAM modules, since they constantly need energy for refreshing the stored information [6]. Consequently, MRAM is more energy efficient. Moreover, it is cost effective to integrate, has an unlimited endurance, and provides fast (≈ 5 ns) random access [6,7]. Another suggested concept aims at overcoming the rigidity of the present hardware by creating logic devices, based on MTJs, that are reprogrammable at run-time (so-called “chameleon” processors [8]).

Semiconductor spintronics, on the other hand, offers a possible direction towards devices that perform logic, communication, and storage within the same materials technology [9]. A drawback of metal-based spintronics is that metals have no band gap; therefore, the carrier density cannot be as precisely controlled as in semiconductors (by doping, electrostatic gating, or band gap engineering). Moreover, the room-temperature spin relaxation times in nonmagnetic semiconductors are three orders of magnitude longer than in nonmagnetic metals [9]. Semiconductor spintronics devices compare favorably in speed and efficiency to electronic devices [10]. They could further reduce the power consumption by staying out of equilibrium for a long time (of the order of the spin relaxation time) instead of undergoing frequent equilibration (energy dissipation) [9]. Magnetic transistors could combine nonvolatility and reprogrammability with amplification [9]. Moreover, this technology could open a door towards quantum computation [9].

An interesting subtopic of semiconductor spintronics is Si spintronics [5], not least for technological and economical reasons: Si is an abundant material, the de-facto standard semiconductor, and our industry and technology is highly optimized for Si processing. In addition, the Si crystal structure has inversion symmetry, and its isotopes ^{28}Si (92 %) and ^{30}Si (3 %) have no nuclear spin [5]. Consequently, the spin relaxation in Si is dominated by the Elliott-Yafet mechanism (in contrast to the D'yakonov-Perel' mechanism, which plays an important role in III-V semiconductors like GaAs [11]), a combined action of momentum scattering, e.g., by phonons or impurities, and spin-orbit interaction, which is small in Si [5]. From an analysis of the Elliott and Yafet processes it has been shown recently that the phonon-induced spin relaxation time τ of the conduction electrons in bulk Si behaves as $\tau \sim T^{-3}$, reaching large values of 1 μs for $T = 60$ K and around 10 ns for $T = 300$ K [12]. Conventional doping reduces the spin relaxation time due to a more frequent impurity scattering [5]. Room-temperature spin injection, spin manipulation, and spin detection in Si has been demonstrated recently [13] as well as the nonlocal detection of a spin accumulation [14].

Rational materials design for future spintronics devices requires a thorough, detailed understanding of materials properties and their interdependence in various fields; for instance, (i) the dependence of structural properties on the growth protocol, (ii) the interplay of structural, electronic, and magnetic properties, (iii) the interactions between the different constituents of the material, e.g., between a semiconductor host and magnetic impurities embedded therein, and (iv) the dependence of transport and spincaloric properties on the electronic and magnetic structure. Modern experimental methods allow for an investigation of these points on the atomic scale. Therefore, contact can be made to state-of-the-art computational physics, which provides detailed insight on the same length scale. This enables one to analyze systems on a fundamental, quantum-mechanical, *ab initio* level, far beyond phenomenological approaches.

In this spirit, the present thesis addresses several topics related to magnetism on the atomic scale in the fields of semiconductor and metal-based spintronics, such as ferromagnetic thin films, dilute magnetic semiconductors (DMS), or MTJs with half-metallic electrodes. As a common theme of all topics, the semiconductor Si is combined in different ways with the abundant and "magnetic" 3*d* transition metals (TMs).

The present thesis is organized as follows: After this introduction (**Chapter 1**), the employed numerical methods are presented (**Chapter 2**), e.g., spin-polarized density functional theory (DFT) within the plane wave framework, ultrasoft pseudopotentials, semilocal and hybrid exchange-correlation functionals, on-site electronic correlation, the simulation of spin-

polarized scanning tunneling microscopy (SP STM) images from the DFT electronic structure in the spirit of the Tersoff-Hamann approximation, and ballistic quantum transport in the Landauer-Büttiker formalism. Other methods like *ab initio* thermodynamics, Monte Carlo simulations, or the Sivan-Imry approach to spincaloric properties are described later in this thesis. Subsequently, the results are discussed in three different chapters. Each of these chapters begins with a more profound introduction, which provides an integration of the results into the current state of research.

One possible route to fabricate a spin injector (or a spin detector) is by growing a ferromagnetic thin film of TMs on a semiconductor substrate (e.g., Si). The high reactivity of the Si surfaces leads to the immediate formation of TM silicides. **Chapter 3** treats several topics concerning different 3*d* TM silicides as bulk material and as film structure of different thickness grown epitaxially on the Si(111) surface. Structural, electronic, and magnetic properties of bulk 3*d* TM monosilicides, which largely crystallize in the complex B20 structure, are presented. In particular, the band gap, which is known to exist for FeSi, is shown to be a universal property of this class of materials. It is linked to the B20 crystal structure and separates two opposing groups of bands with different predominant TM 3*d* orbital character. MnSi and FeSi are also studied in the rock-salt and zinc-blende crystal structures. The validity of the present theoretical description of these materials, which exhibit nontrivial electronic correlations, and the applicability of *ab initio* thermodynamics is assessed.

