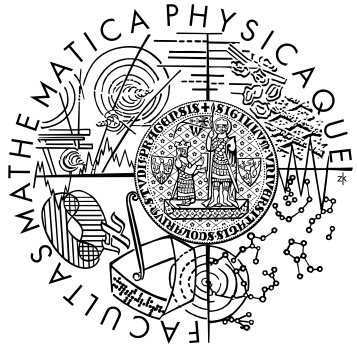
Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS



Rudolf Sýkora

Relativistic Theory of Electron Transport in Magnetic Layers

Department of Condensed Matter Physics

Supervisor: doc. RNDr. Ilja Turek, DrSc.

Study programme: Physics
Specialization: Theoretical Physics, Astronomy and Astrophysics

Prague 2012

Acknowledgements

The path which ended up in writing this thesis has been far from straight. Chronologically, my first thanks go to Ing. Jiří Hošek, CSc., who directed me to the field of condensed matter physics. I have never regretted turning my attention to this amazingly wealthy part of physics.

Next, I thank Doc. RNDr. Ilja Turek, DrSc., who took the responsibility of supervising and guiding my work. He introduced me to the realm of *ab initio* calculations of electronic structure and his codes have formed a basis for my own codes. He was always ready to answer questions, and helped to overcome complications that I encountered. As a newcomer to the field, I greatly appreciated his broad overview and tried to learn from him some of his orientation in the subject. Invaluably, he also pointed in directions in which interesting phenomena might occur.

Regarding the codes, I thank Karel Carva, who passed to me his knowledge of properly running the whole chain of already developed codes. Skilfulness in this procedure is crucial for obtaining any sensible results, while it is fairly easy to go astray at virtually any moment (we joked a few times that only Mr. Turek can run the codes confidently).

I feel indebted for the care I have received from Lenka Nosková, who kept supporting me in almost every respect. She, nonetheless, has always taught me that health and life is more than any degree.

Thanks also belong to my brother, Tomáš Sýkora, who found a way to support me financially by making use of my programming skills in a with-this-thesis unrelated physical task.

Finally, I would like to pay tribute to the creators of the UNIX and Plan9 operating systems. The simplicity and generality of the related tools are, I dare say, yet to be beaten. I have used many of these (awk, sed, rc, yacc, lex, sam, acme, ...) with success for machine-assisted preparation of input files for our programs. I felt so overwhelmed by the unmatched beauty of the tools that I decided to use one of the ancient UNIX programs, troff (*1973), to typeset this thesis.§

§ It is true that TeX's qualities, at least regarding mathematical text, cannot be reached by troff. Nonetheless, the way UNIX typesetting utilities were designed and organized is, not only in my opinion, markedly more elegant.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In ..... date .....

signature

Název práce: Relativistická teorie elektronového transportu v magnetických vrstvách
Autor: Rudolf Sýkora
Katedra / Ústav: Ústav teoretické fyziky
Vedoucí doktorské práce: doc. RNDr. Ilja Turek, DrSc.,
Katedra fyziky kondenzovaných látek

Abstrakt: V práci jsou rozebrány detaily teorie funkcionálu hustoty s využitím pohledu Levyho a Lieba. Jsou odvozeny Kohnovy−Shamovy rovnice, nejprve pro nejjednodušší nerelativistický bezspinový případ, pak s přihlédnutím k existenci spinu a možnému spinovému magnetizmu, a nakonec jsou odvozeny plně relativistické Kohnovy−Shamovy−Diracovy rovnice. Diskutována je lineární muffin-tinová (LMTO) metoda výpočtu elektronové struktury se zmíněním nutných změn pro možnost započtení efektů spin-orbitální (SO) interakce do jinak skalárně relativistické (SR) teorie. Je odvozen vztah pro výpočet vodivosti vrstevnatého systému s využitím Landauerova rozptylového pohledu a s použitím jednoduchých nerovnovážných Greenových funkcí. Vybudovaný formalizmus je následně použit ke studiu vrstevnatých kovových systémů složených z lehkých prvků Co, Ni, Cu a vrstevnatých systémů obsahujících tunelovací bariéru, Fe/MgO/Ag a Fe/GaAs/Ag. Pro tyto systémy jsou diskutovány efekty spin-orbitální interakce ve vztahu k tzv. GMR poměru a/nebo velikosti tunelovací anisotropní magnetoresistence (TAMR). Jsou ukázány jednak systémy, ve kterých jsou tyto efekty malé, rovněž však i systémy, ve kterých je SO interakce významná.

Klíčová slova: Teorie funkcionálu hustoty, Kvantový transport, Magnetické multivrstvy, Tunelovací anisotropní magnetoresistence, Spin−orbitální interakce

Title: Relativistic Theory of Electron Transport in Magnetic Layers
Author: Rudolf Sýkora
Department / Institute: Institude of Theoretical Physics
Supervisor of the doctoral thesis: doc. RNDr. Ilja Turek, DrSc.,
Department of Condensed Matter Physics

Abstract: We review the density-functional theory (DFT) in detail using the Levy−Lieb approach. The Kohn−Sham scheme is discussed, starting from the simplest spinless non-relativistic case, then including spin and considering potential spin magnetism, and finally deriving the full Kohn−Sham−Dirac relativistic scheme. The Linear Muffin-Tin Orbital (LMTO) method for electronic-structure calculation is presented, together with mentioning the necessary changes to include the spin-orbit (SO) interaction effects to an otherwise scalar-relativistic (SR) theory. Derivation of an electronic-conductance formula for a layered system is given, based on the Landauer scattering picture and using simple non-equilibrium Green functions. The formalism is applied to layered metallic systems of light elements Co, Ni, Cu elements, and to layered systems with a tunnelling barrier, Fe/MgO/Ag and Fe/GaAs/Ag. The effects of the SO interaction on the Giant Magnetoresistance (GMR) ratio and/or the Tunnelling Anisotropy Magnetoresistance (TAMR) for these systems are discussed. Both, systems with large and small effects are presented.

Keywords: Density functional theory (DFT), Quantum transport, Magnetic multilayers, Tunnelling anisotropic magnetoresistance (TAMR), Spin−orbit interaction

CONTENTS

Rydberg Atomic Units

Unless otherwise stated we use the Rydberg atomic system of units, which is especially suitable for calculations on atomic scales. It sets $\hbar = 2m = e^2/2 = 4\pi\varepsilon_0 = 1$—$e > 0$ is the absolute value of the electron charge, m is the electron mass—rendering majority of quantities dimensionless (any one whose SI-unit dimension is a combination of kg, m, s, A). We then prefer plain values to values with an 'a.u.' appended to avoid confusion with both 'arbitrary units' and the Hartree atomic units. Rydberg atomic units of important quantities follow; defining SI formulæ have a symbol [SI] appended, e.g. $e'^2 = = \frac{e^2}{4\pi\varepsilon_0}[\mathrm{SI}] = 2$. Thus a unit of **angular momentum** is $\hbar \approx 1.055\cdot 10^{-34}$ Js, of **mass** is $2m \approx 1.822\cdot 10^{-30}$ kg, of **charge** is $e/\sqrt{2} \approx 1.133\cdot 10^{-19}$ C, of **length** is the Bohr radius $a_0 = \frac{\hbar^2}{me'^2}[\mathrm{SI}] = 1 \approx 5.29\cdot 10^{-11}$ m, of **energy** is the Rydberg energy $E_{\mathrm{Ryd}} = \frac{e'^2}{2a_0}[\mathrm{SI}] = \frac{me'^4}{2\hbar^2}[\mathrm{SI}] = 1 \approx 13.6$ eV, of **time** is (ratio of angular momentum and energy) $t_0 = \frac{\hbar}{E_{\mathrm{Ryd}}} = \frac{2\hbar^3}{me'^4}[\mathrm{SI}] = 1 \approx 4.84\cdot 10^{-17}$ s, of **speed** is (ratio of length and time) $\frac{a_0}{t_0} = = \frac{e'^2}{2\hbar}[\mathrm{SI}] = 1 \approx 1.094\cdot 10^6$ m/s. The **speed of light** in vacuum is then $c \approx \frac{2.998\cdot 10^8 \text{ m/s}}{1.094\cdot 10^6 \text{ m/s}} \approx 274$ (or also $c = \frac{2}{\alpha}$, where $\alpha = \frac{e'^2}{\hbar c}[\mathrm{SI}] = \frac{2}{c} \approx \frac{1}{137}$ is the fine-structure constant). The **Bohr magneton** has the value $\mu_{\mathrm{B}} = \frac{e\hbar}{2m}[\mathrm{SI}] = \sqrt{2}$.

Introduction and Motivation

Majority of electronic devices today exploit only the invariable characteristics of electrons, their electric charge, and regularly do not utilize the other intrinsic electron's property, its spin. While it is true that for usual applications like electric lights or engines, where electrons just mediate electric power, the spin is a rather unimportant quality, the situation is becoming quite different when devices processing or storing and retrieving information are considered. We illustrate this trend on probably the most revolutionary (and most spoken-about) development pertaining to information storage in today's computers.

Contemporary computers use several kinds of memories. First, there is a non-volatile memory represented by hard drives (or solid-state drives), which are slow but keep the information even when unpowered. Second, there is a fast but volatile dynamic RAM (DRAM), the main memory used by running programs. Third, there are even faster but more complicated and therefore with much smaller capacity static RAMs (SRAMs), which are also volatile and are used for buffering (caching) purposes, e.g. inside a processor unit. Among all these memories, only the first mentioned—a classical hard drive using a platter with ferromagnetic material—uses to a certain degree the electron's spin. All the others (including solid-state, i.e. flash, drives) are ignorant about spin.

Concerning classical hard drives, one could say that they have always used the spin, because the phenomenon of ferromagnetism is microscopically spin-based. The same could then be said also about ordinary magnetic tapes. Actually, the first hard drives used the same mechanism for information writing and retrieving as magnetic tapes did, namely the magnetic induction phenomenon. The whole principle, once one has accepted the existence of magnetic domains and the ability of magnetic field to rotate domains' magnetization, can be explained 'macroscopically', without a need to dive into microscopic principles. This is why people usually do not consider that such devices deploy the spin degree of freedom. The situation with hard drives has, however, changed in time, first when Anisotropic Magnetoresistance (AMR) effects were utilized, and then rather dramatically after the discovery of Giant Magnetoresistance (GMR) in 1988 by P. Grünberg and independently by A. Fert.

Magnetoresistance (MR) is a general name for a phenomenon in which resistance of a sample changes under influence of a magnetic field. Since a reading head of a hard drive is just to detect the magnetization of the platter underneath it, a material with high-enough magnetoresistance, preferably functioning at room temperature, would be an ideal building block for the head.

First signs of magnetoresistance, in bulk iron and nickel were observed as soon as in 1856 by Lord Kelvin, with a few-per-cent change reported when the magnetic field was rotated. The explanation of the effect came much later, see e.g. (Smit51, Camp70, McGu75), and goes along the lines that magnetic field aligns the spin system, and the spins affect the conductivity through the spin-orbit (SO) interaction. This kind of magnetoresistance, in which magnetization basically points in the same direction in the whole sample, is termed anisotropic, AMR.*

With the advent of fabrication techniques such as the Molecular Beam Epitaxy (MBE), it became possible to manufacture thin layered systems in which layers of one material are deposited on layers of different material. Importantly, the number of deposited atomic planes as well as their quality could be well-enough controlled. Layered systems with distinct interfaces opened an

* Sometimes the diffusive regime of electron transport is also assumed to prevail in material when speaking about AMR, since regularly the effect is measured in a macroscopic bulk of material. However, we will not adopt this terminology.

avenue to study completely new phenomena, such as transport of electrons through structures composed of alternating layers of ferromagnetic (FM) and nonmagnetic (NM) materials, in the simplest case through a trilayer schematically written as FM/NM/FM. For considered cases, there was always a method to achieve either antiparallel (AP) or parallel (P) configuration of the FM sections, e.g. using the oscillatory exchange coupling, (Grü86), for the AP configuration, and external field for the P configuration. It was in such a multilayer, composed of Fe and Cr, that the mentioned Giant Magnetoresistance (GMR) was first observed, (Baib88, Bina89).

In the GMR effect there are always (at least) two ferromagnetic metallic materials whose magnetizations can be controlled, in particular set parallel or antiparallel, and a nonmagnetic metallic material in between. Usually, the resistance of the system in the antiparallel alignment is higher than in the parallel alignment, and the difference can be sizeable (the reason for calling the effect giant); the original work by Fert *et al.* showed $R_{AP} \approx 2R_P$ at helium temperature for a superlattice 40[Fe 30 Å / Cr 9 Å]. Theoretical explanation of the effect can be found in (Zahn98, Bass99). Unlike the AMR, the GMR effect does not rely on the existence of the spin-orbit interaction, and can be, to a greater extent, explained by spin-dependent scattering at the material interfaces; nevertheless spin still plays the central role.

The discovery and following effort to utilize the GMR effect lead to an emergence of a new branch of solid-state physics, called spintronics. This fairly young field investigates the role of spin in transport of electrons and aims at designing devices that would exploit the spin degree of freedom in addition to, or instead of, the electron charge. Specifically, the GMR effect was applied in a new generation of reading heads for hard drives, which could suddenly increase their capacity thanks to the increased sensitivity and decreased size of the heads. Thus hard drives became the first serious, and by-far the most famous, application of spin-dependent electron transport.

If the nonmagnetic metallic layer of the GMR geometry is replaced by a thin-enough insulator layer (≈ 1 nm), so that electrons can tunnel through it from one ferromagnetic layer to the other, yet another effect, called tunnelling magnetoresistance (TMR), appears. The effect, again brought about by changing the alignment of the ferromagnetic layers between parallel and antiparallel, was originally discovered in 1975 by M. Jullière in Fe/Ge-O/Co junctions at 4.2 K with the relative change of resistance of about 14%. In the 1990s experiments using amorphous Al_2O_3 showed a relative change in tens of per cent at room temperature. In 2008 effects of up to 600% at room temperature and more than 1100% at 4.2 K were observed in junctions of CoFeB/MgO/CoFeB, (Iked08). As far as applicability is concerned, the use of TMR is quite similar to that GMR, with some advantages over GMR. Apart from displaying a much larger magnetoresistance effect, a TMR structure incorporates more easily into a circuit since its resistance can be optimized just by changing the width of the isolating layer.

In spite of similarities of TMR and GMR, their underlying mechanisms are, however, distinct. A simplified description of the TMR effect is given by the Jullière model, (Jull75). Perhaps noteworthy is that like GMR, TMR would exist even without the spin-orbit interaction.

Multilayers with a strong TMR effect seem to be promising building blocks for a new generation of memories, called magnetic RAMs (MRAMs). These could combine many of the positive qualities of different types of RAMs listed at the beginning of this introduction. The information would be stored as a stable configuration (P or AP) of two magnetic parts of a small multilayer (each multilayer would store one bit), and the storage would be thus non-volatile. Reading would be achieved by passing current and measuring the resistance of the structure, and this process can be as fast as for SRAMs. The density of information, on the other hand, since the structure itself is small, could reach the values for DRAMs. Finally, although writing information into the structure so far can only be effectively achieved via an inductive process, there is an ongoing research to use other more subtle phenomena like spin-transfer torque that would allow even further size decrease of the structure. To the knowledge of the author, some prototypes of MRAMs already exist (though only with the inductive writing).

There exist quite a few other magnetoresistance effects and we mention one more in which we will be particularly interested, the Tunnelling Anisotropic Magnetoresistance (TAMR). This type of a magnetoresistance is studied in systems with only one metallic magnetic layer whose

magnetization can be rotated, and which is separated from another non-magnetic metallic part by a tunnelling (insulating) barrier; schematically FM/I/NM. Resistance of the structure may change with FM-layer-magnetization orientation, which is, similarly to the AMR we started our tour with, a consequence of the spin-orbit interaction.

So far we have mainly mentioned the practical use and possible impact of newly-born spintronic devices, which is something an engineer is after, rather than a physicist. Therefore, we should perhaps emphasize that spintronic devices also allow a direct study of fundamental quantum physics. The devices are rather small (they are often termed mesoscopic, which means they are much larger than microscopic objects like atoms, but not large enough to be 'ohmic') and phenomena like quantum interference, Anderson localization, etc., are all within reach.

In this thesis we try to contribute a little to the understanding of the role the spin-orbit interaction plays in electron transport within mesoscopic layered structures. Certain phenomena, like the mentioned TAMR, are directly conditioned by the existence of this interaction, while other phenomena, like the GMR, can be to a greater or lesser extent influenced by it. The knowledge of how it can affect a system can then be quite crucial for an appropriate design of new spintronic devices.

Particularly, we investigate the situation in metallic layered structures composed of Co, Ni, Cu elements, where it seems that the more disordered system we prepare (by varying the widths of individual one-material sections) the more relevant the SO-interaction effects can be. Further, we study systems with a tunnelling barrier, such as Fe/GaAs/Ag, and discuss some subtle spin-orbit-induced phenomena which may contribute to sizeable values of TAMR.

PART I — Theory

1

Electronic Structure

The phrase ‘Electronic structure of a material’ refers to the energy levels of the material’s electrons and to their distribution in space and in momentum/energy. Electrons hold nuclei in a solid, liquid, molecular states, and it is their configuration that primarily determines observed electrical, optical, thermal, magnetic, and many other properties of a material. Therefore, characterization, computation, and measurement of electronic structure is the central problem of solid state physics.

In this theoretical thesis we are interested in computer-aided calculation of electronic structure (and from it derived electronic-transport properties) of layered magnetic systems. Such calculations may proceed on qualitatively different levels. On the one hand there are methods more or less based on phenomenology, such as empirical tight-binding models, which frequently contain number of loosely controlled approximations alongside with several tuning (fitting) parameters. These models, often fast and genuinely simple, help to gain necessary insight by capturing the essence of studied phenomena. On the other hand there are so-called *ab initio* approaches, in which as few parameters as possible enter the calculation (usually only the atomic numbers of present atoms and their positions). These methods are robust, fairly complicated and incorporating many effects at once (and so, as opposed to the former methods, interpretation of obtained results can be more difficult). In this work we will exclusively use—and try to enrich—the latter (*ab initio*) methods, pursuing also the fundamental goal of constructing reliable techniques that would agree well with experimental findings.

1.1. Density Functional Theory (DFT)

In quantum mechanics, any physical system is affiliated with an appropriate mathematical (complex) Hilbert space, and any (pure) state of the system corresponds to a ray in this space. Such description is fundamentally more complex than that of classical physics: A single featureless particle confined in a volume Ω is classically described by its position and momentum, i.e. by 6 real numbers (in the ordinary 3D space). In the quantum world the proper Hilbert space for the case is that of the $L^2(\Omega)$-functions, already an infinite-dimensional space (the price for possible delocalization, interference, etc.). If more particles are considered, classically the needed space dimension grows linearly with the number of particles ($6n$ in 3D), while quantum-mechanically for interacting particles it is an exponential growth (‘∞^n’).†‡

† In our chosen case the 1-particle space is infinite-dimensional. Generally, it does not have to be, and practically, when calculations are carried out, never is. Then the ∞ is to be replaced with the finite dimension, in which case the mentioned scaling gains a proper mathematical sense.

‡ Actually, this scaling would be right only if the particles were so-called distinguishable (apart from featureless). In fact, our systems will generally consist of indistinguishable electrons, which moreover have an inner degree of freedom—spin. While the spin only doubles the fundamental 1-particle-space dimension, the indistinguishability implies that quantum statistics (the exclusion principle) limits accessible states, so the dimension of true ‘physical’ space is somewhat less than the naïve analysis presented would give. Nonetheless, the main point is that the problem gets complicated quickly with growing particle number.

Solid state physics deals mainly with systems containing a large number of particles.* Direct attempt to solve either classical or quantum equations of motion including all particles in general ambient conditions is plain hopeless.** One must both appeal for help of simplifying circumstances (room temperature is effectively almost zero temperature from the point of view of electrons in metals; symmetry considerations in the case of crystals, etc.) and decide how detailed knowledge we actually wish to acquire. Often we are only interested in ground-state (or thermodynamically averaged) values of 1- and/or 2-particle observables such as the total energy, average speed or spin of particles, average potential energy per particle, and alike. Concentrating on quantum systems this means that instead of the full N-particle wavefunction $\Psi(1,2,3,\ldots,N)$,§ or more generally the full density matrix $D^{[N]}=\sum_j p_j|\Psi_j><\Psi_j|$—to allow a mixture of several pure states j; $\sum_j p_j=1$—, we need only reduced 1- and/or 2-particle density matrices $D^{[1]}(1,2)$, $D^{[2]}(12,34)$, obviously much simpler objects than $D^{[N]}$. Unfortunately, calculation of $D^{[n]}$ generally requires the knowledge of $D^{[n+1]}$, thus an approximation necessarily has to be applied by cutting the chain at some point (effectively neglecting higher-order correlations). Alternatively, if a general classification (parametrization) of admissible $D^{[2]}$ existed, we could also start from the other end: from $D^{[2]}$ we calculate $D^{[1]}$ and then evaluate the average value of the system's Hamiltonian $<H>$ (assuming H contains at most 2-particle operators). Finding stationary points of $<H>$ would then correspond to the eigenstates of H. Unfortunately still, to the knowledge of the author, there exists no such classification for $D^{[2]}$.

In view of this situation, the early attempts of L. H. Thomas (Thom27) and E. Fermi (Ferm27, Ferm28), and the extensions suggested by P. A. M. Dirac (Dira30) and C. F. von Weizsäcker (Weiz35), which express the ground-state energy in terms of the density alone (i.e. as a function of $D^{[1]}(1,1)$), have to be considered as a heuristic wish to cut the chain of the many-particle correlations indicated above at its very beginning.

Having said all this, it may come as a surprise that in 1964 P. Hohenberg and W. Kohn provided a formal proof (Hohe64) that under quite general conditions¶ the ground-state energy of a system is a functional of the ground-state density, and, not less importantly, that the ground-state density can be calculated—without recourse to the Schrödinger equation—from a variational principle involving just the density. I.e., the chain of correlations can actually be disentangled completely.

Before continuing we believe a note is in place. It should not at all be surprising that there exists *some* simple function (a real function of position) that can embrace so much information about a system. After all, the complete information about a system (see the next section for precise definitions) is from the beginning contained in the external potential, itself being a simple function (the system is *defined* by the potential and we, of course, cannot get out more than we have put in). In this sense everything is a functional of the external potential. However, such a statement is both next to trivial and useless.‖ The (first) Hohenberg–Kohn theorem tells quite more. Unlike the external potential, which defines the *external* set-up, the theorem specifies the extent of details to which we must know the system's *internal* configuration (i.e., its 'response' to the external potential) in order to be able to proceed with finding the ground-state energy and not miss any important information along the way. We know from the theorem that the description by density is sufficient. (This means that there is an intimate relationship in which the external potential maps itself onto the system's density response; see also Fig. 1.1.)

Almost all ab initio condensed-matter calculations of these days are based on the

---

* A cubic centimetre of iron contains $\approx 10^{23}$ Fe atoms.

** W. Kohn in his Nobel lecture, (Kohn99), argues that traditional wave-function methods which are to provide 'required' chemical accuracy are generally limited to molecules with a small total number of chemically active electrons, $N\approx O(10)$; he calls this limitation an exponential wall.

§ We use the notation $1\equiv \boldsymbol{r}_1\sigma_1$, etc., σ variables denote spin degrees of freedom.

¶ We provide more details in the following section.

‖ At least until somebody provides us with a concrete form of the functional (which is not precluded but seems improbable today).

Hohenberg–Kohn theorems and their extensions. Calculations presented in this thesis are not an exception. Corresponding theories, for their use of density as the basic variable, are called density-functional theories (DFTs). A review of DFT can be found e.g. in (Jone89), (Jone06) or (Fiol03).