Calculations for bulk 3*d* TM silicides under biaxial strain provide materials properties in the limit of thick (e.g., several hundred monolayers) epitaxial films on Si(111). For MnSi, a strain-induced volume expansion, increased magnetic moments and Curie temperature, and a phonon softening are found. FeSi makes a surprising transition from a nonmagnetic semiconductor to an almost half-metallic ferromagnet if matched epitaxially to Si(111), which means that it displays a finite density of states at the Fermi energy only in one spin channel. These findings make MnSi and FeSi especially interesting. On the other hand, it is explicitly demonstrated for MnSi/Si(111) heterostructures that the magnetic behavior of thin (e.g., just a few monolayers) 3*d* TM silicide films can deviate significantly from that of thicker films due to quantum confinement and/or interface/surface effects.

Subsequently, the growth mode, the atomic, and the magnetic structure of epitaxial MnSi thin films on Si(111) are analyzed in more detail. Interface properties are presented for MnSi/Si(111) and FeSi/Si(111). The atomic structure of the recently observed $\sqrt{3} \times \sqrt{3}$ and 3×3 surface reconstructions of MnSi films is revealed by combining experimental atomic-resolution STM imaging and first-principles simulations. A comparison of results for films grown by different techniques provides evidence that their internal stacking depends on the growth protocol. Since the magnetic properties of MnSi films are closely related to their atomic structure, the occurrence of different stacking sequences implies that strict control of the growth conditions is required to reproducibly fabricate MnSi films with specified properties. Finally, it is shown that the competing formation of Mn_5Si_3 can be excluded in the present experiments.

A second route towards new materials for spintronics is the doping of semiconductors or insulators with magnetic impurities. The systematic, reliable investigation of these DMS systems is a highly problematic and nontrivial task due to the many possible pitfalls. An example shall be given: Since high doping concentrations ($\sim 1\,\%$) of magnetic impurities are necessary, DMS samples can suffer from clustering and segregation of the impurities. This makes it hard (or impossible) for conventional experimental methods to isolate the different

types of magnetic interactions present in the sample (e.g., impurity-host interactions of isolated impurities vs. impurity-impurity interactions of impurity pairs or clusters; moreover, signals from segregated bulk phases of the impurity material). First-principles simulations, on the other hand, can suffer from the band gap problem or a deficient energy level ordering. **Chapter 4** introduces a new strategy to study experimentally the unbiased, *bulklike* impurity-host and impurity-impurity interactions of magnetic impurities in semiconductors and insulators on the atomic scale by using *surface* methods like SP STM in conjunction with passivated host surfaces. This strategy can lead to new insights on a fundamental, atomic level in the field of DMS.

The proof of principle is given by large-scale *ab initio* computer simulations for 3*d* TM impurities in Si, but the suggested approach is generalizable to other DMS systems like Co-doped ZnO. Different structural, electronic, and magnetic properties of interstitial and substitutional Cr, Mn, and Fe impurities in bulk Si and below the H/Si(111) surface are explained. Kinetic considerations concerning the injection and migration of subsurface impurities are made. It is demonstrated how SP STM in conjunction with passivated host surfaces can be used to image and measure the extension of impurity wave functions, the anisotropic spin polarization induced in the host electronic structure, and, *quantitatively*, the exchange coupling constants between neighboring impurities. This provides, for instance, a visual explanation why a ferromagnetic coupling is more difficult to achieve in Si than in GaAs. It is shown that semilocal and state-of-the-art hybrid exchange-correlation functionals lead to similar electronic and magnetic structures, with the exception of the substitutional Fe impurity. It is demonstrated how the latter can be used as a benchmark for the applicability of hybrid functionals in DMS simulations.

The new technique can also be applied to impurity δ layers and two-dimensional impurity clusters, which are studied afterwards. In particular, it is found that (111) δ layers of interstitial Cr impurities in Si exhibit a strong ferromagnetic interaction and are half-metallic. More generally, Cr-doped Si seems to be an interesting candidate material for a DMS. Moreover, it is shown how a ferromagnetic semiconductor could be constructed from two-dimensional Fe clusters embedded in Si.

The last part of the chapter is devoted to the experimentally motivated question whether or not it is possible to generate subsurface Fe impurities by deposition of organic molecules (e.g., iron phthalocyanine, FePc) on the H/Si(111) surface. It is concluded that the implantation is possible, but the necessary steps in the reaction are quite improbable due to the high binding energy of Fe in FePc. Gas phase calculations of phthalocyanine molecules with different TM centers employing different exchange-correlation functionals underline that these high binding energies are reasonable. It is shown that van-der-Waals corrections to semilocal DFT are very important for the correct description of the FePc adsorption on H/Si(111). Besides, silicon phthalocyanine is found to be an interesting example of d_0 magnetism in molecules.

Magnetic materials containing Si can be synthesized in many ways. Instead of diluting Si with TMs, one can also add Si to an alloy of TMs. An example is the Heusler alloy Co_2MnSi, which is a half-metallic ferromagnet. **Chapter 5** focuses on electronic transport and spincaloric properties of epitaxial MTJs with half-metallic Co_2MnSi Heusler electrodes and MgO tunneling barriers. Such heterostructures can be employed in a suggested "thermo-MRAM" module, which uses the magneto-Seebeck effect to read out the magnetic state of a MTJ without an applied voltage or a flowing charge current. Moreover, a Co_2MnSi/MgO/se-

miconductor setup in conjunction with an applied voltage or a temperature gradient can be used as spin injector that avoids the conductivity mismatch problem.

Given that the Co_2MnSi electrodes are ferromagnetic half-metals, the relative magnetization of the two electrodes has striking consequences for the transport properties of the MTJ. For the case of parallel electrode magnetization, electronic transport, spincaloric properties, and the dependence of both on the interface atomic structure are analyzed. In addition to conventionally obtained, approximate Seebeck coefficients, a more general route to spincaloric properties is presented that directly provides the response of the system (current or voltage) to arbitrary thermal gradients. It is demonstrated that the conventional Seebeck coefficient can be understood as first-order Taylor expansion coefficient of the voltage response. Thermal variations of the chemical potential in the leads and finite-bias effects can be readily included in this method. It is shown that a targeted growth control of the MTJs can be used to tailor their spincaloric properties (e.g., magnitude and sign of the thermally induced voltage).