Let us now have a closer look at the theorems by Hohenberg and Kohn.

1.2. DFT for a Nonrelativistic System in a Local Spin-Independent External Potential

Let us consider a system of N identical particles in a volume Ω, each of mass m and possibly non-zero spin s, obeying non-relativistic quantum mechanics, and moving under influence of an external potential V_{ext}. Let us limit the form of the external potential to $V_{\text{ext}} = \sum_i v_{\text{ext}}(\boldsymbol{r}_i)$, i.e. to a sum of spin-independent local contributions from individual particles at points $\boldsymbol{r}_i$; $v_{\text{ext}}(\boldsymbol{r})$ is a potential function defined on Ω (due to nuclei in most practical cases). Equivalently expressed, $V_{\text{ext}} = \int v_{\text{ext}}(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$, with $n(\boldsymbol{r})$ being the particle density at point $\boldsymbol{r}$. The Hamiltonian H of the system has then three parts: 1) a '*kinetic-energy*' term* T, that is a sum of the individual-particle kinetic-energy operators $(-\nabla_i^2/2m)$ (in the x-representation), 2) a term U describing the *interaction* between the particles (we need not specify U in detail here; it can be whatever we like, even velocity-dependent, non-local, not pairwise, etc., although usually it is just the Coulomb interaction), and 3) the already mentioned *external potential* term V_{ext}. In short, $H = T + U + V_{\text{ext}}$.

While the first two terms of the Hamiltonian depend only on the nature of considered particles (their mass and mutual interaction), the external potential V_{ext} expresses the influence of the environment. For example, both a neutral silicon atom Si and a C_2H_2 molecule have 14 electrons (thus the T and U parts of the two Hamiltonians are exactly the same), but these electrons move in different external potentials created by a single Si nucleus or two C nuclei, respectively.‖

In the just mentioned 14-electron case it is rather intuitive that the different potentials lead to quite different ground-state particle densities. However, it may not be clear if different potentials *always* lead to different ground state densities. This is where the first Hohenberg–Kohn theorem comes into play. Here we use its extended version by Levy (Levy79, Levy82) and independently by Lieb (Lieb82) to show that, under the mentioned conditions on V_{ext}, it is really so.

First we note that shifting the external potential by a constant does not have any physical impact. Thus, we *define* two potentials to be different only if they differ by more than an additive constant. Now, let us calculate the ground-state energy E_g of our N-particle ($\int_\Omega n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} = N$ for any n in the following) system. Denoting a general quantum N-particle state with ψ, we write§

$$\begin{aligned} E_g &= \min_{\psi} \langle \psi | H | \psi \rangle = \min_{n} \min_{\psi \to n} \langle \psi | T + U + V_{\text{ext}} | \psi \rangle = \\ &= \min_{n} \left[\min_{\psi \to n} \langle \psi | T + U | \psi \rangle + \int v_{\text{ext}}(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \right] = \\ &= \min_{n} \left[F_{\text{LL}}[n] + \int v_{\text{ext}}(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \right] = \min_{n} E_{v_{\text{ext}}}[n], \end{aligned} \tag{1.1}$$

where the overall minimization has been split into two steps: first we minimize over all ψ giving a certain particle density profile $n(\boldsymbol{r})$§§ and only then we minimize over all possible profiles $n(\boldsymbol{r})$ (that integrate to the total of N particles). The splitting has naturally lead to the definition of $F_{\text{LL}}[n] = \min_{\psi \to n} \langle \psi | T + U | \psi \rangle$, which is—by its construction—a universal, i.e. independent of v_{ext},

---

* The reason for the quotation marks will be explained below (there is a subtlety when a system gets spontaneously magnetized).

‖ We assume there is not significant magnetic (i.e. spin-dependent) interaction between the nuclei and electrons, so that the potential can be considered spin-independent, as requested.

§ To keep the notation light we quite freely denote operators and their expectation values with the same letter (e.g. the $n(\boldsymbol{r})$ in the definition of V_{ext} [an operator] and in equation (1.1) [an expectation value]). What is meant on each occasion should hopefully be clear from the context; or is explicitly mentioned and/or emphasized with a $\hat{}$.

§§ There are often more ψs giving the same n, recall e.g. that all plane-waves $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ yield the same (constant in this case) density.

existing but to-date unknown functional of the density n; the index LL stands for Levy-Lieb.† The final line of (1.1) claims:

i) The ground-state energy E_g is obtainable by a minimization of $E_{v_{\text{ext}}}[n]$ running over densities $n(\boldsymbol{r})$ that integrate to N particles.

ii) If we know that $n_g(\boldsymbol{r})$ is the system's ground state density, then
   - the system's ground-state energy is $E_g = F_{\text{LL}}[n_g] + \int v_{\text{ext}}(\boldsymbol{r})\, n_g(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$, and
   - the external potential can be determined from $v_{\text{ext}}(\boldsymbol{r}) = -\dfrac{\delta F_{\text{LL}}}{\delta n_g(\boldsymbol{r})}$, (up to an additive constant).¶‡

These points contain (better: generalize) the claims of the original two theorems by Hohenberg and Kohn [which are, 1st: for a system of interacting particles in an external potential $v_{\text{ext}}(\boldsymbol{r})$ and with a non-degenerate ground state, the potential $v_{\text{ext}}(\boldsymbol{r})$ is determined uniquely, up to a constant, by the ground state particle density $n_g(\boldsymbol{r})$; and 2nd: for an external potential $v_{\text{ext}}(\boldsymbol{r})$, there is a functional $E_{\text{HK}}^{v_{\text{ext}}}[n] = F_{\text{HK}}[n] + \int n v_{\text{ext}}$, which assumes a minimum (in the space of densities which integrate to the total of N particles and are v-representable, see below) for the true ground-state density, and F_{HK} is v_{ext}-independent, i.e. universal, functional].

The ii) point confirms that different external potentials lead to different ground state densities (if it were not the case, we would have a common density for two differing potentials, but according to the ii) point the density necessarily identifies the causing potential, which is a contradiction). The relation between external potentials and corresponding quantum ground states and densities is depicted in Fig. 1.1.

From the pragmatic point of view, point i) is of the foremost importance, principally embracing the computational program. If we were given some prescription for F_{LL}, we might then try different densities to minimize $E_{v_{\text{ext}}}[n]$, and eventually obtain an estimate for the ground-state energy. For reasons mentioned below, however, we will not use such a straightforward method. Instead we will be based on the Kohn–Sham scheme, described further in the text.

Having shown the essentials of Levy and Lieb's proof of the Hohenberg–Kohn theorems

---

† Their formulation is for obvious reasons called the constrained-search formulation of DFT.

¶ In the ground state, $0 = \left.\dfrac{\delta E_{v_{\text{ext}}}}{\delta n_g(\boldsymbol{r})}\right|_{\int n = N}$; using a Lagrange multiplier μ to take care of the constraint we get $\dfrac{\delta F_{\text{LL}}}{\delta n_g(\boldsymbol{r})} = -v_{\text{ext}}(\boldsymbol{r}) + \mu$, and we ignore the μ, because it is only a constant shift. If there are several ground-state densities, any one can be used in the variation, always to return the same v_{ext}.

‡ In the beginning, it was the author's wish to carefully present the DFT's details at least for the non-relativistic case. However, it turned out to be too ambitious within the available time and space for several reasons. The theory's mathematical background is rather complex, see (Lieb82). Further, many points were discussed only in not-easily-accessible sources, and today's physical texts generally neglect them altogether (Drei90, Esch03 being exceptions). Thus, eventually, the author was forced to give up for the moment and only note, both generally here and later at some points explicitly, that matters are more complicated than is shown. For instance, already the existence of the minima (as opposed to just infima) in (1.1) is not obvious; we just assume they exist. Next, the existence of the functional derivative of F_{LL}, as presented in the ii) point, is actually unknown (despite the common habit of just writing it out; even worse, sometimes using the original Hohenberg–Kohn functional). In (Engl84, Theorem 6.1) it is proved that a derivative exists for a related Lieb's functional (usually denoted by F_{L}; Lieb himself used F_{DM} in Lieb82) at the so-called ensemble-v-representable densities and nowhere else; thus, had we started with F_{L}, we would have been on a bit safer side, since physical minima are naturally ensemble-v-representable. [F_{L} is more complicated than F_{LL}—the reason why we eventually opted for F_{LL} in the main text—, nonetheless, F_{L} is otherwise preferred for its several important properties, major one being that it is a convex functional (F_{LL} is not), which is of practical importance when differentiability is to be proved.] But even F_{L} is not really sufficient for a full theory reformulation in terms of the usual functional derivatives since the mentioned existing derivative of F_{L} is still only a derivative within the space of N-particle densities, while the usual functional derivative is meant in a δ-function direction and thus essentially leaves the domain of exactly N-particle densities (this point is further mentioned in the main text). Hence yet another extension is needed. Optimistically, it seems that the mentioned and related difficulties have been overcome and the theory is well grounded. On the operational level we can pretend that everything just works smoothly. Finally, we note that if one is only concerned with a theory living on a grid, majority of such difficulties cease to exist (see e.g. Chay85).

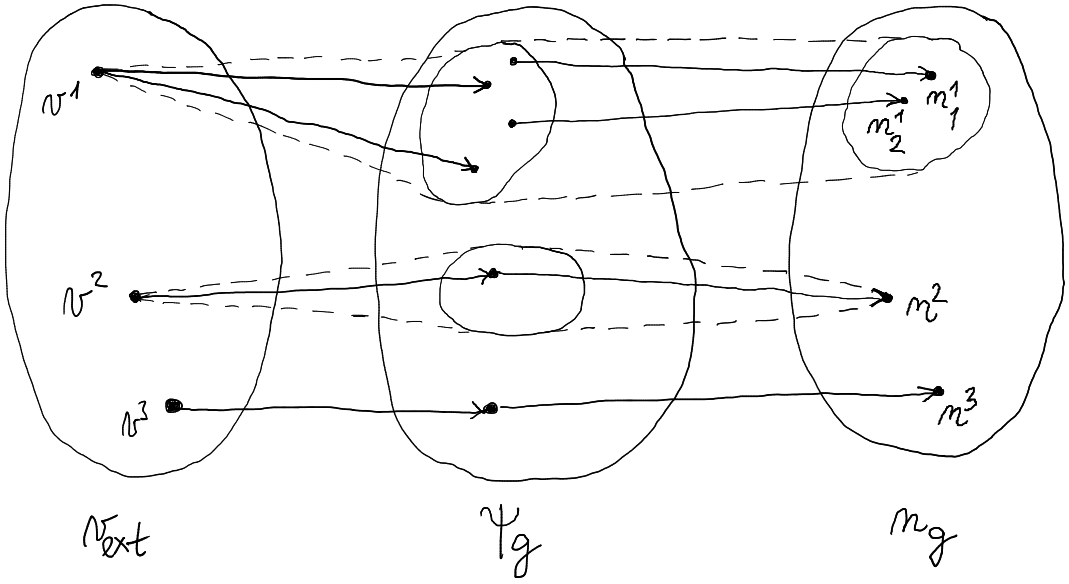


Figure 1.1: The relation of external potentials to corresponding ground states and ground state densities. In general (as for v^1) a potential may lead to a degenerate set of quantum ground states, each possibly having a different density ($n_1^1 \neq n_2^1$; then the external-potential contributions to the total energy, $\int nv$, differ, however, this difference is compensated for by the T and U terms so that the total energy stays the same, i.e. the ground-state energy). A nice example demonstrating the general case is a noninteracting ($U=0$) Li atom with 4 degenerate ground states: $1s^2 2s$, $1s^2 2p^+$, $1s^2 2p^0$, $1s^2 2p^-$. The first three yield different densities, the last one the same as the second one. On the other hand, in the simplest case (as for v^3) a potential leads to a non-degenerate ground state and thus also to a unique density. The intermediate case (v^2) has a degenerate set of ground states, yet all these lead to the same density. Note that the original Hohenberg–Kohn theorem only discussed the simplest case à la v^3.

several related notes are in order.

First, unlike the original theorems, which required the ground state to be non-degenerate, the scheme just presented has no such limitation.† Second, the original theorem proved the existence of a minimized density functional very like our $E_{v_{\text{ext}}}[n]$, we denoted it $E_{\text{HK}}^{v_{\text{ext}}}[n]$ above, however the domains of the two functionals fundamentally differ. The former, $E_{v_{\text{ext}}}$, being based on F_{LL}, is defined for any density that can be obtained from some quantum state ψ;‡ such $n(\boldsymbol{r})$ is called *N*-representable. T. L. Gilbert (Gilb75) and later others* have shown that any reasonable density (non-negative, normalizable, with $\int |\nabla n^{1/2}|^2 < \infty$) is *N*-representable. Thus on the practical level, F_{LL} is defined for any n. On the other hand, the latter functional, $E_{\text{HK}}^{v_{\text{ext}}}$, based on the functional $F_{\text{HK}}[n]$,¶ which has only been defined for densities that correspond to a ground state of some external potential; such densities are termed (pure-state) *v*-representable. As pointed out by Levy (Levy82) and Lieb (Lieb82) (one class of examples), and H. Englisch and R. Englisch (Engl83) (another, more serious class of examples) there exist perfectly well-behaved functions $n(\boldsymbol{r})$ that nonetheless cannot represent a ground state of any external potential; such ns are thus not *v*-representable and F_{HK} is not defined for them.§ Still worse, because no characterization of *v*-

† Actually, as noted in (Kohn85) and discussed in (Drei90, ch. 2.2) the non-degeneracy requirement can actually be relaxed even in the original Hohenberg–Kohn approach.

‡ We stress that ψ can be arbitrary in this context, not necessarily a ground state of some external potential.

* For details and further references see e.g. chapter 4.4 of (Esch03) and chapter 2.3 of (Drei90).

¶ $F_{\text{HK}}[n] = \langle \psi_n | T+U | \psi_n \rangle$, where ψ_n generates density n, and n is required to be a ground state density for some external potential (otherwise F_{HK} is undefined). Thanks to assumed non-degeneracy in the original theorem ψ_n is unique, see Fig. 1.1.

§ Any such n is, in accord with the aforementioned, still *N*-representable, i.e., there is at least one (but possibly more than one) wavevector ϕ generating the density. $F_{\text{LL}}[n]$ keeps existing.

representable densities is known, the formal application of the original theorem may easily become problematic: while looking for a minimum of $E_{\mathrm{HK}}^{v_{\mathrm{ext}}}$ one can easily leave the functional's domain, i.e., simple variational method is not justified for $E_{\mathrm{HK}}^{v_{\mathrm{ext}}}$ since the varied density may possibly not be *v*-representable. Of course, any real ground-state density is by definition *v*-representable and the two functionals' values equal there ($F_{\mathrm{LL}} = F_{\mathrm{HK}}$ whenever F_{HK} is defined [thus F_{LL} is a reasonable extension of F_{HK}] $\Rightarrow E_{v_{\mathrm{ext}}} = E_{\mathrm{HK}}^{v_{\mathrm{ext}}}$) and also both realize their minima (a direct consequence of the Rayleigh−Ritz principle).

An interesting and important consequence of the Hohenberg-Kohn theorems is that since a ground-state density uniquely identifies the external potential $v_{\mathrm{ext}}(\boldsymbol{r})$, the whole Hamiltonian is identified, too. Any 'internal' property of the system is then fixed (including e.g. the system's excitation energies). As a result, for instance, any thermodynamically averaged equilibrium value of any observable is fixed by the system's ground-state density, too.¶ Pictorially:

$$n_g(\boldsymbol{r}) \xrightarrow{\text{1st HK theorem}} \underset{\text{(up to a constant)}}{v_{\mathrm{ext}}(\boldsymbol{r})} \longrightarrow H \longrightarrow \underset{\text{(e.g. } \rho = e^{-\beta H}/Z \text{ in equilibrium)}}{< O > \; = \; \mathrm{Tr}\, \rho O.}$$

We mentioned the pleasant property of the functional $F_{\mathrm{LL}}[n]$ that it is defined for practically all densities $n(\boldsymbol{r})$, and we somewhat tacitly assumed the density integrates to an integral particle number *N*. However, for the above-indicated variational formulation to have the expected meaning we really must extend the functional domain to non-integral densities, $\int n(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = N = M + \omega$, with *M* integral and $0 \le \omega < 1$. Only then we can talk about $\delta/\delta n$ in the usual sense of a derivative in the δ-function direction. Such an extension is possible and is described in (Drei90, ch. 2.4); therein they also take account of existing discontinuities in $\mu(N) = \partial E_g(N)/\partial N$ when *N* crosses integral values (such discontinuities are important in small systems and can be hopefully neglected in large ones we are concerned with). We will not go into details here.

In this section we have discussed systems in spin-independent external potentials. Nonetheless, a system itself may consist of particles featuring non-zero spin (electrons). Such systems can get, thanks to the inter-particle interaction, spontaneously magnetized.† A spontaneously magnetized system necessarily breaks the symmetry of its Hamiltonian, hence its ground state has to be degenerate: The symmetry break-down can happen in several equivalent 'directions' and if we calculated the usual thermodynamic average of the magnetization, we would obtain zero, since we would average over all these states with equal weight. Although a single ground state possesses a non-zero magnetization, the full (naïve) thermodynamic average must respect the symmetry of the Hamiltonian and be zero.

Regarding systems with broken symmetry, let us note, that the fact that the average obtained from the density-matrix operator $\rho = e^{-\beta H}/Z$, $Z = \mathrm{Tr}\, e^{-\beta H}$, $\beta = 1/kT$, is not what is observed (we really observe one particular, somehow chosen, ground state and a non-zero magnetization) indicates, that such a ρ is not really relevant in the considered situation. The principle leading to such ρ assumes that all states with equal energy are equiprobable, which is further based on the idea that all such states are 'visited' for equal amounts of time (ergodicity). But although the individual ground states necessarily have the same energy, there is no easy path that would connect them

¶ We mention the thermodynamic equilibrium because then the occupation of individual system's states is known: the density matrix is $\rho = e^{-\beta H}/Z$, i.e. a function of *H*, which is a functional of *n*. Generally, we may proceed with any (if given) occupation distribution. Or, perhaps, the expectation value of the operator under study may be partially independent of the exact states' occupation (we can redistribute the quantum amplitude among states in which the operator has the same expectation value). An example of the latter is the ground-state Hamiltonian expectation value, which is insensitive to the actual choice of a ground state from a possibly degenerate set. On the other hand, situation for other operators may be different. For instance, a magnetization $\boldsymbol{m}$ of a hydrogen atom, if we only know that it is in a further-unspecified ground state, cannot be a functional of the density alone: the ground state is twofold degenerate (spin 'up' and 'down') with exactly the same density $n(\boldsymbol{r})$ in both cases. The picture is $n \to v_{\mathrm{ext}} \to H \to$ degenerate ground state, each member state possibly with a different expectation value of the studied operator, here $\boldsymbol{m}$, and we need some further information to determine the expectation value.

† Perhaps surprisingly, the interaction does not have to be spin-aware; actually, it is the plain Coulomb interaction (plus statistics) that stands behind the usual magnetic behaviour.

when spontaneous symmetry-breaking takes place.‡ For example, tilting of magnetization of an isotropic ferromagnet can be managed equienergetically only by rotating all the microscopic moments simultaneously, which is highly improbable to happen only due to fluctuations. The situation in a real ferromagnet, which has a crystalline structure, is even simpler. Thanks to the crystalline anisotropy there is only a finite number of directions in which the ground state magnetization may point (easy magnetization axes). Here, there is not an equienergetic path connecting the individual ground states at all, and the probability of magnetization switching from one easy direction to another is utterly negligible. This explains why the thermodynamic averaging via the ordinary density-matrix ρ does not work. The way how to actually proceed with averaging in cases when spontaneous symmetry-breaking may happen was studied e.g. by N.N. Bogoliubov (Bogo70), who introduced the concept of the so-called quasi-average (as opposed to the ordinary, or regular, average). Briefly, he explicitly adds a symmetry-breaking term to the Hamiltonian, which effectively chooses the 'direction', and only then computes the limit of the studied quantity with the extra term going to zero. If the system is susceptible to a symmetry-breaking, the limit, say of magnetization, is finite.* Otherwise it is zero (it is instructive to compare the behaviour of a ferro- and para- magnet in this process). Having explicitly added the symmetry-breaking term to the Hamiltonian, $H \to H' = H + \delta$, one may then safely use the new equilibrium density-matrix operator $\rho' = e^{-\beta H'}/Z'$ to compute now physically relevant expectation values (taking finally the limit $\delta \to 0$). Namely, the magnetization is a functional of the ground-state density (and a function of temperature).

If a non-zero magnetization exists in a system, the kinetic-energy operator T, as defined, is actually no longer the correct kinetic-energy operator, since it ignores the contribution from the magnetic field.‖ This is why we used the quotation marks when we defined the T term. However, what really matters is the sum of $T+U$ (this combination enters the definition of F_{LL}), see (1.1), so as long as U covers what is missing in T (i.e. terms $\propto \boldsymbol{p}\cdot\boldsymbol{A}$, $\boldsymbol{A}\cdot\boldsymbol{p}$, $\boldsymbol{A}^2$), the theory is exact, albeit the T operator is then somewhat misnamed; we declared that U can be rather arbitrary, even velocity-dependent, thus everything is all right in principle. Moreover, even if the U part were just the Coulomb interaction alone (as is the case in the so-called spin-only [non-relativistic] density functional theory) and did not compensate for the by-T-missed magnetic terms, the error thus caused is often small.

The theory as so far presented is rather abstract in the sense that as yet there is no existing usable functional $F_{\mathrm{LL}}[n(\boldsymbol{r})]$ (not for an unpolarized, still less for a spin-polarized case). Before showing a possible way around—the Kohn–Sham scheme—we try to offer a sketchy reasons for this situation.

Attempts to construct a theory that would use only density (and not solve the Schrödinger equation), such as the Thomas–Fermi theory, has showed a lack of ability to sufficiently allow for the kinetic energy of particles. To assess the difficulties, let us consider a noninteracting particle system in a periodic external potential. For this case we can calculate the energy spectrum exactly and know it corresponds to a set of bands. Adding particles to the system means gradual filling of the bands. Notably, the increase in the system energy per added particle is quite different when a band is not yet full and when the band has just been filled and we have to cross the gap. While this step-like behaviour is clearly 'visible' in the orbital/band language, the same cannot be easily extracted from the density. Perhaps a bit vaguely put: to anticipate any non-analytic (step-like; e.g.

---

‡ The author believes this is a truly important attribute of spontaneous symmetry-breaking. Imagine a system of noninteracting particles with non-zero spin, possibly in an external spin-*in*dependent potential. Then the system's ground state is highly degenerate (spins can be independently and arbitrarily rotated). Among the many ground states there are also ones in which all spins point in the same direction. However, we would not probably claim that a symmetry-breaking occurs, although such states indisputably have a lower symmetry than the Hamiltonian.

* The situation when a regular average yields zero while the quasi-average gains a non-zero value just because of an infinitesimal change of the Hamiltonian is called the 'instability of the regular average'. (We also note that for the quasi-average trick to work, the thermodynamical limit $\Omega \to \infty$ must be carried out before limiting the breaking Hamiltonian term back to zero.)

‖ The correct full kinetic-energy operator, in the x-representation, would be $\sum_i (-i\nabla_i - q\boldsymbol{A}(\boldsymbol{r}_i))^2/2m$.

at Brillouin zone boundaries) behaviour just from the density, one cannot just observe a local region, but would have to scrutinize the density on the global scale, which would lead to a highly non-local functional. Even more hopeless is the situation for spin-polarized systems if one would like to use only a total-density functional (which is, as we noted, theoretically possible). The Hamiltonian for such a system may not contain a single spin-aware term, yet the functional would be requested to yield a ground state density corresponding to a correct one of a polarized system.