In contrast, an antiparallel magnetization of the half-metallic electrodes should ideally prevent any electronic transport. Nevertheless, a parasitic current flows in room-temperature experiments that reduces the TMR ratio. It is shown by using a statistical approach to the electron-phonon interaction involving randomly distorted structures that the excitation of phonons has no relevant effect on the electronic structure of Co_2MnSi at room temperature and therefore cannot account for the strong decrease of the TMR ratio. On the other hand, it is argued that inelastic tunneling involving thermally activated interface magnons is among the most probable explanations for the observed behavior. First-principles calculations reveal that such interface magnons are much more easy to excite than bulk magnons.

Finally, **Chapter 6** summarizes the contents of this thesis and provides perspectives for future investigations.

Publications

Parts of this thesis have been published:

- B. Geisler, P. Kratzer, T. Suzuki, T. Lutz, G. Costantini, and K. Kern, *Growth mode and atomic structure of MnSi thin films on Si(111)*, Phys. Rev. B **86**, 115428 (2012), Ref. [15]
- B. Geisler and P. Kratzer, *Strain stabilization and thickness dependence of magnetism in epitaxial transition metal monosilicide thin films on Si(111)*, Phys. Rev. B **88**, 115433 (2013), Ref. [16]
- T. Suzuki, T. Lutz, B. Geisler, P. Kratzer, K. Kern, and G. Costantini, *Surface morphology of MnSi thin films grown on Si(111)*, Surf. Sci. **617**, 106 (2013), Ref. [17]

Some publications are in preparation:

- B. Geisler and P. Kratzer, *Atomic-scale detection of magnetic impurity interactions in bulk semiconductors*, submitted to Phys. Rev.
- B. Geisler and P. Kratzer, *Spincaloric properties of epitaxial $Co_2MnSi/MgO/Co_2MnSi$ magnetic tunnel junctions*
- B. Geisler and P. Kratzer, *Cr-doped Si as novel dilute magnetic semiconductor*
- B. Geisler and P. Kratzer, *Physi- and chemisorption of organic molecules on H/Si*(111)

Further publications that are related to this thesis:

- B. Geisler, P. Kratzer, and V. Popescu, *Interplay of growth mode and thermally induced spin accumulation in epitaxial Al/Co_2TiSi/Al and Al/Co_2TiGe/Al contacts*, Phys. Rev. B **89**, 184422 (2014), Ref. [18]
- D. Comtesse, B. Geisler, P. Entel, P. Kratzer, and L. Szunyogh, *First-principles study of spin-dependent thermoelectric properties of half-metallic Heusler thin films between platinum leads*, Phys. Rev. B **89**, 094410 (2014), Ref. [19]

Chapter 2

Methodology

Contents

2.1 The electronic structure

Almost every result that will be presented in this thesis is based on the electronic structure of a certain physical system. This system can be an infinitely extended solid, an isolated impurity in a host material, a molecule, or just a single atom. All relevant physical properties of this system are determined, or significantly influenced, by its electronic structure: The quantum-mechanical many-particle behavior of electrons in the potential of the nuclei.

2.1.1 The many-particle problem in quantum mechanics

On the atomic scale, condensed matter physics is governed by the Schrödinger equation,

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\left|\psi_t\right\rangle = \hat{H}\left|\psi_t\right\rangle, \tag{2.1}$$

where $\hat{H}$ is the Hamilton operator ("Hamiltonian") of the system under consideration and the vector $|\psi_t\rangle$ denotes the quantum-mechanical state the system is in at time t. Solving Eq. (2.1) is equivalent to finding the stationary solutions and their energies, that is, eigenstates (or, similarly, eigenvectors) $|\psi_i\rangle$ and eigenenergies E_i of $\hat{H}$,

$$\hat{H}\left|\psi_i\right\rangle = E_i\left|\psi_i\right\rangle,$$

since the time evolution of the quantum system can be expressed (for time-independent Hamiltonians) by decomposing an arbitrary initial state $|\psi_0\rangle$ into the eigenbasis of $\hat{H}$:

$$\left|\psi_t\right\rangle = \sum_j \exp\left\{-\frac{\mathrm{i}}{\hbar}E_j t\right\}\left\langle\psi_j|\psi_0\right\rangle\left|\psi_j\right\rangle.$$

While a molecule consists of only few atoms, solids usually contain about 10^{23} atoms per cm^3. Each atom consists of a nucleus and several electrons. The full Hamiltonian of such a system, neglecting relativistic effects, external magnetic fields, and quantum electrodynamics, reads:

$$\hat{H} = \hat{T}_\mathrm{n} + \hat{T}_\mathrm{e} + \hat{V}_\mathrm{n\text{-}n} + \hat{V}_\mathrm{e\text{-}e} + \hat{V}_\mathrm{e\text{-}n},$$

with kinetic energy operators (in position representation) for nuclei (n) and electrons (e),

$$\begin{aligned}
\hat{T}_\mathrm{n} &= \sum_I^{N_\mathrm{n}} \frac{\hat{P}_I^2}{2M_I} = -\sum_I^{N_\mathrm{n}} \frac{\hbar^2}{2M_I}\nabla^2_{\vec{R}_I}, \\
\hat{T}_\mathrm{e} &= \sum_i^{N_\mathrm{e}} \frac{\hat{p}_i^2}{2m} = -\sum_i^{N_\mathrm{e}} \frac{\hbar^2}{2m}\nabla^2_{\vec{r}_i},
\end{aligned}$$