1.3. The Kohn–Sham Scheme for a Nonrelativistic System in a Local Spin-Independent External Potential

In the previous section we mentioned that all density functionals available lead to partially deficient results if one determines the ground-state density and energy directly from the described variational principle, and that the major source of deficiencies is an inappropriate representation of the kinetic-energy contribution $T[n]$. To circumvent the problems (or at least make them less pronounced) a method suggested by W. Kohn and L. Sham (Kohn65), known as the Kohn–Sham scheme, can be used.**

The main idea is an introduction of an *auxiliary noninteracting–particle system* with a tight relation to the studied interacting system. Any noninteracting system has the advantage that its properties, including the kinetic energy, can be calculated exactly (in terms of related occupied orbitals—not as an explicit functional of density—and thus incorporating the quantum properties that have no simple relation to the density). If we can devise an external local *potential* V_0 *that generates the same ground-state density in the noninteracting system as is the ground-state density in the original system* (with an external potential V), we can use the noninteracting system to approximate the kinetic energy of the interacting system. Experience then shows that the difference between thus calculated noninteracting kinetic energy T_0 and the full interacting-system kinetic energy T is sufficiently minor to be negligible (or the difference can also be further compensated for within the exchange-correlation functional E_{xc} introduced below). In a sense, the outlined program is a construction of an effective mean-field theory similar to the Hartree theory (since we use only a local V_0),† however with the exchange and correlation built in on the density-functional level.

Let an N-particle interacting system with Hamiltonian $H = T + U + V$, where T represents the kinetic energy, U the inter-particle interaction, and $V = \int v(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}$ the local external potential,* have a ground-state density $n(\boldsymbol{r})$. (The system may have more ground state densities, see Fig. 1.1, here we consider any single one; there is a comment on this below.) The related (interacting) ground-state energy E can be expressed following our desire to utilize the kinetic energy T_0 of the auxiliary system

$$\begin{aligned} E \;&=\; T+U+V \;=\; (T-T_0+T_0) \;+\; (U-U_{\mathrm{Hartree}}+U_{\mathrm{Hartree}}) \;+\; V \;= \\ &=\; [T_0+U_{\mathrm{Hartree}}+V] \;+\; [(T-T_0) \;+\; (U-U_{\mathrm{Hartree}})] \;= \\ &=\; [T_0+U_{\mathrm{Hartree}}+V] \;+\; E_{\mathrm{xc}}\,, \end{aligned} \tag{1.2}$$

where we have also separated from U the Hartree energy, $U_{\mathrm{Hartree}} = \int \frac{n(\boldsymbol{r}')\, n(\boldsymbol{r}'')}{|\boldsymbol{r}'-\boldsymbol{r}''|}\, \mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{r}''$, and defined the so-called exchange-correlation functional $E_{\mathrm{xc}} = (T-T_0)+(U-U_{\mathrm{Hartree}})$.‡¶ [From the

** We note that the scheme is also used (although justification may be questioned) to cope with yet another problem that there is generally no straight way to express properties of a system only from the density, as opposed to the situation when the wavefunction is known. The Kohn–Sham scheme supplies orbitals, and if one has reasons to believe they have some direct physical relevance, this difficulty is removed.

† There are also other choices of auxiliary systems. One particular is proposed already in the the original paper (Kohn65), where a *non*-local effective V_0 is used, leading to equations similar to those of the Hartree–Fock theory.

* For brevity we have dropped the 'ext' label used in the previous section: $V_{\mathrm{ext}} \to V$.

‡ As defined, E_{xc} contains the kinetic-energy correction $(T-T_0)$. Often, however, the 'exchange-correlation energy' found in literature may have a slightly different meaning. See for example U. von Barth's contribution in (Bart94), pp. 21–62, esp. chapter 2, for a comparison of expressions for E_{xc} and the 'exchange-correlation part of the interaction energy' $U_{\mathrm{xc}} = U - U_{\mathrm{Hartree}}$. It is shown there that while $U_{\mathrm{xc}} = \int \frac{n(\boldsymbol{r})\, n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} [g(\boldsymbol{r},\boldsymbol{r}')-1]\, \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}'$, with $g(\boldsymbol{r},\boldsymbol{r}')$ being the

Hohenberg–Kohn theorem we know that E and V are functionals of n; U_{Hartree} is an explicit functional of n; finally T_0 is shown to be a functional of n below. It then follows that also E_{xc} is a functional of n.] While the three terms in brackets on the last line of (1.2) can be, for a given n and known v, evaluated exactly, the last term, $E_{\text{xc}}[n]$ confines all the complicated many-body effects and has to be approximated.

Let us now express the noninteracting-system kinetic energy T_0. The Hamiltonian of the auxiliary system reads

$$\hat{H}_0 \;=\; \hat{T} + \hat{V}_0 \;=\; \hat{T} + \int v_0(\boldsymbol{r})\,\hat{n}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}. \tag{1.3}$$

Since the system is noninteracting, the ground state—which we first assume to be non-degenerate, note the exclamation mark in (1.5)—has the form of the Slater determinant

$$H_0\Psi_0 \;=\; E_0\Psi_0, \qquad\qquad \Psi_0 = \frac{1}{\sqrt{N!}}\det[\phi_1\phi_2\cdots\phi_N], \tag{1.4}$$

of N lowest-energy 1-electron orbitals ϕ_i solving the 1-particle equations

$$[-\nabla^2 + v_0(\boldsymbol{r})]\;\phi_i(\boldsymbol{r}) \;=\; \varepsilon_i\phi_i(\boldsymbol{r}), \qquad\qquad \varepsilon_1 \le \varepsilon_2 \le \cdots \le \varepsilon_N \overset{!}{(<} \varepsilon_{N+1}), \tag{1.5}$$

and the noninteracting ground-state density n_0 and energy E_0 are given by

$$n_0(\boldsymbol{r}) \;=\; \sum_{i=1}^{N} |\phi_i(\boldsymbol{r})|^2, \qquad\qquad E_0 \;=\; \sum_{i=1}^{N}\varepsilon_i. \tag{1.6}$$

The ground-state kinetic energy T_0 can be exactly evaluated by averaging (1.3)

$$E_0 \;=\; T_0 + \int v_0(\boldsymbol{r})\,n_0(\boldsymbol{r}), \tag{1.7}$$

where all terms except T_0 are now known. Recalling that both E_0 and $v_0(\boldsymbol{r})$ are, by the Hohenberg–Kohn theorem, functionals of $n_0(\boldsymbol{r})$, we see that also T_0 is such a functional.

Since E_0 is minimum at the noninteracting ground state, from (1.7) we get a soon-needed variational relation

$$\frac{\delta}{\delta n_0(\boldsymbol{r})}\left[E_0 - \mu_0\left(\int n_0 - N\right)\right] \;=\; 0 \quad\Rightarrow\quad \frac{\delta T_0}{\delta n_0(\boldsymbol{r})} + v_0(\boldsymbol{r}) - \mu_0 \;=\; 0, \tag{1.8}$$

with μ_0 being a Lagrange multiplier (chemical potential) taking care of the particle-number constraint.

Similarly, for E to be the minimum energy of the interacting system, by varying (1.2) we get

---

pair-correlation function, $E_{\text{xc}} = \int \frac{n(\boldsymbol{r})\,n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}[\tilde{g}(\boldsymbol{r},\boldsymbol{r}')-1]\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}'$, where $\tilde{g}(\boldsymbol{r},\boldsymbol{r}') = \int_0^1 g_\lambda(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}\lambda$ and g_λ is a pair-correlation function of a system in which i) the inter-particle interaction has been scaled with λ (i.e., $\lambda = 0$ is a system with noninteracting particles) and ii) the external potential has been concurrently changed in such a way to 'compensate' the scaling and ensure the system maintains the same ground-state density for any value of λ. The latter expression for E_{xc} is usually the starting point for approximations such as LDA. [Alternatively, one may also consult R. Zeller's contribution, A.1, chapter 3, in (Blüg06).]

¶ The separation-out of the Hartree part follows, of course, our intention to apply the Kohn–Sham scheme to electrons in solids, for which the Hartree interaction energy is a major and, importantly, exactly expressible contribution. [In the jellium model (which is a sensible model for the alkali metals) the Hartree energy is completely cancelled by the interaction energy of electrons with ions (smeared into a homogeneous background) and of ions with ions. Our model, however, will not be jellium, and the mentioned cancellation does not occur. A nice discussion about mutual compensation of several numerically fairly large energies, namely (Wigner's) 'boundary-condition' correction, kinetic, Hartree, exchange, correlation, appears in (Ande03), chapter 2.A.2.,3. To give a gist: In a sodium solid the listed energies are, respectively, −3.2, +2, +4.1, −3.1, −1 eV per electron, eventually leading to the cohesion energy of 1.2 eV per electron.] Further, the separation into the Hartree and exchange-correlation energies can be viewed as an approximate separation of long- and short-range effects.

$$\frac{\delta}{\delta n(\boldsymbol{r})}\left[E-\mu\left(\int n-N\right)\right]=0 \;\Rightarrow$$

$$\Rightarrow \frac{\delta T_0}{\delta n(\boldsymbol{r})}+\int\frac{2n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}'+v(\boldsymbol{r})+v_{\mathrm{xc}}(\boldsymbol{r})-\mu=0, \quad (1.9)$$

where $v_{\mathrm{xc}}(\boldsymbol{r})=\dfrac{\delta E_{\mathrm{xc}}}{\delta n(\boldsymbol{r})}$ is the exchange-correlation (local) potential.

If we now require that the noninteracting and interacting systems have the same ground-state density, $n(\boldsymbol{r})=n_0(\boldsymbol{r})$, we may use (1.8) to rewrite the condition (1.9):

$$v_0(\boldsymbol{r})=v(\boldsymbol{r})+\int\frac{2n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}'+v_{\mathrm{xc}}(\boldsymbol{r}), \quad (1.10)$$

where we have left out the immaterial constant term $(\mu_0-\mu)$ (potentials are defined up to a constant).‖

To summarize: If we have a density $n(\boldsymbol{r})$ such that after inserting it into (1.10) we get $v_0(\boldsymbol{r})$, for which we solve (1.5) and evaluate (1.6) only to obtain back the $n(\boldsymbol{r})$ we have started with, then we know that such $n(\boldsymbol{r})$ is a good candidate for an interacting-system ground-state density. For this program to be applicable, we need some (necessarily approximate) prescription for v_{xc} (or E_{xc}). One such will be discussed in the section on the local density approximation (LDA).

We note that with the knowledge of the sum of 'occupied' selfconsistent energies ε_i, E_0 in (1.6), we may express the approximate ground-state energy E_g of the interacting system as

$$E_g=E_0-U_{\mathrm{Hartree}}[n_g]+E_{\mathrm{xc}}[n_g]-\int\mathrm{d}\boldsymbol{r}\,n_g(\boldsymbol{r})\frac{\delta E_{\mathrm{xc}}}{\delta n_g(\boldsymbol{r})}; \quad (1.11)$$

$$\overset{\mathrm{LDA}}{=} E_0-U_{\mathrm{Hartree}}[n_g], \quad (1.12)$$

which can be derived by combining (1.2), (1.7) and (1.10), and using (1.13) for the LDA result.

The set of equations (1.10), (1.5) and (1.6) (with $n_0=n$) are called the *Kohn–Sham equations.* They are usually solved iteratively to reach self-consistency, starting from some sound n (or v_0) and using some suitable mixing algorithm (such as the Anderson (Ande65) or Broyden (Broy65) method; see also chapter 9 and appendix L of (Mart04)) to create a new n (or v_0) for the next iteration step.

It was noted that the interacting system may possibly have several ground states with different densities. Apparently, it is the choice of the starting n (or v_0) and also the character of the iteration that decide into which ground state (if any) the procedure will converge.

1.4. Local Density Approximation (LDA)

The previous section lead to the Kohn–Sham equations to which some approximation of the exchange-correlation (XC) functional has to be supplied.

For weakly inhomogeneous systems already Kohn and Sham (Kohn65) proposed the so-called *local density approximation* (LDA) in which

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n]=\int\mathrm{d}\boldsymbol{r}\,n(\boldsymbol{r})\varepsilon_{\mathrm{xc}}(n(\boldsymbol{r})), \quad (1.13)$$

where $\varepsilon_{\mathrm{xc}}(n)$ is the XC energy per particle of a homogeneous system of density n. This approximation as if replaces the original system with a 'piece-wise homogeneous' system. For a truly homogeneous system (1.13) is exact, and the function $\varepsilon_{\mathrm{xc}}(n)$ can be approximated (even for this simplest system no analytical formula can be given) by a quantum-mechanical many-body calculation. Often, the energy $\varepsilon_{\mathrm{xc}}$ is thought to have two contributions, the first originating from the exchange, the second from correlation, $\varepsilon_{\mathrm{xc}}=\varepsilon_{\mathrm{x}}+\varepsilon_{\mathrm{c}}$.* Then ε_{x} is obtained (defined) by the

‖ Alternatively viewed, μ_0 can just be made equal to μ, so that they cancel, by a constant shift $v_0 \to v_0+\mathrm{const}$.

* The author believes that this separation, together with the prescription for ε_{x}, (1.14), actually defines ε_{c} (as the missing part between well-defined $\varepsilon_{\mathrm{xc}}$ and ε_{x}).

Hartree–Fock method,

$$\varepsilon_x = -3/2\,(3/\pi)^{1/3} n^{1/3} \approx -\frac{0.916}{r_s}, \qquad r_s = \left(\frac{3}{4\pi n}\right)^{1/3}. \tag{1.14}$$

For the correlation part no analytic formula exists. The first quantitative estimate was given in the 1930s by Wigner,‡ followed by considerable work of others, including e.g. the RPA-based calculations of Hedin and Lundquists (Hedi69, Hedi71), and probably the most reliable parametrizations available today, based on quantum Monte Carlo simulations, are provided by Vosko, Wilk and Nusair (Vosk80), and Perdew and Zunger (Perd81).

To the surprise of many, the LDA approximation turned out to be more useful than originally expected, and can actually be used to obtain quantitatively correct description of atoms, molecules and solids. Valuable insight into the reasons for the LDA success was offered by Gunnarsson, Lundquist and Jonson, (Gunn76, Gunn79).

A generalization of the LDA approximation, the so-called local spin-density approximation (LSDA), which is more suited for spin-polarized systems and is used in our codes, will be now described.

1.5. Generalization of the Hohenberg–Kohn Theory to a Spin-Dependent External Potential, SDFT

The results of Hohenberg and Kohn can be generalized, as shown by U. von Barth and L. Hedin (Bart72), to a spin-dependent external potential by replacing the scalar external potential $v(\boldsymbol{r})$ by a spin-dependent potential $v_{\alpha\beta}(\boldsymbol{r})$ and replacing the scalar density $n(\boldsymbol{r})$ by a spin-density matrix $n_{\alpha\beta}(\boldsymbol{r})$ (α,β being indices in the spin space); i.e., the potential term is changed from $\int nv$ to $\sum_{\alpha\beta} \int n_{\alpha\beta} v_{\beta\alpha} = \mathrm{Tr} \int nv$.¶ Having as the basic variable the spin-density, thus extended theory is called a spin-density functional theory (SDFT). It is noteworthy, however, that in such a theory *there is generally no unique mapping between the ground state spin-density $n_{\alpha\beta}$ and the potential $v_{\alpha\beta}$.*‖ In other words, there is no analogue of the first Hohenberg–Kohn theorem. Thus, some of the ideas presented for the spin-independent external potential are not valid now. For example, the picture 1.1 does not hold and the Hamiltonian is not determined by the ground state $n_{\alpha\beta}(\boldsymbol{r})$.

Fortunately, however, the minimization scheme (1.1) (i.e. the constrained-search formulation of DFT by Levy and Lieb) goes through unmodified (but for adding the spin indices). And this is all we need to calculate the ground-state density and energy.§

---

‡ $\varepsilon_c = -\frac{0.88}{(r_s + 7.8)}$. This is a corrected formula, published in (Wign38), of an earlier result (Wign34).

¶ Under a spin-dependent external potential one naturally imagines an applied magnetic field. The magnetic field, however, couples both to the spin and the orbital movement. By making the potential spin-dependent we only account for the former interaction and if some further changes are not made, we ignore the latter. Theories doing so are termed 'spin-only' theories. We note, that in a noninteracting homogeneous electron gas, in the presence of a magnetic field, the contributions to the energy from spin polarization and orbital movement are of the same magnitude (Pauli vs. Landau susceptibility). Nonetheless, when interactions between electrons are included, the spin contribution (i.e., as if the Pauli susceptibility) may be considerably enhanced by exchange (the so-called Stoner enhancement S; e.g. $S(\mathrm{Pd}) \approx 10$, $S(\mathrm{TiBe}_2) \approx 50$)) providing some justification for the spin-only theories, esp. for ferromagnets. [If the need is, on the other hand, to properly account for coupling of magnetic field to the orbital movement, there are so-called Current Density Functional Theories (CDFTs), which can do so.]

‖ A nice example with explanation can be found in (Esch03), chapter 4.7. See also (Esch01) and (Cape01), and a responding comment of W. Kohn in (Kohn04). [It seems to the author that the lack of a unique mapping, at least in the example in Eschrig's book, is a consequence of the fact that the external potential couples only to spin degrees of freedom.]

§ The point to understand is, I believe, the following. Eq. (1.1) says plainly that if we know v_{ext} (denoted presently just by v), which is usually the case, we can find n_g and E_g by minimizing $E_{v_{\mathrm{ext}}}[n]$, i.e., by testing different densities—simple functions of position—and we do not need any (complicated) wavefunction. Whether v_{ext} is a functional of n_g is of secondary importance. (From (1.1) we know that E_g always is a functional of n_g *and* v_{ext}.) In the spin-independent case, point ii) after (1.1) indicates that if the there-written functional derivative exists, v_{ext} is really a functional of n_g and so is then E_g. But this, once we know v_{ext}, is only an unneeded bonus. And actually, the mentioned functional derivative must not always exist for spin-dependent potentials (otherwise $n^g_{\alpha\beta}$ would uniquely determine $v^{\mathrm{ext}}_{\alpha\beta}$, and we now know this does not always happen). [The sequence of arguments by Levy and Lieb is fundamentally different from the original Hohenberg and Kohn's, the latter building crucially upon the existence of the $n_g \to v_{\mathrm{ext}}$ mapping. Von Barth and Hedin,

The spin-density $n_{\alpha\beta}(\boldsymbol{r})$ is, generally, in the second-quantization language, the expectation value of the $\hat{\phi}^{+}_{\beta}(\boldsymbol{r})\hat{\phi}_{\alpha}(\boldsymbol{r})$ operator.¶ But since we will follow the Kohn–Sham scheme (and will never handle the many-particle wavefunction directly), we only need to express this quantity for the auxiliary (i.e. noninteracting) system. There it is

$$n_{\alpha\beta}(\boldsymbol{r}) \;=\; \sum_{i=1}^{N} \phi^{i}_{\alpha}(\boldsymbol{r})\,\phi^{i*}_{\beta}(\boldsymbol{r}), \tag{1.15}$$

or, in the matrix form and dropping the $\boldsymbol{r}$-dependence, $n = \sum_{i=1}^{N} \phi^{i}\phi^{i+}$; the sum runs over the occupied orbitals (we assume zero temperature, or, more precisely, that every orbital is either occupied or not). The matrix is clearly 2×2 and hermitian, hence it can be also written as†

$$n \;=\; \frac{1}{2}\left[\bar{n}1 \;+\; \boldsymbol{m}^{\mathrm{s}}\cdot\boldsymbol{\sigma}\right] \;=\; \frac{1}{2}\begin{bmatrix} \bar{n}+m^{\mathrm{s}}_{z} & m^{\mathrm{s}}_{x}-im^{\mathrm{s}}_{y} \\ m^{\mathrm{s}}_{x}+im^{\mathrm{s}}_{y} & \bar{n}-m^{\mathrm{s}}_{z} \end{bmatrix}, \quad \text{with } (\bar{n},\boldsymbol{m}^{\mathrm{s}}) \;=\; (\mathrm{Tr}\; n, \mathrm{Tr}\; n\boldsymbol{\sigma}). \tag{1.16}$$

Physically, $\bar{n}(\boldsymbol{r})$ represents the density of electrons around $\boldsymbol{r}$, and $\boldsymbol{m}^{\mathrm{s}}(\boldsymbol{r})$ the spin-moment density in units of $\hbar/2$ (or the magnetic-moment density $\boldsymbol{m}$ divided by $(-\mu_{\mathrm{B}})$, $\boldsymbol{m} = -g\mu_{B}\frac{\hbar}{2}\boldsymbol{m}^{\mathrm{s}} = -\mu_{B}\boldsymbol{m}^{\mathrm{s}}$; an electron's magnetic moment is directed opposite to its spin) originating from electrons' spins.

The spin-aware Kohn–Sham equations can be derived in the exactly same manner as for the spin-independent case, leading to

$$\sum_{\beta}\left[-\nabla^{2}\delta_{\alpha\beta} \;+\; v^{0}_{\alpha\beta}(\boldsymbol{r})\right]\,\phi^{i}_{\beta}(\boldsymbol{r}) \;=\; \varepsilon^{i}\phi^{i}_{\alpha}(\boldsymbol{r}), \tag{1.17}$$

$$v^{0}_{\alpha\beta}(\boldsymbol{r}) \;=\; v_{\alpha\beta}(\boldsymbol{r}) \;+\; \delta_{\alpha\beta}\int\frac{2\bar{n}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}' \;+\; v^{\mathrm{xc}}_{\alpha\beta}(\boldsymbol{r}), \quad v^{\mathrm{xc}}_{\alpha\beta}(\boldsymbol{r}) \;=\; \frac{\delta E_{\mathrm{xc}}[n_{\alpha\beta}]}{\delta n_{\alpha\beta}(\boldsymbol{r})}. \tag{1.18}$$

The equations (1.18), (1.17) and (1.15), supplemented by (1.16), are the generalized versions of (1.10), (1.5) and (1.6).

It is instructive to rewrite the last equation in a matrix form (as we did in (1.16)), showing explicitly the potential's i) spin-dependent part, coupling to a magnetic field composed of the externally applied $\boldsymbol{B}$ as well as the (internally generated) exchange field $\boldsymbol{B}_{\mathrm{xc}}$, and ii) spin-independent part (the terms below with a bar), again originating both from the outside, $\bar{v}$, and from the Hartree, $\bar{v}_{\mathrm{Hartree}}$, and the exchange, $\bar{v}_{\mathrm{xc}}$, terms:

$$v^{0} \;=\; \bar{v}^{0}1 \;+\; \mu_{\mathrm{B}}\,\boldsymbol{B}_{\mathrm{tot}}\cdot\boldsymbol{\sigma}, \qquad \bar{v}^{0} \;=\; \bar{v} \;+\; \bar{v}_{\mathrm{Hartree}} \;+\; \bar{v}_{\mathrm{xc}}, \qquad \boldsymbol{B}_{\mathrm{tot}} \;=\; \boldsymbol{B} \;+\; \boldsymbol{B}_{\mathrm{xc}}, \tag{1.19}$$

with

$$(\bar{v},\boldsymbol{B}) \;=\; ½(\mathrm{Tr}\; v, \frac{\mathrm{Tr}\; v\boldsymbol{\sigma}}{-\mu_{\mathrm{B}}}),$$

---

wanting to confront the spin-dependent potential (a decade before Levy and Lieb came up with their formulation of DFT), had to circumvent this possibly non-existing mapping. They showed, but only for a non-degenerate ground state, that there is a mapping directly to the ground state, $n_g \to \psi_g$. This implies that any quantity that can be calculated directly from ψ_g (such as e.g. the pair-correlation function) is a functional of n_g. More importantly it also implies that if we know v_{ext}, we can again only vary density (and not bother with wavefunctions) to find E_g: $<\psi_g|T+U|\psi_g>$ is a functional of $\psi_g[n_g]$, $<\psi_g|V_{\mathrm{ext}}|\psi_g> = \int n_g v_{\mathrm{ext}}$, and the ground-state energy is obtained by minimization of the sum over possible n_gs (note, however, there is the *v*-representability problem present here).]

¶ Here $\hat{\phi}_{\alpha}(\boldsymbol{r})$, $\hat{\phi}^{+}_{\beta}(\boldsymbol{r})$, is, respectively, a field operator that destroys a particle in the spin-state α at position $\boldsymbol{r}$, and creates a particle in the spin-state β at position $\boldsymbol{r}$. These operators should not be confused with the one-particle Kohn–Sham orbitals $\phi^{i}(\boldsymbol{r})$ further used in the text. Also note that the order of indices in $n_{\alpha\beta}$ is opposite to their order for the *operators*—then averages of one-particle operators can be written naturally as traces, see e.g. (1.16). (The order of orbital *functions* in (1.15) can, of course be changed at will.)

† Any 2×2 hermitian matrix A can be written as a linear combination of the unit matrix and the Pauli matrices (together they form a basis), $A = a_0 1 + a_i\sigma_i$. Taking the trace we get $\mathrm{Tr}\; A = 2a_0$, while multiplying the equation with σ_j and taking the trace we obtain $\mathrm{Tr}\; A\sigma_j = 2a_j$ (the Pauli matrices are traceless, and $\sigma_i\sigma_j = 1\delta_{ij} + i\varepsilon_{ijk}\sigma_k$).

$$\bar{v}_{\text{Hartree}} = \int \frac{2\bar{n}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}',$$

$$(\bar{v}_{\text{xc}}, \boldsymbol{B}_{\text{xc}}) = \tfrac{1}{2}(\operatorname{Tr} v_{\text{xc}}, \frac{\operatorname{Tr} v_{\text{xc}}\boldsymbol{\sigma}}{-\mu_{\text{B}}}). \tag{1.20}$$

From (1.19) and (1.16) one may easily check that the total contribution of the v^0 potential to the energy density of the auxiliary system is $\operatorname{Tr} nv^0 = \bar{v}^0\bar{n} + \mu_B \boldsymbol{B}_{\text{tot}}\cdot\boldsymbol{m}^{\text{s}}$, as expected.

1.6. Local Spin-Density Approximation (LSDA)

The spin-density matrix $n_{\alpha\beta}(\boldsymbol{r})$, being hermitian, can be diagonalized at any point $\boldsymbol{r}$ by an application of (unitary, $\boldsymbol{r}$-dependent) spin-rotation matrices; i.e., locally we may choose coordinates so that the spin-moment density $\boldsymbol{m}$ is oriented along the (local) z-axis, $n_{\alpha\beta}(\boldsymbol{r}) \to \operatorname{diag}[n{\uparrow}(\boldsymbol{r}), n{\downarrow}(\boldsymbol{r})]$. When $n_{\alpha\beta}(\boldsymbol{r})$ changes only slowly in space, useful approximations for the XC-energy functional can be given in terms of $n{\uparrow}(\boldsymbol{r})$ and $n{\downarrow}(\boldsymbol{r})$. Such an approach, being the spin-dependent extension of LDA, is called the local spin-density approximation (LSDA).

Often, the following combinations of $n{\uparrow}(\boldsymbol{r})$ and $n{\downarrow}(\boldsymbol{r})$ are used (we shall re-use the letter n for what has been $\bar{n}$, hopefully causing no confusion):

$$\begin{aligned} n &= n{\uparrow} + n{\downarrow} \\ \zeta &= (n{\uparrow} - n{\downarrow})\,/\,n \end{aligned} \quad\Leftrightarrow\quad \begin{aligned} n{\uparrow} &= \tfrac{1}{2}n(1+\zeta) \\ n{\downarrow} &= \tfrac{1}{2}n(1-\zeta)\,. \end{aligned} \tag{1.21}$$

The exchange part of the XC-energy, defined by the Hartree–Fock approximation, is then given by

$$E_{\text{x}} = \int \mathrm{d}\boldsymbol{r}\,(\varepsilon_{\text{x}}^{\uparrow} n{\uparrow} + \varepsilon_{\text{x}}^{\downarrow} n{\downarrow}) = \int \mathrm{d}\boldsymbol{r}\,\tfrac{1}{2}n\left[(1+\zeta)\varepsilon_{\text{x}}^{\uparrow} + (1-\zeta)\varepsilon_{\text{x}}^{\downarrow}\right], \tag{1.22}$$

with the two spin polarizations contributing independently. Generalizing (1.14) to the spin-dependent case, $\varepsilon_{\text{x}}^{\sigma} = -3/2\,(3/\pi)^{1/3}(2n_\sigma)^{1/3}$, $\sigma = \uparrow, \downarrow$,* we get

$$E_{\text{x}} = \int \mathrm{d}\boldsymbol{r}\; n(\boldsymbol{r})\;\varepsilon_{\text{x}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r})), \tag{1.23}$$

where

$$\varepsilon_{\text{x}}(n,\zeta) = (-3/4)(3/\pi)^{1/3} n^{1/3}\left[(1+\zeta)^{4/3} + (1-\zeta)^{4/3}\right]. \tag{1.24}$$

In the literature this expression is often rewritten in another form. Writing for a moment $\varepsilon_{\text{x}}(n,\zeta) = f(n)g(\zeta)$, we derive $f(n)g(\zeta) = f(n)g(0) + f(n)[g(1)-g(0)]\chi(\zeta)$ with $\chi(\zeta) = \dfrac{-g(0)+g(\zeta)}{g(1)-g(0)}$, and hence, denoting $f(n)g(0) = \varepsilon_{\text{x}}^{\text{P}}(n)$ and $f(n)g(1) = \varepsilon_{\text{x}}^{\text{F}}(n)$,

$$\varepsilon_{\text{x}}(n,\zeta) = \varepsilon_{\text{x}}^{\text{P}}(n) + [\varepsilon_{\text{x}}^{\text{F}}(n) - \varepsilon_{\text{x}}^{\text{P}}(n)]\chi(\zeta), \tag{1.25}$$

$$\varepsilon_{\text{x}}^{\text{P}}(n) = (-3/2)(3/\pi)^{1/3} n^{1/3}, \quad \varepsilon_{\text{x}}^{\text{F}}(n) = 2^{1/3}\varepsilon_{\text{x}}^{\text{P}}(n), \tag{1.26}$$

with

$$\chi(\zeta) = \frac{1}{2}\,\frac{(1+\zeta)^{4/3} + (1-\zeta)^{4/3} - 2}{2^{1/3}-1}. \tag{1.27}$$

The superscripts F and P stand for ‘ferromagnetic’ (or fully polarized, $\zeta = 1$) and ‘paramagnetic’ (unpolarized, $\zeta = 0$) case. Apparently, the formula has a form of interpolation between the two extremes.

The situation with the correlation part of the XC-energy is less lucid. We have already mentioned, in the section on LDA, works by Vosko, Wilk and Nusair (VWN) (Vosk80), and Perdew and Zunger (PZ) (Perd81). Both articles propose formulæ also for the spin-polarized situation.

* Note the factor of 2 in front of n_σ. In the non-polarized case, the average contribution of spin-up and spin-down electrons is the same and equal to the average contribution per any electron, $\varepsilon_{\text{x}}^{\uparrow} = \varepsilon_{\text{x}}^{\downarrow} = \varepsilon_{\text{x}}$. But then $n = 2n_\sigma$, for both σ.

Again, we write

$$E_c = \int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r})\, \varepsilon_c(n(\boldsymbol{r}),\zeta(\boldsymbol{r})). \tag{1.28}$$

The PZ article uses the same ζ-dependence of ε_c as is used for the ε_x part, (1.25),‡

$$\varepsilon_c(n,\zeta) = \varepsilon_c^{\mathrm{P}}(n) + [\varepsilon_c^{\mathrm{F}}(n) - \varepsilon_c^{\mathrm{P}}(n)]\chi(\zeta), \tag{1.29}$$

put forward originally by von Barth and Hedin (Bart72). The dependence used by VWN is even more complicated. Anyhow, both parametrizations faithfully copy the results of quantum Monte Carlo simulations, e.g. (Orti94), and are widely used.

In our calculations we used the VWN parametrizations in all cases. See chapter 10.8.1 of (Ture97) for the in-codes implemented formulæ.

Having the LDA and LSDA theories, an interesting question arises, as to which one to use in the case when the external potential is not spin-dependent, i.e., when there is not magnetic field applied from outside. Since we rely on approximations of the XC functionals and these approximations are limited, it shows generally advantageous to use the spin version. For example, Gunnarsson, Lunquist and Wilkins in (Gunn74) show that while the LDA method gives the energy of a hydrogen atom with a 10% error, LSDA method is only 1.6% off. They generally argue that for applications with unpaired valence electrons spin-density formalism should be used (since the LDA theory views the electrons only as a spin-compensated electron liquid). More advocacy of the usefulness of the spin-density theories is presented in (Gunn76), where the authors say that the spin-density formalism offers greater flexibility allowing to build more of the actual physics into the approximate functionals as well as to simplify the description of open electron shells of atoms and thus provide a basis for Hund's rules.

1.7. Systems with Collinear Magnetism

Magnetic systems investigated in this thesis are *supposed* to feature only the so-called collinear magnetism. This means that anywhere in the system the spin-moment density $\boldsymbol{m}^{\mathrm{s}}$ as well as the exchange field $\boldsymbol{B}_{\mathrm{xc}}$ (and generally also the external field $\boldsymbol{B}$, which we do not explicitly consider†) all point in the same direction.

Since $\boldsymbol{m}^{\mathrm{s}}$ points everywhere in the same direction, there is a common quantization axis to all places with respect to which we can describe the system using the above introduced $n_\uparrow(\boldsymbol{r})$ and $n_\downarrow(\boldsymbol{r})$.

Furthermore, thanks to the assumed collinearity the Kohn–Sham equations (1.17), (1.18) can be simplified—the equations for the spin-up and spin-down components decouple:

$$[-\nabla^2 + v_\sigma^0(\boldsymbol{r})]\, \phi_\sigma^i(\boldsymbol{r}) = \varepsilon^i \phi_\sigma^i(\boldsymbol{r}), \quad \sigma = \uparrow,\downarrow, \tag{1.30}$$

$$v_\sigma^0(\boldsymbol{r}) = v_\sigma(\boldsymbol{r}) + \int \frac{2\bar{n}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}' + v_\sigma^{\mathrm{xc}}(\boldsymbol{r}), \quad v_\sigma^{\mathrm{xc}}(\boldsymbol{r}) = \frac{\delta E_{\mathrm{xc}}[n_\uparrow,n_\downarrow]}{\delta n_\sigma(\boldsymbol{r})}. \tag{1.31}$$

Relativistic counterparts of these equations, discussed below, were used for the calculations of this thesis.

---

‡ The author does not see any compelling argument for such a dependence other than that it describes the situation well enough.

† In our considerations we think of the external field as only of an agent used to fix the magnetization direction of the magnetic parts of the sample. Had it been not for magnetic anisotropy (a spin-orbit-interaction effect), the field could be infinitesimally small. With the anisotropy, the field has to be finite to keep the magnetization close to the wanted direction (unless this direction is by chance an easy magnetization axis). We suppose that effects other than fixing the magnetization direction originating from this finite external field can be neglected (in much the same sense as we generally neglect effects of *any* real magnetic field in the spin-only theories). Furthemore, when the magnetization does not point along an easy axis (due to the external field), we neglect the difference between directions of magnetization and the exchange field.

1.8. General Trends in (nonrelativistic) L(S)DA Calculations

Electron density in molecules and solids is generally close to a superposition of atomic densities, which is far from uniform. Therefore, the use of local approximations was initially treated with great reserve. Nonetheless, the approximations do almost always lead to correct picture of binding trends. Structure, bond lengths, vibration frequencies in many systems are satisfactorily reproduced, and deviations from measurements of these and other quantities are often systematic. On the other hand, binding energies of molecules and cohesive energies of solids are usually overestimated, strongly-correlated systems, such as 3d-transition metal oxides are not described well by LSDA, which predicts them to be metallic, the barriers for chemical reactions and band gaps of materials are underestimated, etc.

Here we will not try to speculate about reasons why some predicted quantities are fairly right why others rather off. We delegate the interested reader to dedicated literature which provides more details: (Bart94, Jone06), and from a slightly more general perspective (Cohe08).

Finally, we note that besides the above described L(S)DA approximations to DFT, there is also a large amount of other related approximations which try to eliminate some of the L(S)DA deficiencies. We just name the mostly known Generalized-Gradient Approximation (GGA), but others can be found e.g. in Chapter 8 of (Mart04).

2

Relativistic Issues

(In this chapter we use $\hbar = c = \mu_0 = \varepsilon_0 = 1$, as is usual in quantum field theories.)

Before plunging into a few chosen aspects of the relativistic theory we wish to make several comments.

The fundamental approach to the relativistic electronic-structure theory is based on quantum electrodynamics (QED). QED, although basically describing only electrons, positrons, photons, and their mutual interaction, is a complex theory: it contains particle creation and annihilation, vacuum polarization, radiative corrections and other effects, which when not properly handled—by introduction of renormalization schemes—lead to divergencies. Grasping all the involved logic of QED and correctly applying it to the subject of our current interest, i.e. calculation of electronic structure of solids, is beyond what the author is presently capable of. The reader is referred to, e.g., relevant chapters in (Enge11, Esch03, Drei90) for details. Fortunately, relativity brings about several effects, of differing importance, and the ones which would force us into consideration of the 'difficult' aspects of QED, like vacuum polarization, vacuum fluctuation and radiative corrections, are those less important, at least for valence electrons. Hence, we generally work in the so-called no-sea approximation, in which all just listed effects are ignored. On the other hand, we will concentrate on what seem to be dominant relativistic corrections for the case of solids, effects contained already in *one*-particle Dirac Hamiltonian: relativistic mass-enhancement, Darwin term, and spin-orbit (SO) interaction.

A comprehensive overview of relativistic effects in structural chemistry is given in (Pyyk88). A rather important idea, also presented there, is that although one would expect the relativistic effects to be most important for core electrons (due to their high speeds), the effect on valence electrons may also be large. This is mainly because valence-electrons' wavefunctions have to be orthogonal to the core electrons, hence an effect on a core electron 'propagates' to higher electrons, too (an example of the so-called indirect relativistic effect).

Another set of examples showing how relativistic effects affect optical properties, Fermi-surface properties, and cohesive properties of solids can be found in (Chri84). A review of fully relativistic band-structure calculations for magnetic solids is then given in (Eber99).

The main goal of the following sections is to provide some justification for the so-called Kohn−Sham−Dirac Hamiltonian, (2.41), which includes certain relativistic effects and is used in our calculations in place of the above-considered non-relativistic, Schrödinger-like Hamiltonian appearing in (1.30).

2.1. Form of Interaction between Electrons and Electromagnetic Field

Our current knowledge derives the form of interaction between fields from the principle of *local gauge invariance.*‖ Free electronic (four-component) matter field ψ is described by the

‖ We wish to note that this principle really dictates the *form* of the interaction. Often, in electromagnetism, one proceeds by adding the so-called minimum-coupling interaction to a free (matter and massless vector) field Lagrangian and after-

Lagrangian density§†*

$$L^0 = \bar{\psi}(i\partial\!\!\!/ - m)\psi, \tag{2.1}$$

(in short: $\bar{\psi}=\psi^\dagger\gamma_0$; $\partial\!\!\!/=\gamma^\mu\partial_\mu$; $+---$; for a detailed explanation of the notation see any book on relativistic quantum mechanics) which is invariant under a global (phase) transformation $\psi(x)\to e^{i\varepsilon}\psi(x)$, ($\bar{\psi}\to e^{-i\varepsilon}\bar{\psi}$), i.e. infinitesimally under $\psi\to\psi+i\varepsilon\psi$. According e.g. to S. Weinberg (Wein95 ch. 7.3, mainly equations 7.3.1,15) such a symmetry leads to a conserved current:

> A Lagrangian density $L(x)$—a function of fields $\psi^l(x)$ and their derivatives $\partial_\mu\psi^l(x)$—which is invariant under an infinitesimal transformation $\psi^l(x)\to\psi^l(x)+i\varepsilon F^l(x)$ leads to $j^\mu=-i\dfrac{\partial L}{\partial(\partial\psi^l/\partial x^\mu)}F^l$, which is conserved, $\partial_\mu j^\mu=0$.¶ (2.2)

Applied to our case, ψ^l are $\{\psi,\bar{\psi}\}$, with the latter not contributing to j^μ thanks to non-existent $\partial_\mu\bar{\psi}$ in L^0,

$$j^\mu = -i\frac{\partial L^0}{\partial(\partial\psi/\partial x^\mu)}\psi = \bar{\psi}\gamma^\mu\psi, \qquad \partial_\mu j^\mu = 0. \tag{2.3}$$

Now, as the requirement is imposed that the Lagrangian density be invariant even for arbitrary space-time−non-constant $\varepsilon(x)$ (this is the local-gauge-invariance principle), the above L^0 is not invariant, but produces an extra term

$$\delta L^0 = i\bar{\psi}\left[e^{-i\varepsilon(x)}\partial\!\!\!/\, e^{i\varepsilon(x)}\right]\psi = -\bar{\psi}\gamma^\mu\psi\partial_\mu\varepsilon(x). \tag{2.4}$$

Devising an *interaction* term with a new vector field $\hat{A}_\mu$‖ of the form** (g is a coupling constant)

$$L^{\text{int}} = g\bar{\Psi}\gamma^\mu\Psi\hat{A}_\mu, \tag{2.5}$$

with the $\hat{A}_\mu$ transforming as

$$\hat{A}_\mu \to \hat{A}_\mu + \frac{1}{g}\partial_\mu\varepsilon(x) \tag{2.6}$$

then makes the extended Lagrangian density L^0+L^{int} invariant for any local gauge transformation. Adding the usual kinetic term for a vector field, L^{elm}, we arrive at the local-gauge−invariant Lagrangian density of quantum electrodynamics§§

---

wards shows that this leads to the known equation of motion, i.e. in our case to the Lorentz-force law, which is as if primordial. The local gauge-invariance principle, on the contrary, can be considered as a more fundamental principle explaining why the Lorentz force has the structure it has.

§ Some people prefer a more symmetric density $\bar{\psi}(i\overleftrightarrow{\partial\!\!\!/}/2-m)\psi$, with $\overleftrightarrow{\partial} = \overrightarrow{\partial} - \overleftarrow{\partial}$. However, the two densities differ by a four-divergence and thus, if the fields vanish at infinity, they are equivalent. Our choice seems simpler for the time.

† In the following text we will freely interchange Lagrangian and Lagrangian density, as well as Hamiltonian and Hamiltonian density, four-current and four-current density. It should always be clear from the context whether a density or its space integral is meant.

* We wish to stress that we now use the second-quantized formalism, in which ψ is a field operator, not a wavefunction. Then, the Lagrangian density does not describe just one electron, but any number of them. Later, when interaction with photons is introduced, the number of particles can even change in time. It is only in the next section, eq. (2.16), that we return to an operator acting on a single-particle space.

¶ Sure, any multiple of j^μ is conserved. However, the $(-i)$ makes j^μ (i) Hermitian and (ii) the meaning of $j^0=\psi^\dagger\psi$, which can be then interpreted directly as the particle-density operator.

‖ The hat above A is introduced to remind us that this field is the dynamical field produced within the system by its charges, as opposed to a further considered externally applied field, A_μ^{ext}, and the total field plainly denoted with $A_\mu\equiv\hat{A}_\mu+A_\mu^{\text{ext}}$.

** We now use capital Ψ, $\hat{A}^\mu$, J^μ instead of ψ, a^μ, j^μ to distinguish interacting fields from noninteracting—they obviously have different equations of motion.

§§ The coefficients like the ¼, $g=-q$ have been fixed so as to reproduce the standard Maxwell equations. The equations of motion inferred from our Lagrangian density are $(i\mathcal{D}\!\!\!/-m)\Psi=0$, i.e. the Dirac equation in electromagnetic field, and $\partial_\mu\hat{F}^{\mu\nu}=qJ^\nu$, i.e. the inhomogeneous Maxwell equations. Note that many signs in our equations follow from the $+---$ convention.

$$L^{\mathrm{QED}} = L^{\mathrm{elm}} + L^{0} + L^{\mathrm{int}} = -\frac{1}{4}\hat{F}_{\mu\nu}\hat{F}^{\mu\nu} + \overline{\Psi}(i\partial\!\!\!/ - m)\Psi - q\overline{\Psi}\gamma^{\mu}\Psi\hat{A}_{\mu}, \tag{2.7}$$

where the electromagnetic field tensor $\hat{F}_{\mu\nu} = \partial_{\mu}\hat{A}_{\nu} - \partial_{\nu}\hat{A}_{\mu}$ is manifestly invariant under the local transformation. Introducing a covariant‡ derivative $D_{\mu} = \partial_{\mu} - ig\hat{A}_{\mu} = \partial_{\mu} + iq\hat{A}_{\mu}$, we rewrite

$$L^{\mathrm{QED}} = -\frac{1}{4}\hat{F}_{\mu\nu}\hat{F}^{\mu\nu} + \overline{\Psi}(iD\!\!\!/ - m)\Psi. \tag{2.8}$$

(We see that the 'switching on' of the interaction between the electron and electromagnetic field effectively boils down to replacing a normal derivative by a covariant one, $i\partial^{\mu} \rightarrow (i\partial^{\mu} - q\hat{A}^{\mu})$, which justifies the usual minimal-coupling prescription $\boldsymbol{p} \rightarrow \boldsymbol{p} - q\hat{\boldsymbol{A}} \wedge H \rightarrow H - q\phi$ when one switches to the Hamiltonian formalism.)

Because $\hat{A}^{\mu}$ does not change under the *global* gauge transformation above and because $\hat{A}^{\mu}$ does not couple to the derivatives of Ψ ($\overline{\Psi}$) in (2.7), the conserved current we obtain from the full L^{QED} has the same form as before. Namely,

$$J^{\mu} = \overline{\Psi}\gamma^{\mu}\Psi, \qquad \partial_{\mu}J^{\mu} = 0, \tag{2.9}$$

and the interaction has the form of the conserving four-current times the four-potential

$$L^{\mathrm{int}} = -qJ^{\mu}\hat{A}_{\mu}. \tag{2.10}$$

The L^{QED} Lagrangian (2.7) describes interacting electrons through the field they themselves create. For further application we ask in what way we must change the Lagrangian to also describe interaction with an *externally* applied electromagnetic field. Since electrons do not distinguish external sources from internal ones, one is lead to the identification that $\hat{A}_{\mu}$ *in (2.10) should become the sum* $\hat{A}_{\mu} \rightarrow \hat{A}_{\mu} + A_{\mu}^{\mathrm{ext}}$, where the first term is the field from the system's electrons, while the second describes the external field. One may further wonder, whether also the kinetic term $\propto \hat{F}_{\mu\nu}\hat{F}^{\mu\nu}$ should be changed to include the external field. As we think of A_{μ}^{ext} to be controlled from outside irrespectively of what is happening inside the system, this external field is more a parameter of the Lagrangian and is not a dynamical variable whose time evolution would be described by the Lagrangian. Thus we do not include any such contribution from the external field.