and Coulomb interaction potential operators,

$$\begin{aligned}
\hat{V}_\mathrm{n\text{-}n} &= +\frac{1}{2}\sum_{I\neq J}^{N_\mathrm{n}} \frac{e^2}{4\pi\varepsilon_0}\frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|}, \\
\hat{V}_\mathrm{e\text{-}e} &= +\frac{1}{2}\sum_{i\neq j}^{N_\mathrm{e}} \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\vec{r}_i - \vec{r}_j|}, \\
\hat{V}_\mathrm{e\text{-}n} &= -\sum_{i,I}^{N_\mathrm{e},N_\mathrm{n}} \frac{e^2}{4\pi\varepsilon_0}\frac{Z_I}{|\vec{r}_i - \vec{R}_I|},
\end{aligned}$$

where the factor $1/2$ eliminates double counting. Relativistic effects (like mass modifications or the electron spin) can be included later. The full many-particle wave functions Φ_i depend on the coordinates of all electrons and nuclei and are solutions of

$$\hat{H}\,\Phi_i(\vec{r}_1,\ldots,\vec{r}_{N_\mathrm{e}};\,\vec{R}_1,\ldots,\vec{R}_{N_\mathrm{n}}) = E_i\,\Phi_i(\vec{r}_1,\ldots,\vec{r}_{N_\mathrm{e}};\,\vec{R}_1,\ldots,\vec{R}_{N_\mathrm{n}}). \tag{2.2}$$

It is needless to say that, in general, such a system can be solved neither algebraically nor numerically. Even simple many-particle systems like a He atom or the H_2 molecule cannot be described exactly (although very good approximate results can be achieved here). Therefore, several well-chosen approximations are necessary that will be discussed in this chapter.

2.1.2 The adiabatic approximation

First of all, the motion of the nuclei can be decoupled from the motion of the electrons due to their highly different masses ($m/M \approx 10^{-3}$). This procedure is known as the "adiabatic approximation" (or Born-Oppenheimer approximation [20]). Generally, any many-particle wave function Φ can be decomposed according to

$$\Phi(\{\vec{r}_i,\,\vec{R}_I\}) = \sum_\nu \Psi_\nu(\{\vec{r}_i\}:\{\vec{R}_I\})\;\Xi_\nu(\{\vec{R}_I\}), \tag{2.3}$$

with electronic wave functions $\Psi_\nu(\{\vec{r}_i\}:\{\vec{R}_I\})$ and nucleonic wave functions $\Xi_\nu(\{\vec{R}_I\})$. Considering m/M as a small quantity, the kinetic energy of the nuclei $\hat{T}_\mathrm{n}$ can be regarded as a perturbation to the electronic part of the Hamiltonian $\hat{H}_\mathrm{e}$:

$$\hat{H} = \hat{H}_\mathrm{e} + \hat{T}_\mathrm{n}\;, \qquad \hat{H}_\mathrm{e} = \hat{T}_\mathrm{e} + \hat{V}_\text{n-n} + \hat{V}_\text{e-e} + \hat{V}_\text{e-n}\;.$$

Note that $\hat{H}_\mathrm{e}$ contains no differential operators with respect to the nucleonic positions $\vec{R}_I$. Inserting the wave function $\Phi(\{\vec{r}_i,\,\vec{R}_I\})$ as given in Eq. (2.3) into Eq. (2.2) and using the electronic part of the equation,

$$\hat{H}_\mathrm{e}\;\Psi_\nu(\{\vec{r}_i\}:\{\vec{R}_I\}) = \epsilon_\nu(\{\vec{R}_I\})\;\Psi_\nu(\{\vec{r}_i\}:\{\vec{R}_I\}), \tag{2.4}$$

leads, after multiplication with $\Psi^*_{\nu'}(\{\vec{r}_i\})$ and integration over the electron coordinates $\vec{r}_i$, to the still exact expression

$$(\hat{T}_\mathrm{n} + \epsilon_\nu(\{\vec{R}_I\}))\;\Xi_\nu(\{\vec{R}_I\}) + \sum_{\nu'} A_{\nu,\nu'}\;\Xi_{\nu'}(\{\vec{R}_I\}) = E\;\Xi_\nu(\{\vec{R}_I\}).$$

The adiabatic approximation [21, 22] corresponding to the limit $m/M \to 0$ consists of neglecting the transition matrix elements $A_{\nu,\nu'}$ (which contain the electronic wave functions). It leads to a complete separation of electronic and nucleonic dynamics. The equation for the nucleonic wave functions then reads:

$$(\hat{T}_\mathrm{n} + \epsilon_\nu(\{\vec{R}_I\}))\;\Xi_\nu(\{\vec{R}_I\}) = E\;\Xi_\nu(\{\vec{R}_I\}). \tag{2.5}$$

Thus, the problem is separated into two Schrödinger-like equations (2.4) and (2.5). The electronic equation (2.4) depends only parametrically on the positions of the nuclei. Hence, one can study the interacting electrons in a fixed, "external" potential generated by the nuclei. The central idea behind the adiabatic approximation is that the electron system responds instantaneously to changes of the nucleonic positions due to the mass difference,[1]

[1] Typical time scales for the electrons (nuclei) are femtoseconds (picoseconds).

and that there are no excitations of the electronic system (transitions $\nu \to \nu'$) due to nucleonic dynamics, so that the electrons remain in their (ground) state [22].