Since L^{int} does not contain time derivatives, proceeding from the Lagrangian to Hamiltonian formalism is fairly straightforward: to Hamiltonian densities of the free fields (arising from L^{0} and L^{elm}) we need add the interaction-Hamiltonian density, which is simply the interaction-Lagrangian density with its sign changed:*¶

$$H^{0} = \Psi^{+}(\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta m)\Psi, \qquad \text{where} \quad \alpha^{i} \equiv \gamma^{0}\gamma^{i}, \tag{2.11}$$

$$H^{\mathrm{elm}} = \tfrac{1}{2}(\hat{\boldsymbol{B}}^{2} + \hat{\boldsymbol{E}}^{2}) - \rho\hat{A}_{0}, \qquad \text{where} \quad \hat{\boldsymbol{B}} \equiv \nabla\times\hat{\boldsymbol{A}}, \quad \hat{\boldsymbol{E}} \equiv -\nabla\hat{A}^{0} - \partial_{0}\hat{\boldsymbol{A}}, \tag{2.12}$$

$$H^{\mathrm{int}} = qJ^{\mu}A_{\mu} \equiv qJ^{\mu}(\hat{A}_{\mu} + A_{\mu}^{\mathrm{ext}}). \tag{2.13}$$

(Obtaining the first and the last expression is easy; the second is slightly more complicated and we refer the reader to the derivation given in (Saku67, section 1.4 and appendix A)). Writing the terms together yields†

---

‡ Covariant because it transforms as $D_{\mu} \rightarrow e^{i\varepsilon(x)}\, D_{\mu}\, e^{-i\varepsilon(x)}$.

* It is interesting to compare this to a classical movement of a charged particle in an electromagnetic field (or also to an evolution of a charged quantum Klein-Gordon particle-field), where the interaction naturally contains a term with velocity, $q\boldsymbol{v}\cdot\boldsymbol{A}$, and the transition from the Lagrangian to Hamiltonian formalism is more complicated.

¶ The author finds it interesting that one apparently need maintain the second term in (2.12), even though it is really zero for a truly free-field case ($\rho = 0$)—when there are charges present, this term suddenly does contribute. One might, on the other hand, perhaps naturally (though mistakenly) expect that all contributions related to the charge–elm-field constalation would originate in the interaction term.

† Here presented Hamiltonian is the same as found in (Esch85, eq. 15). For a reason not quite understood, the same author later, in (Esch03, eqs. 9.2–4), presents a similar, though not equivalent Hamiltonian (there is an additional $J^{0}\hat{A}_{0}$ term, which in our case gets subtracted by the second term in (2.12)). The author of this thesis then tends to think the expression in (Esch03) might not be quite right.

$$\begin{aligned} H &= H^0 + H^{\mathrm{elm}} + H^{\mathrm{int}} = \\ &= \Psi^+[\boldsymbol{\alpha}\cdot(\boldsymbol{p}-q\hat{\boldsymbol{A}}) + \beta m]\Psi + \tfrac{1}{2}(\hat{\boldsymbol{B}}^2+\hat{\boldsymbol{E}}^2) + qJ^\mu A^{\mathrm{ext}}_\mu = \\ &\equiv H_{\mathrm{system}} + qJ^\mu A^{\mathrm{ext}}_\mu . \end{aligned} \tag{2.14}$$

[In the non-relativistic case, and putting the interaction with the outside aside, we had just electrons interacting via instantaneous Coulomb interaction. Now the interaction is mediated by the electromagnetic field. Nonetheless, it is easy to show, see e.g. (Cohe97, section I.B.4), that the instantaneous Coulomb interaction is, in our current situation, embodied in the so-called longitudinal part of the electric field (usually denoted by $\hat{\boldsymbol{E}}_{\parallel}$), and is often separated out. The rest of the field (having only transverse components and hence called the transverse field) provides the field for the magnetic interaction among moving electrons and also ensures that the instantaneous Coulomb interaction is corrected to eventually become retarded, in accordance with the finite speed of light.]

2.2. Relativistic Density Functional Theory (RDFT)

Having divided the Hamiltonian (2.14) into a system's part and interaction with the outside, one may ask the 'DFT questions'. In section 1.2 we showed, with a non-relativistic system in mind and interaction of the form $\int n(\boldsymbol{r})v^{\mathrm{ext}}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r}$, that the ground-state energy is a functional of the ground-state density n_g. In section 1.5 we generalized to a spin-dependent external potential $v^{\mathrm{ext}}_{\alpha\beta}$ (which coupled only to spin degrees of freedom) and argued that knowledge of $v^{\mathrm{ext}}_{\alpha\beta}$ and $n^g_{\alpha\beta}$ again determines the ground-state energy. Neither of these considerations actually relied on the form of the system's Hamiltonian, solely on the form of the interaction with the outside. Thus, for example, if A^{ext}_μ in (2.14) has only the 0-component nonzero, we may claim that the ground-state density $n_g = J^0_g$§ determines the ground-state energy of H. In a general case, it is argued e.g. in (Stra98, section 10.14)‡ that the ground-state energy is a functional of the expectation value of the full four-current, i.e., $E_g = E_g[J^\mu]$; we have here an example of a current-DFT (CDFT).

Similarly as we did in section 1.3, Kohn–Sham equations related to the relativistic Hamiltonian (2.14) can be set up: we use an auxiliary non-interacting system derived from (2.14) by switching all 'internal' interactions off, i.e., $\hat{\boldsymbol{A}} = \hat{\boldsymbol{E}} = \hat{\boldsymbol{B}} = 0$, and 'replace' their effect and the effect of the original external field $A^{\mu;\mathrm{ext}}$ by an effective external potential $A^{\mu;\mathrm{eff}} = (V^{\mathrm{eff}}, \boldsymbol{A}^{\mathrm{eff}})$; $H \to \Psi^+[\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta m]\Psi + qJ^\mu A^{\mathrm{eff}}_\mu$.¶ Thus we arrive at a *single*-particle Hamiltonian for the auxiliary system

$$H_0 = \boldsymbol{\alpha}\cdot[\boldsymbol{p} - q\boldsymbol{A}^{\mathrm{eff}}(\boldsymbol{r})] + \beta m + qV^{\mathrm{eff}}(\boldsymbol{r}), \tag{2.16}$$

to which we must add some prescription for $A^{\mu;\mathrm{eff}}$ that would 'connect' the non-interacting system to the interacting one, much in the sense discussed in section 1.3—now so that the expectation value of the four-current is the same in the interacting and the auxiliary system groundstates—, but with two differences: i) now the potential has four components, $A^{\mu;\mathrm{eff}}$, instead of just one v_0, and ii) the participating functionals are functionals of the four-current J^μ instead of just the particle density. It can be argued‖ that a sensible form of the effective potential, generalization of (1.10), is*

---

§ In the text we use J^μ to denote both the operator and its expectation value. From the context it should be clear what is meant at each time.

‡ The original related papers are, e.g., (Raja73), (MacD79), (Vign88).

¶ In section 1.3 we used index 0 (in v_0) to denote the effective field. There this index was suitable since it also reminded us of the fact that the potential is used in otherwise non-interacting system. Here such an index could be mistaken for a 0-component of some four-vector, hence we opt for the use of 'eff' instead.

‖ More details can be found in the referred book by P. Strange.

* Note that this potential is to be coupled with a *charge* density (and current), while v_0 in (1.10) was defined to couple with a *particle* density. I.e., v_0 should be compared with $qA^{0;\mathrm{eff}}$ with the factor of charge (q) present; then, in the atomic units, $q^2 = 2$ for an electron, and this is the factor appearing in the second term of (1.10). Lastly, as far as the second term is concerned, one should not miss that its space components really represent a classical (static) contribution from a current to the vector potential (while its time component, as we already know, is the classical Coulomb contribution).

$$A^{\mu;\mathrm{eff}}(\boldsymbol{r}) \;=\; A^{\mu;\mathrm{ext}}(\boldsymbol{r}) \;+\; q\int \frac{J^{\mu}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}' \;+\; \frac{\delta E_{\mathrm{xc}}[J]}{\delta J_{\mu}(\boldsymbol{r})}\,. \tag{2.17}$$

For a given $A^{\mu;\mathrm{eff}}$ we find eigenfunctions ψ_i and eigenenergies E_i of the Hamiltonian H_0, 'fill' the N lowest electron-like (of positive energy) states, where N is the number of electrons in our system, evaluate the current from

$$J^{\mu} \;=\; \sum_{i=1}^{N} \psi_i^{+}\gamma^0\gamma^{\mu}\psi_i\,, \tag{2.18}$$

and with some prescription for $E_{\mathrm{xc}}[J]$ we may recalculate $A^{\mu;\mathrm{eff}}$ from (2.17). If we find a selfconsistent solution of (2.16), (2.17), (2.18), the related J_{μ} is a candidate for the ground state four-current density; the indicated equations are the so-called *Kohn–Sham–Dirac equations*. To estimate the ground-state energy, there is an analogue of (1.11)†

$$E_g \;=\; \sum_{i=1}^{N} E_i \;-\; \tfrac{1}{2}q^2\int \mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}'\;\frac{J_g^{\mu}J_{g;\mu}}{|\boldsymbol{r}-\boldsymbol{r}'|} \;+\; E_{\mathrm{xc}}[J_g] \;-\; \int \mathrm{d}\boldsymbol{r}\;J_g^{\mu}(\boldsymbol{r})\,\frac{\delta E_{\mathrm{xc}}[J_g]}{\delta J_g^{\mu}(\boldsymbol{r})}\,. \tag{2.19}$$

The main problem of the just-described approach lies in a missing suitable expression for the relativistic exchange-correlation functional $E_{\mathrm{xc}}[J]$. We will now make an approximation of the full theory, following (Raja73) and (MacD79), and will derive equations that look more like the non-relativistic (spin-only) density-functional expressions for magnetic systems. This will allow us to use—as an approximation—non-relativistic exchange-correlation functionals, which were discussed in section 1.6 (the use of these functionals is hardly justifiable rigorously, nonetheless is common).

2.3. Different Currents, an Approximate RDFT

In the last but one section we showed that the interaction between a charged spin-½ particle (e.g. an electron) with electromagnetic field has the form $qJ^{\mu}A_{\mu}$; particularly, the interaction of a system of electrons with an external field is described by the operator $qJ^{\mu}A_{\mu}^{\mathrm{ext}}$, the second term on the last line of (2.14). Despite its simple look, the particle-current density $J^{\mu}=\overline{\Psi}\gamma^{\mu}\Psi$ contains both a contribution from the 'ordinary' particle movement in space and a contribution of a 'quantum' origin related to the particle spin. The derivation of the previous section showed that it is really this total current that is to be considered when electromagnetic interaction is discussed, not just some (usually heuristically deduced) part of it (like just a paramagnetic, diamagnetic, spin part, or a combination; the matter may be obscured by the fact that the full current can be split into parts that are separately conserved, see below). We will now show the connection between this result and some other used expressions for the current operator. To this end the Gordon decomposition§ will be deployed.

From the Dirac equation

$$(i\not{D}-m)\Psi \;=\; 0 \qquad (D_{\mu} \;=\; \partial_{\mu}+iqA_{\mu}) \tag{2.20}$$

we get

$$\Psi \;=\; \frac{i}{m}\overrightarrow{\not{D}}\Psi \quad \text{and} \quad \overline{\Psi} \;=\; -\frac{i}{m}\overline{\Psi}\overleftarrow{\not{D}^{*}}, \tag{2.21}$$

where $\not{D}^{*}=\not{\partial}-iq\not{A}$.‡* Then

$$J^{\mu} \;=\; \overline{\Psi}\gamma^{\mu}\Psi \;=\; \frac{1}{2}(\overline{\Psi}\gamma^{\mu}\Psi \;+\; \overline{\Psi}\gamma^{\mu}\Psi) \;=\; \frac{i}{2m}(\overline{\Psi}\gamma^{\mu}\overrightarrow{\not{D}}\Psi \;-\; \overline{\Psi}\overleftarrow{\not{D}^{*}}\gamma^{\mu}\Psi). \tag{2.22}$$

† This is essentially eq. (10.249) in (Stra98).

§ Named after its discoverer W. Gordon (Gord28).

‡ Classically, A_{μ} is a real field. When quantized, it becomes a hermitian-operator–valued field. In any case, it is sufficient to just write D_{μ}^{*} as compared to D_{μ}^{+}. (They are equal.)

* We note that there is the *full* electromagnetic potential A_{μ} present in the covariant derivative; the electrons move under the influence of both the internal as well as the external field.

Now we decompose J^μ into two parts, $J^\mu_{(1)}$ and $J^\mu_{(2)}$, according to whether the μ index coincides with the summed-over index within $\not{D}$s (such a coincidence leads immediately to disappearance of the γs between $\bar\Psi$ and Ψ, thus one as if separates the contributions that mix the internal particle components and those which do not).

$$J^\mu_{(1)} = \frac{i}{2m}\bar\Psi\left[\gamma^\mu\gamma^\mu\overrightarrow{D_\mu} - \overleftarrow{D^*_\mu}\gamma^\mu\gamma^\mu\right]\Psi \qquad \text{(no summation over } \mu\text{)}. \tag{2.23}$$

Using $(\gamma^\mu)^2 = g_{\mu\mu} = \mathrm{diag}(+---)$ and $g_{\mu\mu}X_\mu = X^\mu$ (no summation)

$$\begin{aligned} J^\mu_{(1)} &= \frac{i}{2m}\bar\Psi\left[\overrightarrow{D^\mu} - \overleftarrow{D^{\mu *}}\right]\Psi = \\ &= \frac{i}{2m}\bar\Psi\left[\overrightarrow{\partial^\mu} - \overleftarrow{\partial^\mu}\right]\Psi - \frac{q}{m}\bar\Psi\Psi A^\mu. \end{aligned} \tag{2.24}$$

The second contribution is (here we sum over $\nu' \neq \mu$)

$$\begin{aligned} J^\mu_{(2)} &= \frac{i}{2m}\bar\Psi\left[\gamma^\mu\gamma^{\nu'}\overrightarrow{D_{\nu'}} - \overleftarrow{D^*_{\nu'}}\gamma^{\nu'}\gamma^\mu\right]\Psi = \frac{i}{2m}\bar\Psi\left[\gamma^\mu\gamma^{\nu'}(\overrightarrow{D_{\nu'}} + \overleftarrow{D^*_{\nu'}})\right]\Psi = \\ &= \frac{i}{2m}\bar\Psi\left[\gamma^\mu\gamma^{\nu'}(\overrightarrow{\partial_{\nu'}} + \overleftarrow{\partial_{\nu'}})\right]\Psi = \frac{i}{2m}\partial_{\nu'}(\bar\Psi\gamma^\mu\gamma^{\nu'}\Psi) = \\ &= \frac{1}{2m}\partial_\nu(\bar\Psi\Sigma^{\mu\nu}\Psi), \end{aligned} \tag{2.25}$$

where we used $\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}$, $\Sigma^{\mu\nu} \equiv \frac{i}{2}[\gamma^\mu,\gamma^\nu]$; full summation over ν is re-allowed on the last line. Altogether

$$\begin{aligned} J^\mu &= \bar\Psi\gamma^\mu\Psi = [J^\mu_{(1)}] + \{J^\mu_{(2)}\} = \\ &= \left[\frac{i}{2m}\bar\Psi\left[\overrightarrow{\partial^\mu} - \overleftarrow{\partial^\mu}\right]\Psi - \frac{q}{m}\bar\Psi\Psi A^\mu\right] + \left\{\frac{1}{2m}\partial_\nu(\bar\Psi\Sigma^{\mu\nu}\Psi)\right\}. \end{aligned} \tag{2.26}$$

Knowing that the total J^μ is conserved and observing that also $J^\mu_{(2)}$ is conserved by itself ($\Sigma^{\mu\nu}$ is antisymmetric) we come to the conclusion that all J^μ, $J^\mu_{(1)}$ and $J^\mu_{(2)}$ are *separately* conserved.

The first term in (2.26) is formally identical to the corresponding result for the Klein–Gordon (i.e., structureless, spin-zero) field, Φ, except that we would have to replace $\bar\Psi$ and Ψ with Φ^+ and Φ, respectively. The second term is a contribution arising from spin degrees of freedom.

Investigation into the components of $J^\mu_{(2)}$,

$$J^0_{(2)} = \frac{1}{2m}\partial_\nu\bar\Psi\Sigma^{0\nu}\Psi = \frac{1}{2m}\partial_i\bar\Psi\Sigma^{0i}\Psi = \frac{i}{2m}\boldsymbol{\nabla}\cdot\bar\Psi\boldsymbol{\alpha}\Psi = -\boldsymbol{\nabla}\cdot\boldsymbol{P}, \tag{2.27}$$

where $\boldsymbol{P} \equiv \frac{-i}{2m}\bar\Psi\boldsymbol{\alpha}\Psi$ and we used $\Sigma^{0i} = i\alpha^i$, and

$$J^i_{(2)} = \frac{1}{2m}(\partial_0\bar\Psi\Sigma^{i0}\Psi + \partial_j\bar\Psi\Sigma^{ij}\Psi) = \frac{1}{2m}(\partial_0\bar\Psi(-i\alpha^i)\Psi + \partial_j\bar\Psi\varepsilon^{ijk}\Sigma^k\Psi),$$

(Σ^{ij}, being antisymmetric in ij can be written as $\varepsilon^{ijk}\Sigma^k$), i.e.,

$$\boldsymbol{J}_{(2)} = \frac{-i}{2m}\partial_t(\bar\Psi\boldsymbol{\alpha}\Psi) + \frac{1}{m}\boldsymbol{\nabla}\times(\bar\Psi\frac{\boldsymbol{\Sigma}}{2}\Psi) = \partial_t\boldsymbol{P} + \boldsymbol{\nabla}\times\boldsymbol{M}, \tag{2.28}$$

where $\boldsymbol{M} \equiv \frac{1}{m}\bar\Psi\frac{\boldsymbol{\Sigma}}{2}\Psi = \frac{1}{2m}\Psi^+\beta\boldsymbol{\Sigma}\Psi$ (we recall that $\beta = \gamma^0 = \gamma_0$), suggests interpretation of the terms as contributions from an internal 'polarization' $\boldsymbol{P}$ and 'magnetization' $\boldsymbol{M}$ densities. (There is a missing factor of charge, q, in these definitions and one must multiply $\boldsymbol{P}$ and $\boldsymbol{M}$ by it if the true 'electromagnetic' densities are wanted.)

Armed with these results we may rewrite the second term on the last line of (2.14), the interaction of the system with the external field. First,

$$J_\mu A^\mu_{\rm ext} \;=\; J_0 A^0_{\rm ext} \,+\, J_i A^i_{\rm ext} \;=\; \overline{\Psi}\gamma_0\Psi A^0_{\rm ext} \,-\, J^i A^i_{\rm ext} \;=$$
$$=\; \Psi^+\Psi A^0_{\rm ext} \,-\, (J^i_{(1)} + J^i_{(2)})\, A^i_{\rm ext}. \tag{2.29}$$

Second, for the *space* components of the current we use the decompositions (2.24) and (2.28):

$$J_\mu A^\mu_{\rm ext} \;=\; \Psi^+\Psi A^0_{\rm ext} \,-\, \boldsymbol{J}_{(1)}\cdot\boldsymbol{A}_{\rm ext} \,-\, (\partial_t \boldsymbol{P} \,+\, \boldsymbol{\nabla}\times\boldsymbol{M})\cdot\boldsymbol{A}_{\rm ext}, \tag{2.30}$$

where

$$\boldsymbol{J}_{(1)} \;=\; \frac{1}{2m}\,\overline{\Psi}[(\vec{\boldsymbol{p}} - q\boldsymbol{A}) - (\overleftarrow{\boldsymbol{p}} + q\boldsymbol{A})]\,\Psi. \tag{2.31}$$

Now, for our purposes, since we are interested in a *stationary* state, the first part of the third term in (2.30), $\propto \partial_t \boldsymbol{P}$, is zero. Furthemore, since (2.30) represents really a Hamiltonian *density*, which needs to be integrated over space, the second part of the third term can be written as

$$(\boldsymbol{\nabla}\times\boldsymbol{M})\cdot\boldsymbol{A}_{\rm ext} \;=\; (\boldsymbol{A}_{\rm ext}\times\vec{\boldsymbol{\nabla}})\cdot\boldsymbol{M} \overset{\int}{=} -(\boldsymbol{A}_{\rm ext}\times\overleftarrow{\boldsymbol{\nabla}})\cdot\boldsymbol{M} \;=\; (\boldsymbol{\nabla}\times\boldsymbol{A}_{\rm ext})\cdot\boldsymbol{M} \;=$$
$$=\; \boldsymbol{B}_{\rm ext}\cdot\boldsymbol{M}, \tag{2.32}$$

where the second equality is meant under the space integration and we further integrated by parts; $\boldsymbol{B}_{\rm ext} \equiv \nabla\times\boldsymbol{A}_{\rm ext}$ denotes the external magnetic field.

Thus, under stationary conditions, (2.30) acquires the form

$$J_\mu A^\mu_{\rm ext} \;=\; nV_{\rm ext} \,-\, \boldsymbol{J}_{(1)}\cdot\boldsymbol{A}_{\rm ext} \,-\, \boldsymbol{B}_{\rm ext}\cdot\boldsymbol{M}, \tag{2.33}$$

where $n \equiv \Psi^+\Psi$ is the particle density, $V_{\rm ext} \equiv A^0_{\rm ext}$ has the form of a scalar potential. So far without any approximation.

The second term in (2.33) reflects interaction of the external magnetic field with the displacement (orbital) current. Neglect of this term would physically mean neglect of diamagnetic effects, which is arguably a rather good approximation for metals. In the following we accept this 'spin-only' simplification, actually the same as was considered in section 1.5. Then (2.14) reads

$$H \;\simeq\; H_{\rm sys} \,+\, qnV_{\rm ext} \,-\, q\boldsymbol{B}_{\rm ext}\cdot\boldsymbol{M} \;\equiv\; H_{\rm sys} \,+\, nv_{\rm ext} \,-\, \boldsymbol{B}_{\rm ext}\cdot\boldsymbol{m}, \tag{2.34}$$

with

$$v_{\rm ext} \;\equiv\; qV_{\rm ext,} \qquad \boldsymbol{m} \equiv q\boldsymbol{M} \;=\; \frac{q}{2m}\,\Psi^+\beta\boldsymbol{\Sigma}\Psi \;=\; -\mu_B\Psi^+\beta\boldsymbol{\Sigma}\Psi. \tag{2.35}$$

Finally, as we did before in the transition from (2.14) to (2.16), we again design a single-particle Hamiltonian of an auxiliary system by replacing the inter-particle interactions hidden in $H_{\rm sys}$ with effective fields, now, however, following the structure of (2.34):

$$H_0' \;=\; (\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta m) \,+\, v^{\rm eff} \,+\, \mu_B\boldsymbol{B}^{\rm eff}\cdot\beta\boldsymbol{\Sigma}. \tag{2.36}$$

This Hamiltonian supplemented with prescriptions for the effective fields (which should ensure that the groundstates of the interacting and auxiliary systems share the same $n(\boldsymbol{r})$ and $\boldsymbol{m}(\boldsymbol{r})$)

$$v^{\rm eff}(\boldsymbol{r}) \;=\; v_{\rm ext}(\boldsymbol{r}) \,+\, \int \mathrm{d}\boldsymbol{r}'\,\frac{n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \,+\, \frac{\delta E_{\rm xc}[n,\boldsymbol{m}]}{\delta n(\boldsymbol{r})}, \tag{2.37}$$

$$\boldsymbol{B}^{\rm eff}(\boldsymbol{r}) \;=\; \boldsymbol{B}_{\rm ext}(\boldsymbol{r}) \,+\, \frac{\delta E_{\rm xc}[n,\boldsymbol{m}]}{\delta \boldsymbol{m}(\boldsymbol{r})}, \tag{2.38}$$

where n and $\boldsymbol{m}$ are to be calculated from the N lowest-energy solutions of $H_0'\psi_i = E_i\psi_i$,

$$n(\boldsymbol{r}) \;=\; \sum_{i=1}^{N} \psi_i^+(\boldsymbol{r})\,\psi_i(\boldsymbol{r}), \tag{2.39}$$

$$\boldsymbol{m}(\boldsymbol{r}) \;=\; -\mu_B \sum_{i=1}^{N} \psi_i^+(\boldsymbol{r})\,\beta\boldsymbol{\Sigma}\psi_i(\boldsymbol{r}), \tag{2.40}$$

form the relevant Kohn–Sham equations for the current situation. These are, as before, to be solved

selfconsistently.

The just-described formalism is quite similar to the one of section 1.5, though there we implicitly considered a non-relativistic system described by a Schrödinger equation. Here we have a 'spin-only' approximation of a fully relativistic theory, with the auxiliary single-particle Hamiltonian being based on the Dirac instead of Schrödinger Hamiltonian. Furthemore, it is important to note that the exchange-correlation functional, E_{xc}, of a relativistic theory generally contains more effects (retardation to name at least one) than its non-relativistic counterpart. In spite of this fact, for relativistic calculations of this thesis—which are based on the above 'spin-only' theory—we use a non-relativistic E_{xc}, in its LDA approximation as discussed in section 1.6.

With the introduction of the spin-orbit interaction one may ask questions about how it can affect orbital magnetism. Actually, with a spin-only theory without the spin-orbit interaction (and without terms that would take into account the so-called shape anisotropy) no orbital magnetism would exist. In this sense, in our current theory, the spin-orbit interaction is the only mechanism which may induce (if spin degeneracy gets lifted) some orbital magnetism, (Erik99). Fully relativistic calculations done within the LSDA theories, however, so far yield, e.g. for Co, results that are by a factor of about three too small (for Co, the experimental value is approximately $0.3\mu_{\mathrm{B}}$, while calculated is $0.09\mu_{\mathrm{B}}$), (Söd92, Erik99, Eber99, Ture12). Some improvement may be achieved in more complicated theories, (Esch05, Chad08).

In this chapter we used a system of units with $c=1$. We close the chapter by rewriting H_0' with the c-factors added. We will call this result a *Kohn–Sham–Dirac (KSD) Hamiltonian*:

$$H_{\mathrm{KSD}} = c\boldsymbol{\alpha}\cdot\boldsymbol{p} + (\beta-1)mc^2 + v^{\mathrm{eff}}(\boldsymbol{r}) + \mu_B \boldsymbol{B}^{\mathrm{eff}}(\boldsymbol{r})\cdot\boldsymbol{\Sigma}\beta, \tag{2.41}$$

where we further subtracted the rest-mass energy of an electron.

3

Electronic Structure by the (L)MTO Method in the ASA

For numerically solving one-electron Kohn–Sham equations we use the linearized muffin-tin-orbital method (LMTO) in the Atomic-Spheres Approximation (ASA).

The method derives from J. C. Slater's simplifying proposition (Slat37) to divide space into spherical regions around atoms, where the potential is considered spherically symmetric (the potential close to an atom is mainly determined by that atom), and an interstitial region with a constant potential. The resulting potential profile, depicted in Fig. 3.1, resembles a muffin tin (MT) and is commonly known under this name. The reason for such a division is that an electronic wavefunction behaves differently near atomic nuclei, where the kinetic energy is high and the wavefunction rapidly oscillates, and between atoms, where the behaviour is much smoother.‖

The non-relativistic Kohn–Sham Hamiltonian to be solved, $H_{\mathrm{KS}}\phi = E\phi$, reads¶

$$\begin{aligned} (-\Delta + V^{\mathrm{eff}}_{\boldsymbol{R}}(r))\,\phi(\boldsymbol{r}) &= E\phi(\boldsymbol{r}) \qquad \text{inside an } \boldsymbol{R}\text{-sphere}, \\ (-\Delta + V_0)\,\phi(\boldsymbol{r}) &= E\phi(\boldsymbol{r}) \qquad \text{in the interstitial region}. \end{aligned} \tag{3.1}$$

In the case of a relativistic calculation, the Hamiltonian is to be appropriately replaced by the Kohn–Sham–Dirac form, (2.41). The potential $V^{\mathrm{eff}}_{\boldsymbol{R}}$ inside the $\boldsymbol{R}$-sphere is not a priori known. It comprises i) an intraatomic part originating from the attraction of the nucleus, the Hartree and exchange interaction among the electrons on the site, and ii) an interatomic part due to all other sites. The latter is, when the true potential from the surroundings is spherically symmetrized (which is what we do to obtain a symmetric $V^{\mathrm{eff}}_{\boldsymbol{R}}(r)$), just a constant shift, called the Madelung contribution.† We further note that $V^{\mathrm{eff}}_{\boldsymbol{R}}$ changes over iterations of solving the Kohn–Sham equations. Thus, in each iteration an electron solution is to be found numerically.** On the other hand, in the interstitial space (with the constant potential) the solution is known analytically.

All the Hamiltonians we will cope with will be real, i.e., $H^{*} = H$.‡ This has the nice

‖ For an illustration see e.g. Fig. 11.2 in (Ashc76).

¶ In the following we use a bit incoherent but light notation in which $X_{\boldsymbol{R}}(r)$ denotes the value of the quantity X 'naturally' centred at $\boldsymbol{R}$ at a distance r from this centre. On the other hand, $X_{\boldsymbol{R}}(\boldsymbol{r})$ denotes the value of X at a point with global coordinates $\boldsymbol{r}$, i.e., *not* at $\boldsymbol{r}_{\boldsymbol{R}} \equiv \boldsymbol{r} - \boldsymbol{R}$.

† See chapter 5 of (Ture97) for details.

** In Appendix C we derive radial equations that have to be solved when an electron moves according to the Dirac equation in a spherical potential. If beside the spherical potential there is also an interaction of an electron's spin with an effective magnetic field, the situation is more complicated. A special (but often relevant case) when the magnetic field points along the z direction and its size only depends on the distance from the sphere's centre is discussed in section 6.2.2 of (Ture97).

‡ This disqualifies ordinary magnetic field, which would lead to a non-real term $\boldsymbol{p}\cdot\boldsymbol{A}$ in the Hamiltonian. Nonetheless we still consider the (presumably for us more important) coupling of magnetic field to spin degrees of freedom. Then, if the 'magnetic' field everywhere lies (or can be made to lie) in the xy plane, the coupling term, $\propto \boldsymbol{B}\cdot\boldsymbol{\sigma}$, is real; this allows us to treat thus limited kind of non-collinear magnetism.

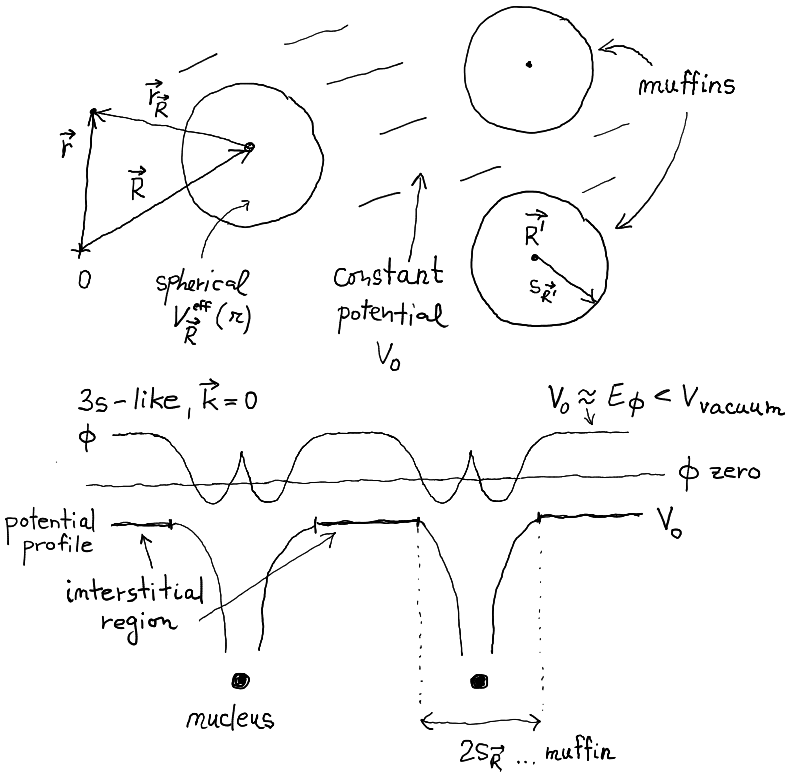


Figure 3.1: Muffin-tin approximation of a potential. An example of the potential profile and a 3*s*-like wavefunction ϕ (with Bloch's $\boldsymbol{k}$-vector zero so that ϕ can be real everywhere and can be drawn) of a valence (sizeable ϕ in the interstitial region) electron as a function of position between two nuclei. The flat character of ϕ in the interstitial region implies a small kinetic energy there, i.e., $\kappa^2 \equiv E_\phi - V_0 \approx 0$.

implication that if ϕ is a solution of (3.1) with energy E, so is ϕ^*. Then the basis of solutions can always be chosen real (if need be, we utilize that $\phi+\phi^*$ and $i(\phi-\phi^*)$ are both real and span the same subspace as ϕ and ϕ^*).

3.1. A Single-Muffin Problem

As a preparatory step we want to find physically relevant¶ solutions $\phi_i(E,\boldsymbol{r})$§, $\boldsymbol{r}\in \mathrm{R}^3$, for a certain energy E of

$$(-\Delta + V(r))\,\phi_i(E,\boldsymbol{r}) = E\phi_i(E,\boldsymbol{r}), \qquad V(r>s) = V_0, \tag{3.2}$$

with $V(r)$ being a given continuous potential function. The spherical symmetry of this equation means that a basis of the solutions exists with its elements' space-dependencies factorized into radial and angular parts:*

$$\phi_{Li}(E,\boldsymbol{r}) = \phi_{li}(E,r)\; Y_L(\hat{\boldsymbol{r}}); \tag{3.3}$$

here $L\equiv lm$ is a compound index comprising the angular (l) and magnetic (m) quantum numbers, and Y_L designates the spherical harmonics.§§ The radial part then (for $r\neq 0$) satisfies

¶ I.e., either normalizable to unity ($\in L^2(\mathrm{R}^3)$) to represent a bound state, or δ-function normalizable to represent a scattering state.

§ The index i will distinguish members of some set, possibly supplementing other indices (E in (3.2), E and L in (3.3)).

* The wording is to avoid saying that any solution with energy E is factorizable. A counter-example is a linear combination of a 2s and 2p state of the Coulomb potential.

§§ As was mentioned above, it's generally possible as well as advantagous to work in a basis of real ϕs. To this end, instead of using the usual (complex) spherical harmonics, often real combinations of these (which span the same space) are used instead. A possible choice of the real combinations is given in Appendix D.

$$\left[-\partial_{rr} + \frac{l(l+1)}{r^2} + V(r) - E\right] r\phi_{li}(E,r) = 0. \tag{3.4}$$

For potentials less singular than r^{-2} at the origin (i.e., including the Coulomb potential), the behaviour of ϕ close to the origin is dominated by the first two terms in the bracket (even for $l=0$), suggesting two possibilities (two values for i): $\phi_l \propto r^l$ and $\propto r^{-l-1}$. The latter, however, is to be abandoned, since it actually does not solve (3.2) at the origin.‖ For a given E and L we thus have one potentially physically relevant solution of (3.2), behaving like r^l at the origin and apparently square-integrable there. We may follow (e.g. by numerical means) this solution outwards, until we reach the sphere boundary, $r=s$.

Outside the sphere the radial part of a wavefunction, $\phi_l(E,r)$, can be expressed analytically as a linear combination of a spherical Bessel (j_l) and Neumann (n_l) function,

$$\phi_l(E,r\geq s) = \alpha_l(\kappa) j_l(\kappa r) + \beta_l(\kappa) n_l(\kappa r), \quad \kappa = \begin{cases} (E-V_0)^{1/2} & [\kappa^2 \geq 0] \text{ if } E \geq V_0, \\ i(V_0-E)^{1/2} & [\kappa^2 \leq 0] \text{ if } E \leq V_0, \end{cases} \tag{3.5}$$

which must smoothly match the solution inside the sphere; this determines α and β.

There is an unbound (scattering-like), δ-function normalizable, solution for any *real* κ (equivalently, $\kappa^2 \geq 0$ or E from the continuous spectrum of $-\Delta + V$). For an *imaginary* κ (equivalently, $\kappa^2 \leq 0$ or E from the discrete spectrum**) there is either a solution normalizable to unity (a bound state) or no physically relevant solution (both j and n exponentially blow up at infinity, and only their combination proportional to $j+in$, i.e. the first spherical Hankel function, usually denoted $h^{(1)}$, is normalizable, with the growing terms in j and n cancelling).

Before we turn our attention to a many-muffin problem, we note that when the scattering picture is relevant ($\kappa^2 > 0$), i.e., when the electron has positive kinetic energy outside the sphere, often the scattering process is described in terms of so-called phase shifts, δ_l. A particle before scattering is described by, e.g., a plane wave, which can be decomposed into a sum of spherical waves,‡

$$e^{i\boldsymbol{\kappa}\cdot\boldsymbol{r}} = \sum_L 4\pi i^l Y_L^*(\hat{\boldsymbol{\kappa}}) \{j_l(\kappa r) Y_L(\hat{\boldsymbol{r}})\} = \sum_L c_L(\hat{\boldsymbol{\kappa}}) \{J_L(\kappa, \boldsymbol{r})\}. \tag{3.6}$$

The decomposition is useful since individual spherical components when being scattered on a spherically symmetric potential do not mix (i.e., L is conserved), hence each such component can be considered separately. For $r \geq s$ the L-component's radial behaviour must be of the form of (3.5). For a large $x \equiv \kappa r$, and denoting $x - l\pi/2 \equiv y$,

$$\begin{aligned} \alpha j_l(x) + \beta n_l(x) &\approx \frac{1}{x}\left[\alpha \sin(x - \frac{l\pi}{2}) - \beta \cos(x - \frac{l\pi}{2})\right] = \\ &= \frac{1}{2ix}\left[(\alpha - i\beta) e^{iy} - (\alpha + i\beta) e^{-iy}\right], \end{aligned} \tag{3.7}$$

from which it follows that $|\alpha - i\beta|$ must equal $|\alpha + i\beta|$ (i.e., α and β cannot be completely arbitrary) if $(\alpha j + \beta n)$ is to represent a stationary scattering of an incoming partial wave, e^{-iy}, into the outgoing one, e^{iy} (in such a process particles are neither created nor lost anywhere and the incoming probability flux $\propto |\alpha - i\beta|^2$ must equal the outgoing flux $\propto |\alpha + i\beta|^2$§); this means $\alpha/\beta \in \mathrm{R}$. The relative phase between the incoming and scattered partial wave can be used to define the *phase shift* $\delta_l \in [-\pi/2, \pi/2]$:

$$\frac{\alpha_l - i\beta_l}{\alpha_l + i\beta_l} \equiv e^{i2\delta_l}. \tag{3.8}$$

‖ See, e.g., chapter 7, topic A, subtopic 2c (p. 781) of (Cohe77). They argue that $\Delta \frac{1}{r^{l+1}}$ involves the lth derivative of $\delta(r)$, which, however, does not appear elsewhere in the equation and thus cannot be compensated for.

** We wish to stress that the discrete spectrum (quantization) is i) determined and ii) a result of the *boundary conditions* the solutions have to fulfil, here square-integrability over the whole R^3.

‡ Particularly noteworthy is that only js but no ns appear in (3.6).

§ The angular dependence of the wavefunction, being the same for the incoming and outgoing wave, does not spoil this

Knowledge of all δ_ls is equivalent to the knowledge of how a given (spherical) potential scatters particles.†

3.2. Infinite Number of Muffins

We will now turn our attention directly to the case of infinitely many muffins. It might perhaps seem more appropriate to consider the case of a finite cluster of atoms first, but this would mean that we would have to take special care of appropriate boundary conditions: electrons in a finite piece of solid are all bound to it (and their wavefunctions are hence, from a global viewpoint, 1-normalizable). However, to attain understanding of transport phenomena in a solid (in contrast to e.g. a molecule) we actually do not need to deal with such details—the (bulk) properties are determined by the situation 'inside' the solid, and are not much susceptible to what is happening at the boundaries.* Thus it seems adequate to consider a formally infinite solid (though not necessarily periodic) and argue that a real sample, when large enough, shares with it the interesting properties.

When an infinite piece of solid is considered, a (formal) electron's energy-eigenstate wavefunction is usually not 1-normalizable in the whole space: For example, if the solid is periodic, one may deploy Bloch's theorem to classify the energy eigenstates, and each such eigenstate leads to an equal probability of finding the electron around any of the sites. The states may then be, e.g., normalized to one in the volume of any Wigner–Seitz cell. The condition of global continuity of the eigenstates, including their 1st space derivatives, and the requirement to fulfil Bloch's condition lead to the solid's band-structure $E(\boldsymbol{k})$.

For the Schrödinger equation (3.1) the interesting energy range begins where an electron can move or tunnel from one atom to the next, i.e. at about ½ Ry below the level of the interstitial potential V_0, and ends at the Fermi level, which is often about ½ Ry above V_0 as well as about ½ Ry below the vacuum level. Electrons occuring at these energies, called valence, determine the electronic behaviour, and are the main target for the electronic-structure calculation. Core states, on the other hand, are rather inert for many purposes.

The (one-electron) methods may be classified according to whether they seek the wavefunctions as an expansion in some set of *fixed basis-functions*, like atomic orbitals, plane waves, gaussians, Slater-type orbitals, or they expand the wavefunctions in a set of energy-dependent *partial waves*, as done in the Wigner–Seitz cellular method, Korringa–Kohn–Rostoker (KKR) method and augmented-plane-wave (APW) method. The former type of methods have the advantage that the determination of the expansion coefficients mathematically leads to a 'simple' (generalized) eigenvalue problem '$\det[H_{ij}-EO_{ij}]=0$', which can be readily handled by linear-algebra software. This is to be contrasted with the latter methods, which generally lead to a much more time-consuming problem of finding roots of '$\det M(E)=0$', where the energy dependence is nonlinear. On the other hand, the methods using energy-independent bases may need many basis functions to achieve wanted accuracy, while the methods using partial waves already use solutions of the related single-site problems, which 'only' have to be correctly matched together; in this sense a basis arising from partial waves is *minimal*, there is just one *s*-like, three *p*-like, five *d*-like, etc., orbitals per atom.§

A direct generalization of the single-site problem of the last section would be to try and

---

argument.

† The factor of 2 has its roots in the following rewrite

$$(e^{-iy} - e^{i2\delta_l}e^{iy}) = -2ie^{i\delta_l}\sin(y+\delta_l), \tag{3.9}$$

showing that when the wavefunction (3.7) is written in such a way that there is just a coefficient of $1/\kappa r$ in front of the (total) incoming wave (this fixes the global phase as well as the normalization), the wavefunction at large r is shifted with respect to the one when no potential would exist by a phase δ_l. Also, with this normalization, the wavefunction shows the same modulo–π δ-ambiguity (see 3.9) as the phase shift itself in (3.8).

* For example, we note that the model of nearly-free electrons sometimes fairly well describes the solid and in this model the fact that electrons are eventually bound by boundaries is of secondary importance.

§ Quite generally we do not need more partial waves of the same character, such as e.g. two *s*-like ones, since these would be (thanks to the orthogonality) energetically remote—the additional partial wave would either be too high to be ever filled or too deep to represent a valence electron.

match, for a given energy, solutions of individual muffin potentials (i.e. partial waves) together by an interstitial-space solution of the same energy. One way to carry out this programme is a construction of so-called (energy-dependent) muffin-tin orbitals (MTOs), smooth and centred around individual sites (but having so-called tails reaching to the other sites), and then searching for their linear combination which would fulfil the so-called tail-cancellation condition, i.e. the condition ensuring that the combination solves the Schrödinger equation on a global scale; see (Ande71, Kaso72) and also (Ande73), which is more verbose as well as general. This at-that-time newly devised so-called Linear Combination of Muffin-Tin Orbitals (LCMTO)* method has been very close to the older KKR multiple-scattering method, and it has been shown in the references that the KKR wavefunction may in fact be interpreted as a one-centre expansion of the related linear combination of MTOs. The LCMTO method has, however, shown to have far better *l*-convergence than the KKR spherical wave expansion. Furthemore, it also opened up a way to include non-MT-like (i.e. not spherically symmetric) potentials.

The straightforward use of energy-dependent orbitals in the LCMTO method like outlined in the last paragraph means such a method features the root-finding difficulty mentioned above. Instead of giving more details about it, we now provide a few ideas on a further historical development, which lead to the LMTO method actually used in our codes.

In (Ande75) O. K. Andersen described so-called *linear methods*,† namely the Linear Augmented Plane-Wave Method (LAPW) and Linear Muffin-Tin Orbital Method (LMTO); for a review see, e.g., (Ande85), Andersen, Jepsen and Šob's contribution in (Yuss87), chapter 17 of (Mart04), but, regarding the LMTO method, probably the best reference for a start is (Skri84).‡ These methods employ energy-*in*dependent basis functions, which are however derived from the partial waves. Thus the linear methods combine desirable features of fixed-basis and partial-wave methods: i) they can principally provide solutions of arbitrary accuracy when applied to the muffin-tin (MT) part of the real potential (thanks to the use of from-partial-waves derived basis), yet they are not limited to MT potentials (thanks to the variational character of the solution-finding), ii) they lead to an 'easy' generalized eigenvalue problem. Furthemore, the LMTO method, at least in its simplest form (single κ), uses a minimal basis.

It seems there have been several courses of thinking driving the development of the LMTO method. The original one, discussed already in (Ande73), is reviewed in chapter 5 of (Skri84). A more comprehensive view (but also much more involved) summarizing with hindsight the twenty-year-long progress of the method (also in relation to the so-called screening transformations, which were not yet known at the time Skriver wrote his book) can be found in (Ande92). Here we will only succinctly provide one fairly intuitive (and far from complete) argument supporting the form of LMTO orbitals, basically following the introduction in (Ande99), and let the reader consult the references for further information.

At each step of the calculation there is a set of energies, usually denoted by $E_{Rl\nu}$ (one for each orbital type at each atom; ν is just a label here, not an index), chosen so that they lie in the energy interval under study, and for which (exact) partial waves, $\phi_{Rl\nu}$, and their energy derivatives, $\dot{\phi}_{Rl\nu}$, are calculated. Then the expression

$$\phi_{Rl}(r;E) = \phi_{Rl\nu}(r) + \dot{\phi}_{Rl\nu}(r)(E-E_{Rl\nu}) + \mathrm{O}((E-E_{Rl\nu})^2), \tag{3.10}$$

truncated after the linear term, can be thought of as an approximation of the radial part of the *l*-character solution around the site $\boldsymbol{R}$ at an energy E close to $E_{Rl\nu}$; ϕ, $\dot{\phi}$ with a missing energy parameter denote, respectively, the wavefunction and its energy derivative at $E=E_{Rl\nu}$. It is seen, e.g. in Fig. 1 of (Ande75), that this Taylor expansion quickly converges for the there-given example, and

* It should not be confused with the LMTO method, described further, where the 'L' in the name stands for a linear approximation, not for a linear combination.

† Note, however, that the linearization idea appeared already in earlier Andersen's works, esp. in (Ande73).