In most cases, speaking of "nucleonic wave functions" is exaggerated, since normally no Schrödinger-like equation, e.g., Eq. (2.5), is solved for the nuclei. Instead, they move on trajectories following classical dynamics, acted upon by forces calculated from the quantum-mechanical electron system (e.g., Hellmann-Feynman forces [23], which are used in this thesis for optimizations, i.e., relaxations, of the different atomic structures). For heavy atoms, this approximation is very good. Although nothing else than this is done in this thesis, it should be mentioned here that there exist other approaches, especially in modern research. For instance, *ab initio* path-integral molecular dynamics can be done, in which both the electrons and the nuclei are treated as quantum particles [24, 25].

2.1.3 Bloch's theorem

The complexity of a quantum many-particle problem is a consequence of the Coulomb interaction between the particles. However, even if the particles are assumed to be "noninteracting" (despite, e.g., some effective, mean-field-like interaction or the inclusion of the Pauli exclusion principle for electrons) and only the electronic parts of the wave functions are considered (for fixed nuclei), the question arises how a solid bulk system of 10^{23} atoms can be described efficiently. The key observation is that the potential generated by the nuclei is invariant under certain translations in a bulk crystal. Therefore, the whole (ideal and infinite) crystal can be constructed from a single unit cell by using translational operations. Bloch's theorem, which is well known in mathematics as Floquet's theorem, states that eigenfunctions of the electronic Hamiltonian $\hat{H}$ of the crystal can be chosen such as to possess the same translational invariance, modulated by a phase factor.[2] It is of fundamental importance in the field of electronic structure theory of solids.

Consider a periodic lattice in three dimensions which is described by real space lattice vectors $\vec{R} = n_1\vec{a_1} + n_2\vec{a_2} + n_3\vec{a_3}$. Since a translation operator $\hat{T}_{\vec{R}}$, which is defined by

$$\hat{T}_{\vec{R}}\, f(\vec{r}) = f(\vec{r} - \vec{R})$$

for arbitrary functions f, commutes with an electronic, single-particle Hamiltonian[3] in the case of a periodic potential,

$$\hat{H} = -\frac{\hbar^2\nabla^2}{2m} + V(\vec{r}), \qquad V(\vec{r} + \vec{R}) = V(\vec{r}) \quad \forall\vec{R},$$

the eigenstates of $\hat{H}$ can be chosen such that they are simultaneously eigenstates of all possible translation operators $\hat{T}_{\vec{R}}$. The translation operators are unitary; thus, they have complex eigenvalues with unit norm that can consequently be written as $\mathrm{e}^{-\mathrm{i}\vec{k}\vec{R}}$, which introduces the real vector $\vec{k}$. This can equivalently be deduced from their group properties. The eigenvalue equation,

$$\hat{T}_{-\vec{R}}\, \psi_{n\vec{k}}(\vec{r}) = \psi_{n\vec{k}}(\vec{r} + \vec{R}) = \mathrm{e}^{\mathrm{i}\vec{k}\vec{R}}\, \psi_{n\vec{k}}(\vec{r}),$$

turns out to be Bloch's theorem already. It is common to make a Bloch-like wave function *ansatz*,

$$\psi_{n\vec{k}}(\vec{r}) = \mathrm{e}^{\mathrm{i}\vec{k}\vec{r}} u_{n\vec{k}}(\vec{r}) \quad \text{with} \quad u_{n\vec{k}}(\vec{r} + \vec{R}) = u_{n\vec{k}}(\vec{r}), \tag{2.6}$$

[2] Note that this does not mean that *every* wave function of the crystal has translational invariance.

[3] We will be using single-particle equations and wave functions later in the Kohn-Sham context, e.g., Eq. (2.10).

since

$$\psi_{n\vec{k}}(\vec{r}+\vec{R}) = \mathrm{e}^{\mathrm{i}\vec{k}(\vec{r}+\vec{R})} u_{n\vec{k}}(\vec{r}+\vec{R}) = \mathrm{e}^{\mathrm{i}\vec{k}\vec{R}}\,\mathrm{e}^{\mathrm{i}\vec{k}\vec{r}} u_{n\vec{k}}(\vec{r}) = \mathrm{e}^{\mathrm{i}\vec{k}\vec{R}} \psi_{n\vec{k}}(\vec{r}).$$

If we insert the Bloch-like wave function *ansatz* into a Schrödinger-like equation,

$$\hat{H}\psi_{n\vec{k}}(\vec{r}) = \left[-\frac{\hbar^2\nabla^2}{2m} + V(\vec{r})\right] \mathrm{e}^{\mathrm{i}\vec{k}\vec{r}} u_{n\vec{k}}(\vec{r}) = \varepsilon_{n\vec{k}}\,\mathrm{e}^{\mathrm{i}\vec{k}\vec{r}} u_{n\vec{k}}(\vec{r}) = \varepsilon_{n\vec{k}}\,\psi_{n\vec{k}}(\vec{r}),$$

apply the operator in brackets and cancel the phase, we end up with the following partial differential equation for the $u_{n\vec{k}}$:

$$\left[\frac{\hbar^2}{2m}(\vec{k} - \mathrm{i}\nabla)^2 + V(\vec{r})\right] u_{n\vec{k}}(\vec{r}) = \varepsilon_{n\vec{k}}\, u_{n\vec{k}}(\vec{r}). \tag{2.7}$$

As we see, $\vec{k}$ enters the equation as a parameter; thus, we can solve it separately for each $\vec{k}$ we are interested in. Since the $u_{n\vec{k}}$ have to fulfill the lattice periodicity constraint, solving Eq. (2.7) becomes a boundary value problem on one single unit cell spanned by the vectors $\vec{a}_i$. Therefore, the energies $\varepsilon_{n\vec{k}}$ will be discrete. The index n labels the different solutions for a given $\vec{k}$ and is usually referred to as band index.