‡ The author may only express his sadness about not getting hold of this book until very recently. Skriver seems to have expanded and discussed in detail and (importantly) with certain comprehensibility the original knowledge compressed in Andersen's papers as well as never published notes.

we will assume it is generally the case. The idea behind the linearized methods is then to construct an energy-*in*dependent basis set, $\chi_{RL}(\boldsymbol{r})$—of *smooth* (continuous and once differentiable) orbitals centred at individual sites $\boldsymbol{R}$ (and also labelled with angular momentum L whose meaning will become clearer below)—, in such a way that for each exact solution Ψ_i with energy E_i there exists a linear combination of the basis members that approximates Ψ_i around each site correctly to the linear order.¶§ Mathematically, for the i^{th} solution there need be coefficients $c_{RL,i}$ such that, for $\boldsymbol{r}$ in any muffin sphere,

$$\sum_{RL} \chi_{RL}(\boldsymbol{r})\, c_{RL,i} \;=\; \sum_{RL} \left[\phi_{RL}(\boldsymbol{r}) \;+\; (E_i - E_{Rl\nu})\, \dot{\phi}_{RL}(\boldsymbol{r}) \right] c_{RL,i} \;\approx \tag{3.11}$$

$$\approx \;\sum_{RL} \phi_{RL}(E_i, \boldsymbol{r})\, c_{RL,i} \;=\; \Psi_i(\boldsymbol{r}).$$

Hereafter ϕ and $\dot{\phi}$ are considered set to zero outside their mother muffin, $\phi_{RL}(\boldsymbol{r} \notin \boldsymbol{R}) = 0$, $\dot{\phi}_{RL}(\boldsymbol{r} \notin \boldsymbol{R}) = 0$. It is straightforward to see, by a direct substitution into (3.11), that orbitals

$$\chi_{RL}(\boldsymbol{r}) \;\equiv\; \phi_{RL}(\boldsymbol{r}) \;+\; \sum_{R'L'} \dot{\phi}_{R'L'}(\boldsymbol{r}) [H_{R'L',RL} - E_{Rl\nu}\, \delta_{R'L',RL}] \;+\; \chi^{\mathrm{I}}_{RL}(\boldsymbol{r}) \tag{3.12}$$

behave as needed by (3.11) if $c_{RL,i}$ are eigenvectors of H with eigenvalues E_i.‖ The last term, χ^{I}_{RL}, describes the orbital in the interstitial region (it is not present in (3.11) since there we only examine the situation *inside* the spheres).

An orbital χ_{RL} is effectively defined by its *envelope*, which determines (perhaps up to a factor) the orbital's behaviour in the interstitial region, see Fig. 3.2. The orbital's behaviour inside the spheres is then dictated by the smoothness requirement. We will now concentrate on a particular form of orbitals, called *unscreened* LMTOs. Each such χ_{RL} has as its envelope a spherical harmonics with the L-character about the orbital's centre, $\boldsymbol{R}$, the envelope is regular at infinity and solves the wave-equation for a particle in a constant potential everywhere (except at the central point $\boldsymbol{R}$).

We have required that each χ_{RL} be smooth in all space, however, it is perhaps not obvious that this can be achieved with the presented form, i.e., that $H_{R'L',RL}$ in (3.12) can be chosen so that all orbitals χ_{RL} are smooth. Regarding this point we refer the reader to the literature, but note here that so far we have nowhere specified the normalization of $\phi_{RL}(E)$, and this freedom is, together with the freedom in $H_{R'L',RL}$, sufficient to meet the smoothness requirement, at least for the choice of the envelope functions made in the (L)MTO method.‡

¶ Nevertheless, see (Ande99) and Andersen, Jepsen and Krier's contribution in (Kuma94) for generalization of the LMTO method to achieve higher-order accuracy, so-called N^{th}-order (or Exact) Muffin-Tin–Orbital Method (NMTO, EMTO).

§ Eq. (3.10) suggests that using, for each site, two orbitals, one behaving like ϕ and the other like $\dot{\phi}$ around the site would just meet the requirement. However, this would lead to twice as many basis elements compared to the basis actually used in the LMTO method.

‖ H is sometimes called a first-order Hamiltonian of the problem.

‡ One may always write the (inside-muffin) solution with an arbitrary energy E as (dropping everywhere the RL indices) $\phi(E) = N(E)\tilde{\phi}(E)$ with $\tilde{\phi}(E)$ being—for any energy—always normalized to one in its sphere, $\int \mathrm{d}\boldsymbol{r}\, \tilde{\phi}^2(E,\boldsymbol{r}) = \int \mathrm{d}r\, r^2 \tilde{\phi}^2(E,r) = 1$. Then $\dot{\phi}(E) = \dot{N}(E)\tilde{\phi}(E) + N(E)\dot{\tilde{\phi}}$, and, if we choose $N(E_\nu) = 1$, $\dot{\phi}(E_\nu) = \dot{\tilde{\phi}}(E_\nu) + \tilde{\phi}(E_\nu)\dot{N}(E_\nu)$, in short $\dot{\phi} = \dot{\tilde{\phi}} + \tilde{\phi} o$, where $o \equiv \dot{N}(E_\nu)$ is a real Rl-diagonal matrix. It can be shown that radial dependencies of $\tilde{\phi}$ and $\dot{\tilde{\phi}}$ always have different (radial) logarithmic derivatives at the sphere's surface (see, e.g., equation 2.9. in (Ande75), or equations 2.83 and 2.20 in (Ture97)), and hence by a choice of the o value one may adjust $\dot{\phi}$ to have an arbitrary logarithmic derivative there. Further, an important quality of envelopes used in the (L)MTO theory is that each one, say N_{RL} (centred at $\boldsymbol{R}$), can be within any other sphere $\boldsymbol{R}'$ expressed as a linear combination of functions Js (their precise definition is unimportant here) thus: $N_{RL}(\boldsymbol{r} \in \boldsymbol{R}') = -\sum_{L'} J_{R'L'}(\boldsymbol{r})\, S_{R'L',RL}$, where the function $J_{R'L'}$ has the L'-character around $\boldsymbol{R}'$ and S is a (so-called structure-constants) matrix; we utilize this theorem later in (3.25) for the Ns and Js relevant in the ASA. The importance of the expansion theorem is that it shows that irrespective of where the mother sphere of an envelope is, $\boldsymbol{R}$, its L'-component around $\boldsymbol{R}'$ will have, at the surface of the $\boldsymbol{R}'$-sphere, the logarithmic derivative given by that of $J_{l'}(r)$, i.e., it will be *the same for all envelopes* (particularly, in the ASA case with $\kappa^2 = 0$ and the conventional unscreened envelopes N^0, it is $+l'$). This fixes the logarithmic derivative of all $\dot{\phi}$s (see Fig. 3.2) and thus determines all o_{Rl}s. The elements $H_{RL',RL}$ are then adjusted so that the logarithmic derivative of the orbital's head is the same as that of the envelope for each L'-component present in the envelope (in the ASA with $\kappa^2 = 0$ and the

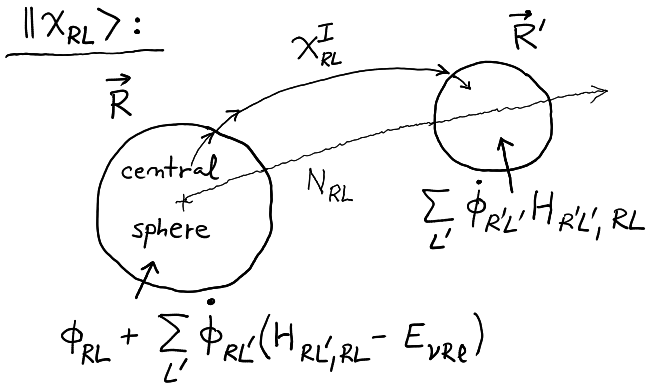


Figure 3.2: Structure of a general linear muffin-tin orbital (LMTO). Each orbital is 'centred' at a certain site, here $\boldsymbol{R}$, has a part within its mother (or central) sphere, called a *head*, and a part outside the sphere, called a *tail*. The orbital is effectively defined by its *envelope* function, often named $N_{RL}(\boldsymbol{r})$ (or K_{RL}). It is always a function centred at $\boldsymbol{R}$, regular at $r=\infty$, and at the same time a solution of a wave-equation for a particle in a constant potential at least in the interstitial region (the possibility that it does not have to be such a solution inside the foreign spheres is important for the screening phenomenon discussed later) with a suitably chosen kinetic energy κ^2 (see the text). In the interstitial region the envelope function equals (possibly up to a factor) the orbital's $\chi^{\rm I}_{RL}$ part. For an *unscreened* LMTO, the envelope is a single L-like spherical harmonics about $\boldsymbol{R}$; the value of RL then gives the orbital its name. (In the screened case the envelopes N_{RL} do *not* have just a single L component around their centres, yet the L component is the dominating one close to the centre of the orbital.) In the unscreened case, the head, in order to smoothly match the envelope, may only contain the same L-component as the envelope has, and thus consists of the solution ϕ_{RL} plus some $\dot{\phi}_{RL}$ (i.e., $H_{RL,RL'}$ in (3.12) must be diagonal in Ls). On the other side, when the envelope enters spheres around other sites, say for concreteness the $\boldsymbol{R}'$ sphere, it is to be smoothly matched to a combination of $\dot{\phi}_{R'L'}$s inside this sphere, with generally all L's (around $\boldsymbol{R}'$) present.

Hereafter, χ will implicitly denote the unscreened LMTOs.

A convenient and succinct way of writing (3.12) is, after (Ande92),

$$||\chi> \;=\; |\phi> \,+\, |\dot{\phi}>h \,+\, |\chi^{\rm I}>, \tag{3.13}$$

where the different 'brackets' show the extent of each term: $||\cdot>$ spans all the space, $|\cdot>$ is non-zero only in one's own muffin, and $|\cdot>$ is non-zero only in the interstitial region. Further, each such symbol is to be interpreted as a row vector of its 'RL-components', e.g., $|\phi> \equiv |\phi_{R_1L_1}, \phi_{R_2L_2}, \cdots>$, and analogously with the others. Similarly, we will use $<\cdot||$, $<\cdot|$, and $<\cdot|$ to represent related column vectors; thus $<\phi|\phi>$ is a matrix with elements $<\phi_{RL}|\phi_{R'L'}>$. In the second term of (3.13) h is the matrix $h \equiv H - E_\nu$ with the elements $h_{R'L',RL} \equiv H_{R'L',RL} - E_{Rl\nu}\delta_{R'L',RL}$, and the vector–matrix multiplication between $|\dot{\phi}>$ and h is implied.

Equation (3.11) offers a lucid picture of how overlapping of energy-independent orbitals χ_{RL}, i.e., letting the tails from the 'other' sites enter the 'central' site, leads to the correct (to the first order) energy dependence within the spheres.

Using $\tilde{\phi}_{RL}(E)$, which is $\phi_{RL}(E)$ (re)normalized to one in its sphere, see the last ‡ footnote, and the relation between $\dot{\tilde{\phi}}$ and $\dot{\phi}$, $\dot{\phi}_{RL} = \dot{\tilde{\phi}}_{RL} + \tilde{\phi}_{RL}\, o_{Rl}$, we may rewrite (3.13) as

$$||\chi> \;=\; |\tilde{\phi}>(1+oh) \,+\, |\dot{\tilde{\phi}}>h \,+\, |\chi^{\rm I}> \;=$$

conventional unscreened envelopes N^0 there is just one participating component, L, with the logarithmic derivative equal to $-l-1$), and the envelope is then normalized to match the head in value (alternatively, we could have the normalization of the envelopes fixed but keep the freedom in not setting $N(E_\nu)$ to one). Last, $H_{R'L',RL}$ are adjusted to match in value the envelope originating at $\boldsymbol{R}$ to its tail part in $\boldsymbol{R}'$.

$$= |\tilde{\phi}>\Pi + |\dot{\tilde{\phi}}>\Omega + |\chi^{I}>, \qquad \text{where} \quad \Pi\equiv 1+oh, \quad \Omega\equiv h, \tag{3.14}$$

which is sort of a 'standard' form of the orbitals' prescription.

A nice thing about the $\tilde{\phi}$ and $\dot{\tilde{\phi}}$ functions is that they are orthogonal to each other in their sphere, a property obtained by energy differentiation of $\int_{\boldsymbol{R}} \mathrm{d}\boldsymbol{r}\, \tilde{\phi}_{Rl}(E,\boldsymbol{r})^2 = 1$; i.e., $<\tilde{\phi}|\dot{\tilde{\phi}}> = 0$. Then, if the inverse of Π exists, which we assume, we may create a new set of so-called *nearly-orthogonal* orbitals,

$$||\theta> \equiv ||\chi>\Pi^{-1} = |\tilde{\phi}> + |\dot{\tilde{\phi}}>\Omega\Pi^{-1} + |\chi^{I}>\Pi^{-1}, \tag{3.15}$$

which are apparently *just linear combinations of the original orbitals* $||\chi>$. Their overlap matrix is

$$O^{\theta} \equiv <\theta|| \, ||\theta> = 1 + h^{\theta} p h^{\theta} + <\theta^{I}|\,|\theta^{I}>, \tag{3.16}$$

where $h^{\theta}\equiv\Omega\Pi^{-1} = h\Pi^{-1} = h(1+oh)^{-1}$, $p\equiv <\dot{\tilde{\phi}}|\dot{\tilde{\phi}}>$, and $|\theta^{I}>\equiv|\chi^{I}>\Pi^{-1}$.* The second term is usually small‖ and the last term will be 'annihilated' by making the atomic-spheres approximation (ASA) described below. Hence the overlap of two different θ-orbitals is almost zero, justifying their name. In the following we will be neglecting their overlap altogether.

The *Atomic-Spheres Approximation* (ASA), which all our calculations use,

- gets rid of the interstitial region by inflating the muffin spheres to the extent that the sum of their volumes just equals the total volume; integration over all space is then approximated by integration over the spheres and by summing up these contributions. By construction, there are necessarily places where the spheres overlap as well as places not covered by any sphere. Although it seems to be a drastic approximation, it both simplifies calculations since one does not have to do integrations over the complicated-in-shape interstitial space *and* the overlap can actually improve results, since it can be shown, see (Ande92, section V) that the physical meaning of the mathematical overlap is a superposition of potentials from the overlapping spheres and thus superposed potential is often closer to the reality than a flat interstitial potential of the conventional non-overlapping MTO theory.§
- sets the kinetic energy, κ^2, of the particles in the interstitial volume (or, better, of the kinetic energy of the envelope functions, since now we actually do not have any interstitial volume, but still have the envelopes, via which the orbitals (and their smoothness) are defined) to zero, $\kappa^2 = 0$.¶ First, some *fixed* value has to be chosen, because the orbitals must be energy-*in*dependent (in order to lead to an eigenvalue problem). Second, the zero value is essentially the simplest choice and lies in the middle of the energy range we are interested in. Third, the very value used is actually much less crucial in the space-filled-by-spheres configuration than in the non-overlapping one; see e.g. (Ande92, section III) for the reasons and for the comparison with the KKR method (where the κ-dependencies of relevant-theory elements are strong but actually cancel as unphysical).

We now have a set of orbitals, either χs or θs, whose combination may approximate correct solutions inside the muffins to the first order in $(E_i - E_{Rl\mathrm{v}})$. First-order energies seldom suffice, and in the LMTO method use is made of the variational principle for the Hamiltonian, so that errors of the order $(E_i - E_{Rl\mathrm{v}})^2$ in the basis set merely give rise to errors of order $(E_i - E_{Rl\mathrm{v}})^4$ in the energies (or to the order $(E_i - E_{Rl\mathrm{v}})^3$ if we use the θ-orbitals and neglect terms containing p). The

---

* On the RHS of (3.16) we had better write the second term as $h^{\theta+}ph^{\theta}$. However, after (3.20) it will be clear that h^{θ} is Hermitian.

‖ Realistic values of $p^{-½}$ are several Rydbergs. Surrounding p by h^{θ}s, where h^{θ} are 'small' when operating on wavefunctions with energy close to E_{v} (this follows from (3.20), 3.22)) makes the second term in (3.16) negligible.

§ One should realize that the most space-filling lattices, face-centred cubic (fcc) and hexagonal close packed (hcp), have the space-filling factor (with touching spheres) 0.74, so about 1/3 of the total volume is interstitial when spheres do not overlap and the potential in this volume is potentially badly approximated .

¶ When the ASA is not used, the role of envelopes is clear. When it is used, the envelopes are still necessary to connect the functions from one sphere to another, although this connection may actually (where the spheres overlap) involve some backing off.

variational energies $\overline{E}_i$ and eigenvectors $\overline{c}_{RL,i}$ are obtained as solutions of the generalized eigenvalue problem

$$\sum_{RL} [<\chi_{R'L'}||\hat{H}||\chi_{RL}> - \overline{E}_i <\chi_{R'L'}|| \, ||\chi_{RL}>] \; \overline{c}_{RL,i} = 0, \tag{3.17}$$

or of an equivalent but simpler (thanks to the orthogonality of θs in the ASA with the neglect of *p*-terms) *ordinary* eigenvalue problem

$$\sum_{RL} [<\theta_{R'L'}||\hat{H}||\theta_{RL}> - \overline{E}_i \delta_{R'L',RL}] \; \overline{c}^{\theta}_{RL,i} = 0. \tag{3.18}$$

In these equations $\hat{H}$ denotes the Kohn–Sham(–Dirac) Hamiltonian operator, see (3.1).

The term $<\theta||\hat{H}||\theta>$ in (3.18) can be using (3.15) rewritten in terms of $|\tilde{\phi}>$ and $|\dot{\tilde{\phi}}>$. Recalling that $|\tilde{\phi}>$ are normalized-to-one eigenfunctions of $\hat{H}$ inside the spheres, $\hat{H}|\tilde{\phi}(E)> = E|\tilde{\phi}(E)>$,[×] evaluated at $E = E_\nu$, and $|\dot{\tilde{\phi}}>$ are the related energy derivatives, for which $\hat{H}|\dot{\tilde{\phi}}> = |\tilde{\phi}> + E_\nu|\dot{\tilde{\phi}}>$ (obtained by energy derivating [×]), we get†

$$\begin{aligned} <\tilde{\phi}|\hat{H}|\tilde{\phi}> &= E_\nu, & <\tilde{\phi}|\hat{H}|\dot{\tilde{\phi}}> &= 1, \\ <\dot{\tilde{\phi}}|\hat{H}|\tilde{\phi}> &= 0, & <\dot{\tilde{\phi}}|\hat{H}|\dot{\tilde{\phi}}> &= E_\nu p. \end{aligned} \tag{3.19}$$

Then

$$<\theta||\hat{H}||\theta> = E_\nu + h^\theta + h^{\theta +} E_\nu p h^\theta. \tag{3.20}$$

Since $\hat{H}$ is a Hermitian operator,** it is—in any basis—necessarily represented by a Hermitian matrix. E_ν is a real diagonal matrix, so is p, hence both the first and the third RHS terms are Hermitian. It follows that h^θ must be Hermitian, too, and we can as well write just $h^\theta E_\nu p h^\theta$ in (3.20).§ We now again neglect the term $\propto p$ and define

$$H^\theta \equiv E_\nu + h^\theta. \tag{3.21}$$

A non-trivial solution of (3.18) exists if

$$\det(<\theta||\hat{H}||\theta> - \overline{E}_i) \approx \det(H^\theta - \overline{E}_i) = 0. \tag{3.22}$$

Thus the variational estimates of eigenenergies of the system's Hamiltonian $\hat{H}$ are the eigenvalues of the matrix H^θ. We will now discuss how H^θ can be parametrized by means of quantities readily available in an electronic-structure calculation.

The matrix H^θ is, via its construction, fundamentally related to the smoothness of the LMTOs. From the point of view of any one orbital, there are two boundaries of interest: the matching at the surface of the orbital's central sphere, and at all the other spheres. In both cases we are to match a function within a sphere to the orbital's envelope. The unscreened LMTOs, which we now denote simply by χs, have been defined to have simple envelopes. On the other hand, the θ orbitals, being combinations of χ*s*, see (3.15), already do not have this quality. It is therefore easier, when smoothness is discussed, to work with the unscreened χs (and keep in mind that when these are smooth, any combination of them is smooth, too), for them determine the relevant *h* and *o* matrices, and only eventually determine H^θ.

For an unscreened LMTO, the matching at the central sphere, ***R***, is easy, because the envelope

† These are essentially equations 2.68 in (Ture97).

** We note that $\hat{H}$ is necessarily Hermitian when considered over all space. When studying $\hat{H}$ in a limited region, like the inside of an ***R***-sphere, one should be a bit cautious, since there it is *not* Hermitian. E.g., the relation $<\tilde{\phi}_{RL}|\hat{H} - E_\nu|\dot{\tilde{\phi}}_{RL}> = 1$ (which can easily be obtained from the last equation before (3.19); it is also equation 2.9 in (Ande75)), or even just a look at the second and third equations in (3.19), might be surprising at first glance.

§ We defined h^θ as $h(1+oh)^{-1}$. This can be turned around into an inverse relation $h = h^\theta(1 - oh^\theta)^{-1} = = h^\theta + h^\theta o h^\theta + h^\theta o h^\theta o h^\theta + \cdots$, and we see (*o* is real and diagonal) that also the original matrices *h* and *H* are necessarily Hermitian.

has by definition only one L-component with respect to $\boldsymbol{R}$. The head then must also have only this L (c.f. Fig. 3.2), and the matching is possible if the radial logarithmic derivative, D,‡ of the inner function is the same as that of the envelope. We note that it is only the derivative *at the surface*, $|\boldsymbol{r}-\boldsymbol{R}|=s_R$, that matters for the successful matching. In this respect, we primarily need know this derivative.

In the ASA, the kinetic energy of envelopes is set to zero, $\kappa^2=E-V_0=0$, and the envelopes thus solve the Laplace equation $\Delta N_{RL}(\boldsymbol{r})=0$ in the interstitial region, see (3.1). They are required to be regular at infinity and an unscreened envelope has everywhere a single L component around its central site. Then the envelopes of unscreened LMTOs must inevitably be proportional to r^{-l-1} and are defined as (we use the superscript ^{0} to distinguish the unscreened objects)

$$N^0_{RL}(\boldsymbol{r}) \equiv N^0_l(r)\,Y_L(\hat{\boldsymbol{r}}_{\boldsymbol{R}}), \qquad N^0_l(r) \equiv (\frac{r}{w})^{-l-1} \quad \rightarrow \; D[N^0_l(r)] = -l-1, \tag{3.23}$$

where $\hat{\boldsymbol{r}}_{\boldsymbol{R}} \equiv (\boldsymbol{r}-\boldsymbol{R})^0$, and the quantity w is introduced so that N^0s are dimensionless; results do not depend on the exact value, but usually w is taken to be the average radius of the spheres (or the lattice constant). Clearly, N^0s are irregular (i.e., have a pole) at their respective centres. The other solutions of the Laplace equation, which together with the N^0s form the basis of all solutions, are the functions

$$J^0_{RL}(\boldsymbol{r}) \equiv J^0_l(r)\,Y_L(\hat{\boldsymbol{r}}_{\boldsymbol{R}}), \qquad J^0_l(r) \equiv \frac{1}{2(2l+1)}(\frac{r}{w})^{l} \quad \rightarrow \; D[J^0_l(r)] = l, \tag{3.24}$$

which are regular at their origins but diverge at infinity.*¶

From the last paragraph it follows that in the head of the χ_{RL}-orbital there must be a combination of $\tilde{\phi}_{RL}$ and $\dot{\tilde{\phi}}_{RL}$ that has its logarithmic derivative equal to $-l-1$ (and either this combination or the envelope should be renormalized so that they also meet in value).