It is possible to restrict $\vec{k}$ to the first Brillouin zone. Since the reciprocal lattice vectors $\vec{G} = m_1\vec{b_1} + m_2\vec{b_2} + m_3\vec{b_3}$ fulfill (or can be defined via) $\mathrm{e}^{\mathrm{i}\vec{R}\vec{G}} = 1$, any $\vec{k}$ can be folded back to the first Brillouin zone by using an appropriate $\vec{G}$ without inducing any changes. This means that the problem of solving the Schrödinger equation for an infinite crystal reduces to the study of a single unit cell for different $\vec{k}$. As a consequence of the periodic boundary conditions, the number of $\vec{k}$ points in the first Brillouin zone is identical to the number of unit cells in the crystal volume that is considered. In the limit of an infinite system, the eigenvalues $\varepsilon_{n\vec{k}}$ become quasi-continuous functions of $\vec{k}$ and form bands.

2.2 Density functional theory

In the previous section it was shown how large periodic crystals can be described efficiently using Bloch's theorem, which states that the electronic structure of the infinite solid can be derived from calculations of single unit cells containing a small atomic basis. Single atoms and medium-sized molecules like porphyrins or phthalocyanines are not periodic, but contain a comparable number of electrons which move in an external potential given by the fixed nuclei. Unfortunately, already these small quantum systems are far too complicated to be solved directly. A possible alternative is the extremely successful density functional theory (DFT), for which W. Kohn received the 1998 Nobel Prize in Chemistry. Even though DFT was introduced nearly fifty years ago, there is still ongoing research concerning fundamental aspects. Its applications are extremely widespread. DFT is the quantum-theoretical method of choice for systems that are too large to be treated by the more exact quantum-chemical methods like configuration interaction or coupled cluster (small molecules), but still small enough to be in range for modern high-performance computers. Nowadays, systems containing around 1000 atoms can be handled.[4] Very large biomolecules or polymers

[4] See, for example, Refs. [26,27]. Note that actually the number of *electrons*, not the number of atoms, is the decisive factor. Thus, the feasibility of a system depends on the involved elements, the basis set (localized vs. delocalized), whether an all-electron or a pseudopotential approach is used, the necessary quality of the Brillouin zone sampling (metallic/insulating/isolated system), etc.

with several thousands of atoms are beyond DFT and can only be described by empirical methods like classical molecular dynamics employing so-called force fields.

Density functional theory is closely related to the Thomas-Fermi method which was already proposed in 1927 and employs the charge density $n(\vec{r})$ as its basic variable. However, it is more sophisticated in its mathematical foundation and the treatment of kinetic energy, exchange, and correlation of the interacting particles.

2.2.1 The theorems of Hohenberg and Kohn

Density functional theory is based on two mathematical theorems published by Hohenberg and Kohn in 1964 [28]. They can be applied to any system of N interacting particles in an external potential $V_{\text{ext}}(\vec{r})$, thus also (and especially) to interacting electrons in a potential given by the nuclei.

1. The external potential $V_{\text{ext}}(\vec{r})$ is, except for a constant shift, a unique functional of the ground state density $n_0(\vec{r}) = \langle\Psi_0|\sum_{i=1}^{N}\delta(\vec{r}-\vec{r}_i)|\Psi_0\rangle$. Since the external potential determines the Hamiltonian $\hat{H}$ completely, all properties (observables) of the system follow from $n_0(\vec{r})$.

2. The ground state density $n_0(\vec{r})$ minimizes the total energy functional

$$E[n] = F_{\text{HK}}[n] + E_{\text{ext}}[n] = F_{\text{HK}}[n] + \int \mathrm{d}^3r\; n(\vec{r})\, V_{\text{ext}}(\vec{r}), \tag{2.8}$$

   where $F_{\text{HK}}[n] = T[n] + E_{\text{e-e}}[n]$ is a *universal* functional (independent of any external potential) as defined by Hohenberg and Kohn [28]. $T[n]$, $E_{\text{e-e}}[n]$, and $E_{\text{ext}}[n]$ account for the kinetic energy, the electron-electron potential energy, and the electron-nucleus potential energy, respectively.

The astonishingly simple proof normally assumes a system with nondegenerate ground state, but can be extended to the degenerate case [22, 29]. The alternative, more general foundation of DFT proposed by Levy and Lieb [30, 31] already comprises this aspect. It is important to note here that these theorems found DFT as an *exact* many-particle theory without approximations, even though the coordinate space dimension is reduced from $3N$ to 3 for N particles in three dimensions.

In summary, the theorems of Hohenberg and Kohn imply that if we have found the global minimum n_0 of $E[n]$ by using, for example, the variational principle, we know the exact ground state energy $E[n_0] = \langle\Psi_0|\hat{H}|\Psi_0\rangle$ and all other properties we are interested in of the interacting many-body system.

2.2.2 The Kohn-Sham scheme

The next step, which turned DFT into the successful technique in electronic structure theory it is nowadays, was made by Kohn and Sham (KS) in 1965 [32]. They replaced the many-body problem of interacting electrons and nuclei by an auxiliary single-electron system that leads to the same ground state density $n_0(\vec{r})$ as the real system.

The Kohn-Sham equations

The task is to minimize the energy functional (2.8). From occupied, fermionic single-particle wave functions $\psi_i(\vec{r})$ building up a Slater determinant Ψ the density can be derived as

$$n(\vec{r}) = \sum_{i=1}^{N} \langle \Psi | \delta(\vec{r} - \vec{r}_i) | \Psi \rangle = \sum_{i=1}^{N} |\psi_i(\vec{r})|^2.$$

The kinetic energy of such a system is, in Hartree units,[5]

$$T = \langle \Psi | \left[-\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 \right] | \Psi \rangle = \sum_{i=1}^{N} \int \mathrm{d}^3 r \; \psi_i^*(\vec{r}) \left[-\frac{1}{2} \nabla^2 \right] \psi_i(\vec{r}) = \frac{1}{2} \sum_{i=1}^{N} \int \mathrm{d}^3 r \left| \vec{\nabla} \psi_i(\vec{r}) \right|^2,$$

where in the last step integration by parts has been used, omitting the surface term. We can also calculate a Hartree term which arises from the "classical" Coulomb interaction of an electron density $n(\vec{r})$:

$$E_{\mathrm{H}} = \frac{1}{2} \int \mathrm{d}^3 r_1 \int \mathrm{d}^3 r_2 \, \frac{n(\vec{r_1}) \, n(\vec{r_2})}{|\vec{r_1} - \vec{r_2}|}.$$

The universal part of the energy functional (2.8) can now been written as

$$F_{\mathrm{HK}}[n] = T + E_{\mathrm{H}}[n] + E_{\mathrm{xc}}[n],$$

which defines formally the so-called exchange-correlation functional, one of the most important components in DFT:

$$E_{\mathrm{xc}}[n] = F_{\mathrm{HK}}[n] - T - E_{\mathrm{H}}[n] = (T[n] - T) + (E_{\text{e-e}}[n] - E_{\mathrm{H}}[n]).$$

It contains the difference between the original kinetic energy functional and the single-particle kinetic energy defined above, which is presumably small [29], and the nonclassical part of the electron-electron interaction. The total energy functional now reads:

$$E[n] = T + E_{\mathrm{H}}[n] + E_{\mathrm{xc}}[n] + E_{\mathrm{ext}}[n]. \tag{2.9}$$

By minimizing this functional with respect to the density n, the Kohn-Sham equations can be derived. The constraint $\int \mathrm{d}^3 r \; \psi_i^*(\vec{r}) \psi_j(\vec{r}) = \delta_{ij}$ of orthonormal single-particle wave functions has to be included via Lagrange multipliers ε_{ij}:

$$\Omega[n] = E[n] - \sum_{i,j=1}^{N} \varepsilon_{ij} \left(\int \mathrm{d}^3 r \; \psi_i^*(\vec{r}) \psi_j(\vec{r}) - \delta_{ij} \right).$$

The variational equation (minimization condition) leads to

$$\begin{aligned} 0 \quad \stackrel{!}{=} \quad \frac{\delta \Omega[n]}{\delta \psi_i^*(\vec{r})} &= \frac{\delta T}{\delta \psi_i^*(\vec{r})} + \left[\frac{\delta E_{\mathrm{H}}}{\delta n(\vec{r})} + \frac{\delta E_{\mathrm{xc}}}{\delta n(\vec{r})} + \frac{\delta E_{\mathrm{ext}}}{\delta n(\vec{r})} \right] \frac{\delta n(\vec{r})}{\delta \psi_i^*(\vec{r})} - \sum_{j=1}^{N} \varepsilon_{ij} \psi_j(\vec{r}) \\ &= -\frac{1}{2} \nabla^2 \psi_i(\vec{r}) + \left[V_{\mathrm{H}} + V_{\mathrm{xc}} + V_{\mathrm{ext}} \right] \psi_i(\vec{r}) - \sum_{j=1}^{N} \varepsilon_{ij} \psi_j(\vec{r}), \end{aligned}$$

[5] Hartree units: $e = \hbar = m = (4\pi\varepsilon_0)^{-1} = 1$ vs. Rydberg units: $e^2/2 = \hbar = 2m = 1$

where the chain rule has been used. Unitary transformation of the ψ_i [29] leads to the well-known form of the Kohn-Sham equations,

$$\left[-\frac{1}{2}\nabla^2 + V_{\mathrm{KS}}(\vec{r})\right]\psi_i(\vec{r}) = \varepsilon_i\psi_i(\vec{r}), \tag{2.10}$$

where the effective Kohn-Sham potential is given by

$$V_{\mathrm{KS}} = V_{\mathrm{H}} + V_{\mathrm{xc}} + V_{\mathrm{ext}}. \tag{2.11}$$

These eigenvalue equations are partial differential equations of the Schrödinger type and are much easier to handle than the original many-particle system. They describe the movement of a single electron in an effective potential generated by the other electrons (through their charge density) and the nuclei. This is the common mean-field interpretation of the Kohn-Sham equations. In the case of periodic boundary conditions, Bloch's theorem can be applied, which leads to equations like Eq. (2.7).

Solving the Kohn-Sham equations: The self-consistency scheme

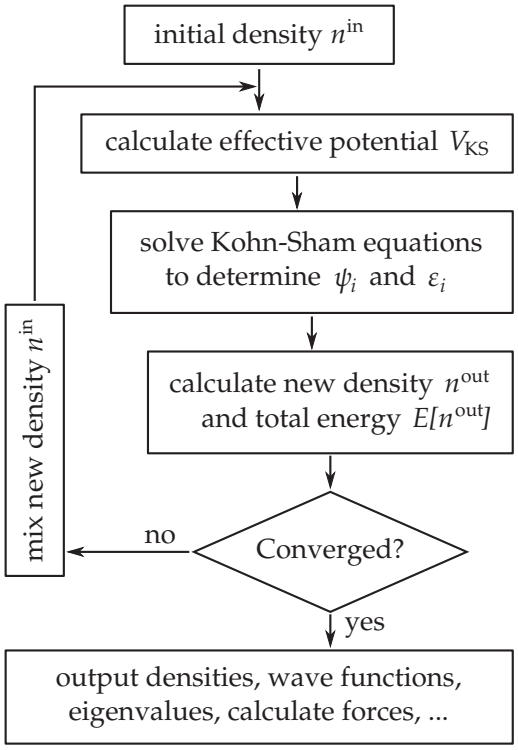


Figure 2.1 – DFT-SCF flowchart for solving the Kohn-Sham equations.

The effective Kohn-Sham potential V_{KS} depends via the Hartree potential,

$$V_{\mathrm{H}}(\vec{r}) = \frac{\delta E_{\mathrm{H}}}{\delta n(\vec{r})} = \int \mathrm{d}^3r' \, \frac{n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \, ,$$

on the density n and thus on the wave functions ψ_i, which again depend on V_{KS}. Thus, the Kohn-Sham equations (2.10) have to be solved iteratively, as outlined in Fig. 2.1.

Starting from an initial density n_1^{in} (calculated, for example, from superimposed atomic orbitals), the ψ_i are determined as solutions of Eqs. (2.10). From these wave functions a new density n_1^{out} is derived. By using some mixing algorithm, a density n_2^{in} is generated from n_1^{out} and (several) old densities, and Eqs. (2.10) are solved again. This *self-consistent field* (SCF) loop is repeated until the density n (and the corresponding energy $E[n]$) is converged.[6]

Converging the density is, in principle, a chapter of its own. Several mixing algorithms exist, like simple linear mixing,

$$n_{i+1}^{\mathrm{in}} = \beta n_i^{\mathrm{out}} + (1-\beta) n_i^{\mathrm{in}},$$

or the more elaborate (and in most cases faster converging) Broyden mixing, a Krylov-like method that uses data of several preceding densities and will be employed predominantly in

[6] "Converged" means that the following density changes are only small; for instance, the norm of the difference of two subsequent densities is below a chosen threshold.

this thesis. Linear mixing with $\beta = 1$ corresponds to the naive case where one only uses the new density, a strategy that mostly fails. Other common schemes are Pulay mixing, which works similarly to the Broyden method, and Kerker mixing.

Spin-polarized systems

For a spin-polarized system with collinear magnetic moments, the effective Kohn-Sham potential can be written as

$$\begin{aligned} V_{\mathrm{KS}}^{\sigma}(\vec{r}) &= V_{\mathrm{ext}}(\vec{r}) + V_{\mathrm{H}}(\vec{r}) + V_{\mathrm{xc}}^{\sigma}(\vec{r}) \\ &= V_{\mathrm{ext}}(\vec{r}) + \frac{\delta E_{\mathrm{H}}}{\delta n^{\sigma}(\vec{r})} + \frac{\delta E_{\mathrm{xc}}}{\delta n^{\sigma}(\vec{r})}, \end{aligned}$$

where $\sigma \in \{\uparrow, \downarrow\}$. The generalization of density and kinetic energy expressions is very simple, since just an additional summation over the two spin states σ has to be performed. In addition, a spin density can be defined:

$$m(\vec{r}) = n^{\uparrow}(\vec{r}) - n^{\downarrow}(\vec{r}).$$

The Hartree potential,

$$V_{\mathrm{H}}(\vec{r}) = \frac{\delta E_{\mathrm{H}}}{\delta n^{\sigma}(\vec{r})} = \int \mathrm{d}^3 r' \, \frac{n^{\uparrow}(\vec{r}\,') + n^{\downarrow}(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} = \int \mathrm{d}^3 r' \, \frac{n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} \qquad \forall \sigma,$$

is spin-independent. The extension of Eq. (2.7) to the spin-polarized, collinear case is straightforward:

$$\left[\frac{1}{2} (\vec{k} - \mathrm{i}\nabla)^2 + V_{\mathrm{KS}}^{\sigma}(\vec{r}) \right] u_{n\vec{k}}^{\sigma}(\vec{r}) = \varepsilon_{n\vec{k}}^{\sigma} \, u_{n\vec{k}}^{\sigma}(\vec{r}). \tag{2.12}$$

Here one simply has two sets of effective potentials V_{KS}^{σ}, energy eigenvalues $\varepsilon_{n\vec{k}}^{\sigma}$, and wave functions $u_{n\vec{k}}^{\sigma}$. The spin index σ in the latter two quantities is often absorbed in n or in $\vec{k}$ by a doubling of the bands or the $\vec{k}$ points, respectively.

Interpretation of the Kohn-Sham eigenvalues

It is tempting to interpret the Kohn-Sham eigenvalues ε_i as energy levels of the studied system. In fact, this is actually the common procedure. However, care has to be taken. First of all, they arise only as auxiliary quantities in the Kohn-Sham scheme, and at first sight there is no obvious reason why any connection to physical observables should exist. However, by using the adiabatic connection concept it is possible to show that differences of Kohn-Sham eigenvalues can be interpreted as well-defined "zeroth-order" approximations to excitation energies in the interacting system [33], which means that they can be compared, for example, to results from angle-resolved photoemission spectroscopy (ARPES) experiments. One exception that allows for a rigorous statement is the highest occupied energy eigenvalue in a *finite* system, which can be shown to be equal to minus the ionization energy [22, 34], a statement that is similar to Koopmans' theorem in Hartree-Fock theory. Problematic are the facts that (i) the electron-electron interaction is in general not small, which weakens the perturbative statement above, and (ii) the exact exchange-correlation functional is not known (see below).