The matching of N^0_{RL} to the other spheres at $\boldsymbol{R}'\neq\boldsymbol{R}$ then proceeds under the control of the famous expansion theorem§

$$N^0_{RL}(\boldsymbol{r}\in\boldsymbol{R}') = -\sum_{L'} J^0_{R'L'}(\boldsymbol{r})\,S^0_{R'L',RL}, \tag{3.25}$$

where the quantities $S^0_{R'L',RL}$ are so-called *canonical structure constants*.† The very prescription for them can be found as the equation 2.9 in (Ture97), but here the important quality is that they describe and only depend on the *geometry of the problem*, and the matrix S^0 is really the only place where geometry enters the formalism (as far as mutual positions of the spheres' centres are concerned; geometrically, beside S^0 we still need the spheres' radii, s_Rs). The matching of N^0_{RL} to an $\boldsymbol{R}'$-sphere then means that in the $\boldsymbol{R}'$-sphere the combination of $\tilde{\phi}_{R'L'}$ and $\dot{\tilde{\phi}}_{R'L'}$ (specified by the o matrix) must have the logarithmic derivative equal to $+l'$.

---

‡ We recall that a logarithmic derivative is defined thus: $D[f(r)] \equiv \dfrac{\partial \ln f(r)}{\partial \ln r} = r\dfrac{f'(r)}{f(r)}$.

* We note that the radial behaviour of J^0s and N^0s is necessarily related to the radial behaviour around $r=0$ of solutions of the single-site problem discussed after (3.4). There, at the origin both the potential and the value of energy did not play any role, while in the current situation V_0-E is zero everywhere. Hence, J^0s and N^0s behave, respectively, as r^l and r^{-l-1} also everywhere.

¶ The letters J, N denoting the two types of solutions are not chosen arbitrarily, but are derived from the usual letters used for the Bessel and Neumann spherical functions (j and n), of which J^0s and N^0s are zero-energy limits. [It might be of interest to note that while js are all regular at infinity, their (renormalized!) zero-energy limits, J^0s, are divergent there; basically, $J^0_l(r) \propto \lim_{\kappa\rightarrow 0}\kappa^{-l} j_l(\kappa r)$.]

§ The expansion theorem is valid for general spherical Bessel and Neumann (and Hankel) functions, not just for their limits J^0 and N^0 used by us. In this regard, the very same theorem enters the formalism of energy-dependent MTO formalism. [An analogical expansion expressing a basis function centred at one site in terms of basis functions centred at another site exists also for polynomials multiplied by gaussians ($r^\beta e^{-\alpha r^2}$) and Slater-type orbitals ($r^\beta e^{-\alpha r}$). See also footnote 1 of Section 15.2 in (Mart04).]

† With definitions (3.23) and (3.24), it can be seen from the explicit form given as equation 2.9 in (Ture97) that S^0 is a real symmetric matrix.

The important message the previous few paragraphs were meant to convey is that the information needed for the construction of the matrix H^θ is contained in i) the (radial) logarithmic derivatives of $\tilde{\phi}_{Rl}$ and $\dot{\tilde{\phi}}_{Rl}$ at the spheres' surfaces, ii) the canonical structure-constant matrix S^0 (it does not play a role in matching the logarithmic derivatives but is important in matching the functions in value), and iii) the spheres' radii, s_Rs. In short, there is a mapping

$$\{D,\dot{D},S^0,s\} \;\rightarrow\; H^\theta, \tag{3.26}$$

where $D \equiv \{D[\tilde{\phi}_{Rl}(r=s_R)] \;\forall\; Rl\}$ and $\dot{D} \equiv \{D[\dot{\tilde{\phi}}_{Rl}(r=s_R)] \;\forall\; Rl\}$. Instead of the set $\{D,\dot{D},s\}$ one may use quite a few other sets of parameters. One such, with a physically motivated background,‖ is the set of so-called *standard LMTO potential parameters* $\{C,\Delta,\gamma\}$. These are all Rl-diagonal matrices, and their relation to $\{D,\dot{D},s\}$ can be found as equation 2.84 in (Ture97). Using the new set, the Hamiltonian matrix H^θ can be written thus:‡

$$H^\theta \;=\; C \;+\; \sqrt{\Delta}\, S^\gamma \sqrt{\Delta}\,, \qquad S^\gamma \equiv S^0(1-\gamma S^0)^{-1}. \tag{3.27}$$

With a prescription for the Hamiltonian matrix (including the knowledge to calculate its parameters from the logarithmic derivatives) one may theoretically set up a selfconsistency loop for finding a solution (with an energy close to E_ν), much along the lines of the paragraph after (1.10): We start with some estimate of the potential within spheres, calculate $\tilde{\phi}$s and $\dot{\tilde{\phi}}$s, from these we determine the C, Δ and γ parameters, find the eigenvalues and eigenvectors of (3.27), calculate electron densities in the spheres, recalculate potentials in the spheres, if needed adjust E_ν, and start over.*

The raw programme outlined in the previous paragraph has deficiencies: i) it represents an infinitely large problem when the number of sites is infinite, ii) the S^γ matrix entering the Hamiltonian matrix has unnecessarily long range (although much shorter than S^0), and iii) it is not suitable for dealing with disorder.

The first point cannot, of course, be solved in general, however, e.g., if the system features periodicity, one may deploy the Fourier transform, classify solutions by a $\boldsymbol{k}$ vector in the reciprocal space, and thus for each $\boldsymbol{k}$ decrease the size of the eigenvalue problem. In the case of a three-dimensional (3D) periodicity, the size of the problem is given by the number of atoms in the basis, n_b, times the number of orbitals used per atom; if *spd*-like orbitals are used, the size of the Hamiltonian matrix for each $\boldsymbol{k}$ is $18n_b \times 18n_b$. In layered systems, on the other hand, we may only count with a two-dimensional (2D) periodicity along layers, while in the direction across layers the periodicity is lost, see Figure 3.3. We may still introduce a reciprocal vector $\boldsymbol{k}_\parallel$ parallel to the atomic layers and do the Fourier transform in the *xy*-plane, by means of which we effectively, for each $\boldsymbol{k}_\parallel$, reduce the eigenvalue problem to an infinite one-dimensional (along the *z*-axis) system. Finally, we will only be interested in systems in which the number of atomic layers in the 'active'—or sample—part will be finite, while the semi-infinite leads attached to the sample possess periodicity also along the *z*-direction. It is then reasonable to expect that the electronic structure well within leads is bulk-like and can be precalculated with a program expecting a full 3D symmetry. If we then add several layers from the lead material (not necessarily the same on the two sides) to the ends of the central part, we may take the electronic structure fixed and bulk-like in the entirety of the leads and let the electronic-structure program establish only the parameters in the central part, relying on the few added lead-material layers being enough to appropriately relax and connect realistically to the fixed leads' electronic structure. (Of course, one may always check things are all right by trying to add or remove several such buffer layers and ensure the sample characteristics do not change.) By these steps we happily arrive at a finite-sized problem, although considerably bigger than in a simple bulk case.

‖ The physics behind these parameters is well described in (Skri84). There the parameters are shown to have a direct relation to a bandstructure of periodic solids.

‡ It might be of value to note that the structure-constant matrix S^0 is the only object in (3.27) that actively uses the magnetic quantum number m.

* For more details see Section II.6 of Andersen, Jepsen and Šob's contribution in (Yuss87).

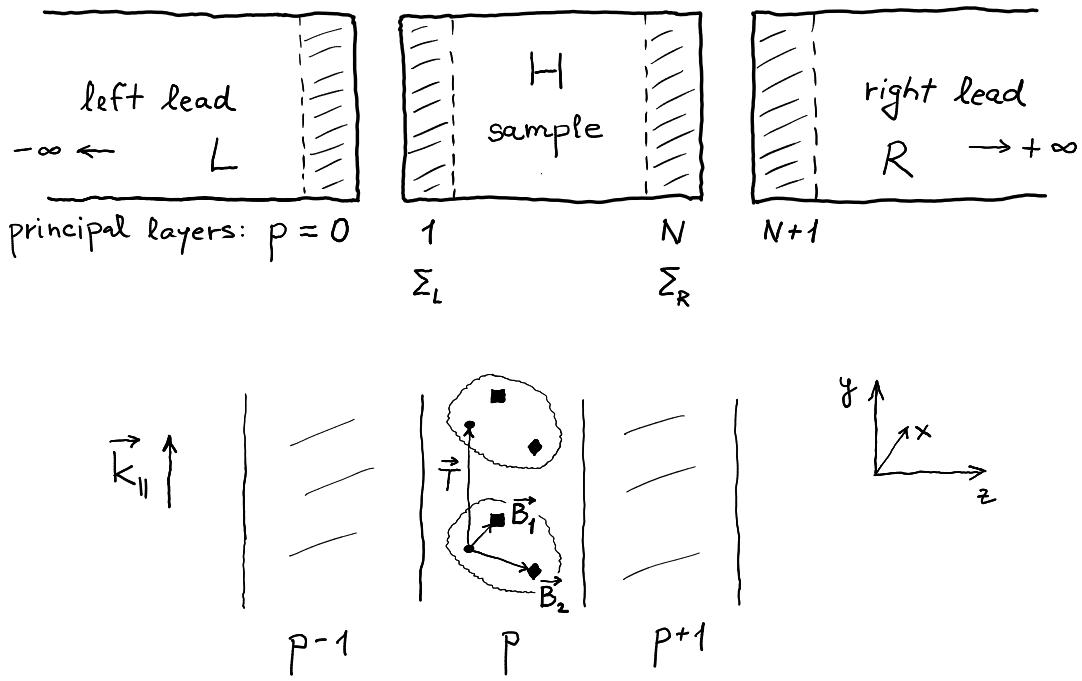


Figure 3.3: A sketch of a layered system, which consists of left and right semi-infinite 3D-periodic leads, and a central part called a sample. Indicated is the grouping of atomic layers into principal layers, *p*s, discussed in the main text. Each atomic (hence also principal) layer is supposed to have a 2D periodicity (in the *xy* plane), which is used in classifying wavefunctions with a $\boldsymbol{k}_{\parallel}$ vector. Position of an atom in the structure can be specified by indicating its principal layer *p*, the translation (2D) vector $\boldsymbol{T}$ of the 'primitive' cell the atom is in, and the position of the atom within this cell, specified by a vector $\boldsymbol{B}$; in short, $\boldsymbol{R}\equiv p\boldsymbol{TB}$. In the given example, the primitive cell of the principal layers contains three sites.

The second mentioned deficiency is related to the extent of LMTOs, which is simply determined by the decay characteristics of the envelopes. For the unscreened orbitals the envelopes, N^0s, decay fairly slowly as r^{-l-1}, which is essentially reflected in the slow decay of the canonical structure constants, $S^0_{R'L',RL} \propto (\frac{w}{|\boldsymbol{R'}-\boldsymbol{R}|})^{l+l'+1}$. For the θ orbitals (in terms of which H^θ represents the Hamiltonian), however, the situation is already different. The envelopes of these orbitals are not simply the N^0 functions, but their combination specified by the Π^{-1} matrix, as is clear from (3.15). It can be argued that the extent of θ orbitals is smaller than that of the original LMTOs (χs). Thus, by making linear combinations of the original unscreened long-range envelopes we may obtain another set of envelopes, and from-these derived orbitals, with a shorter range. The nearly-orthogonal θ orbitals are just one example of this. While the original envelopes have a pole only in the origin of their mother sphere, making their combination necessarily introduces poles also in other spheres (which were used in the combination). This is, however, not important, because as an envelope enters any sphere, it is replaced by a proper $\phi,\dot{\phi}$ combination. Different combinations of the original envelopes show different screening properties, and usually an effective exponential screening is possible (in closely packed structures), in which envelopes beyond 2nd-nearest neighbours can be neglected. We arrive at an *ab initio* tight-binding (TB) picture.

The 'screening' procedure was originally, for the $\kappa^2 = 0$ case, inspired by electrostatics, where the potential generated by a single charge and higher multipoles has the same form as the N^0_l functions (the electrostatic field satisfies the Laplace equation everywhere but for points where the charges are), (Ande84). By surrounding e.g. a single charge of strength one by charges with an opposite sign and total charge of minus one, the total field (still satisfying the Laplace equation everywhere but for points where the charges are) will decay more quickly (at worst as r^{-2}) at larger

distances than originally (r^{-1}).

Although originally derived for the $\kappa^2 = 0$ situation, screening is not limited to this case.

Screening of envelopes may be achieved in many different ways.¶ Inspired by electrostatics one may require that expansion of the new envelopes around foreign sites should not only include the regular J^0 functions (which by themselves, being regular, cannot account for any 'screening charge at a foreign site) but also some portion of the irregular, N^0, solution. It is then fairly natural to define a *screened function* J^α thus:

$$|J^\alpha_{RL}> \equiv |J^0_{RL}> - |N^0_{RL}>\alpha_{Rl}. \tag{3.28}$$

The original unscreened envelope, $||N^0>$ can be using (3.25) and ignoring the interstitial part be written as

$$||N^0> = |N^0> - |J^0>S^0. \tag{3.29}$$

Expressing $|J^0>$ from (3.28) and substituting it into the last equation we get

$$||N^0> = |N^0> - (|J^\alpha> + |N^0>\alpha)S^0 = |N^0>(1-\alpha S^0) - |J^\alpha>S^0. \tag{3.30}$$

If $(1-\alpha S^0)$ has an inverse, then

$$||N^0>(1-\alpha S^0)^{-1} = |N^0> - |J^\alpha>S^0(1-\alpha S^0)^{-1}, \tag{3.31}$$

which, by defining *screened envelopes* $||N^\alpha>$ and a *screened structure-constants matrix* S^α

$$||N^\alpha> \equiv ||N^0>(1-\alpha S^0)^{-1}, \quad S^\alpha \equiv S^0(1-\alpha S^0)^{-1}, \tag{3.32}$$

can be rewritten as

$$||N^\alpha> = |N^0> - |J^\alpha>S^\alpha. \tag{3.33}$$

From (3.32) we see that the new envelopes are just a linear combination of the old ones. The existence of the inverse of $(1-\alpha S^0)$ further ensures that the new envelopes span the same space as the old ones. By considering the boundedness of so-called canonical bands, see e.g. (Skri84), it can be argued that sufficiently small αs always lead to an invertible $(1-\alpha S^0)$.†

Different αs lead to envelopes with different screening. The task is to find such values that the screening is efficient, i.e., that the range of the S^α is small. As already mentioned, values can be found, see (Ande84), for which matrix elements of S^α are negligible beyond the second nearest neigbours. Then S^0 can be expressed in terms of α and S^α, (3.32), and the orthogonal Hamiltonian (3.27) can be written as

$$H^\theta = C + \sqrt{\Delta}\, S^\gamma \sqrt{\Delta}, \qquad S^\gamma \equiv S^\alpha[1+(\alpha-\gamma)S^\alpha]^{-1}. \tag{3.34}$$

First, we see that the nearly-orthogonal representation is a special case of screening, with αs set equal to the γ parameters. Second, the orthogonal Hamiltonian may now be expanded as a power series in S^α and, in this way, when using αs leading to the shortest range, we arrive at an approximation to H^θ in which the two-centre objects (i.e. S^αs) have conveniently short range. More details on this can be found in section 3.2.3 of (Ture97).

Eventually, the last mentioned deficiency, that the formalism is not prepared to handle disorder (essentially because we work with wavefunctions and these have no simple 'averaging properties') is not really of great importance to us here as we will not consider any disorder in this thesis. Nonetheless, the codes actually used to carry out our calculations have evolved from codes created for bulk calculations and the latter allow for the Coherent-Potential-Approximation (CPA) dealing with (substitutional) disorder. This means that the codes do not operate with Hamiltonians and global wavefunctions, but instead with (one-particle) Green functions, which have the needed averaging properties. Putting the usual form of disorder aside, Green-function formalism represents also a

¶ See, e.g., the discussion about equation 3.24 of Andersen et al.'s text in (Yuss87), where they discuss particularly why α must be a diagonal matrix in the LMTO theory.

† Basically, if αs are smaller than the reciprocal value of the biggest eigenvalue of S^0, the inverse exists.

convenient way for dealing with layered structures. This we will now briefly indicate, with more details to be found in Section 3.4 (and more generally in the whole Chapter 3) of (Ture97).

We define the *Green function (matrix)*

$$G(z) = (z - H^{\theta})^{-1}, \tag{3.35}$$

where z is a complex parameter. Apparently, the Green function and the Hamiltonian carry principally the same information, but the extra dependency of Green functions on z let us express certain quantities more easily compared to working just with the Hamiltonian. For example, the charge density $\rho_R(\boldsymbol{r})$ inside the $\boldsymbol{R}$-sphere can be expressed as

$$\rho_R(\boldsymbol{r}) = \sum_{LL'} \int_{-\infty}^{E_F} \mathrm{d}E\, \tilde{\phi}_{RL}(\boldsymbol{r},E)\, n_{R,LL'}(E)\, \tilde{\phi}_{RL'}(\boldsymbol{r},E),$$

where $n_{R,LL'}(E)$ is the local density of states matrix at energy E and can be given in terms of the *on-site* block of the above-introduced Green function matrix

$$n_{R,LL'}(E) = -\frac{1}{\pi} \mathrm{Im} G_{RL,RL'}(E + i0).$$

Worth noticing about this expression is that local properties of a system are encoded in local characteristics of the Green function (which nonetheless incorporates effects from the whole system). This observation has a more general validity. Thus, when investigating properties of a sample sandwiched between the leads in a layered system, Fig. 3.3, we are mainly interested in Green-function matrix elements localized in the sample. Although the parts outside the sample—the left, L, and right, R, leads—do contribute to the $RL, R'L'$ matrix elements of G within the sample, their effects can be included via so-called selfenergies Σ_L, Σ_R thus:

$$G(z)_{\text{sample}} = (z - H^{\theta}_{\text{sample}} + \Sigma_L + \Sigma_R)^{-1}. \tag{3.36}$$

All objects in this equation are restricted to the sample and have finite sizes. The technique based on selfenergies is called partitioning, is thoroughly described in the given reference, and will also be justified from a different perspective in the chapter discussing transport. For calculation of self-energies we use the renormalization-decimation technique described in Section 10.6.2 of (Ture97).**

Finally we note, that from the codes' perspective it is in certain moments advantageous to work with a so-called *auxiliary Green function (matrix)* instead of the Green function matrix introduced above. Its definition looks like

$$g^{\alpha}(z) = [P^{\alpha}(z) - S^{\alpha}]^{-1}. \tag{3.37}$$

Here $P^{\alpha}(z)$ is a so-called potential function, represented by a diagonal matrix, and all objects are in the screened α-representation. Without going into details here about meaning of g and P, we only want to point out that for calculation of the indicated inverse it is of great numerical importance that the S^{α} matrix is short-ranged. If we introduce a conceptual division of the structure into so-called *principal layers*, as sketched in Fig. 3.3, defined so that matrix elements of S^{α} are negligible between sites across one principal layer and further, the S^{α} matrix is then tri-diagonal in the principal-layer indices, and for such matrices efficient numerical procedures for calculating an inverse exist.

** During our work we several times encountered difficulties which the renormalization-decimation technique. It turned up that if one is unlucky, one may during the repetitive procedure of the technique happen to be trying to invert a matrix with one very small eigenvalue, which then completely spoils the result. This was usually manifested by not having all the eigenvalues of the resulting $\mathrm{Im}G^r$ negative. We have not found other solution than to use sufficiently large 'infinitesimal' ε and check at each step that the signature of $\mathrm{Im}G^r$, see Appendix E, is all right.

3.3. Inclusion of the Spin–Orbit Interaction

So far in this chapter, having been basically based on a nonrelativistic formalism, we have virtually ignored the role of relativity and spin, and how it may complicate things, since spin and orbital moment are not independent as soon as relativistic domain is entered. On the other hand, in the previous chapter we derived the full-fledged relativistic Kohn–Sham–Dirac Hamiltonian incorporating one-particle relativistic effects exactly (the quality of treatment of many-particle effects depends on the form of the chosen exchange-correlation functional). The latter formalism, using the four-component (bispinor) notation, is rather heavy, and although regularly required for correct description of core electrons, it seems often unnecessarily complicated for valence electrons, especially for light atoms we consider.

In (Ture08) the authors review—starting from the Kohn–Sham–Dirac Hamiltonian—, a simplified treatment of relativistic effects including the spin-orbit (SO) interaction within the TB LMTO method, carried in a two-component, non-relativistic-like spinor formalism using an orbital basis set of the so-called scalar-relativistic (SR) approximation (SRA), (Koel77). In the SR approximation spin and orbital movement are decoupled and conserved, i.e., notably, spin stays a good quantum number. Yet, relativistic mass-velocity enhancement and the Darwin term are included in the approximation. As mentioned in (Ande75, Section V), see also (Chri84), for valence electrons there is usually an order-of-magnitude energy separation between strength of scalar-relativistic effects and the SO interaction. This effectively allows us to treat the SO interaction perturbatively on top of a SR calculation. The reference (Ture08) then shows that incorporation of SR effects does not modify the non-relativistic formalism (see equation 13 therein, which is identical to our (3.27)). Furthermore, the Hamiltonian including the SO interaction can be represented in the orthonormal LMTO-SRA basis, $|\theta_{RLs}>$, by a slight modification of (3.27), or (3.34) matrix

$$\overline{H}^{\theta} = \overline{C} + \sqrt{\Delta}\, S^{\gamma} \sqrt{\Delta}, \qquad S^{\gamma} \equiv S^{\alpha}[1+(\alpha-\gamma)S^{\alpha}]^{-1}, \qquad \overline{C} \equiv C + \overline{\xi}, \tag{3.38}$$

where the matrix $\overline{\xi}$ has elements

$$\overline{\xi}_{RLs,R'L's'} \equiv \delta_{RR'}\,\delta_{ll'}\,\xi_{Rl,ss'} < Ls|\boldsymbol{L}\cdot U^{+}\boldsymbol{S}U|L's'> \tag{3.39}$$

and the parameters ξ denote atomic-like SO parameters which are given by

$$\xi_{Rl,ss'} = 2\int_0^{s_R} \mathrm{d}r\; r\phi_{Rls}(r)\phi_{Rls'}(r)[\omega_R^{-1}(r)]', \quad \omega_R(r) \equiv 1+c^{-2}[E_0 - V_R(r)]. \tag{3.40}$$

In the last equations we introduced several new objects. First, we should note that the spin index $s=\uparrow,\downarrow$ is a 'local' index meant with respect to a site-local quantization axis chosen parallel to the effective magnetic field at the site. Theoretically, for a system with non-collinear magnetism, this axis can point in different directions for different sites. In our systems, which only feature collinear magnetism, this axis is common to all sites, though it still may have an arbitrary orientation with respect to a global coordinate system, which is suitably chosen along crystal axes. We use the global system to express components of the orbital variables ($\boldsymbol{L}$), and the U matrix in (3.39), acting only on the spin indices, is used to express the local spinors in the global frame so that the $\boldsymbol{S}$ operator can be taken as halves of the Pauli σ matrices. Explicitly, $U=\exp(-i\phi\sigma_z/2)\exp(-i\theta\sigma_y/2)$, with $\boldsymbol{n}\equiv(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$ being components (in the global system) of a vector pointing along the magnetization. Second, the ϕs appearing in (3.40) are the radial parts of the regular solutions inside individual spheres (normalized there to one). Third, the ω factor, of whose inverse radial derivative appears in (3.40), is called a mass-enhancement factor. It is clearly large close to the atomic nuclei. The exact value of E_0 is unimportant but is to lie within the valence band we want to study.

It is noteworthy to realize that (3.39) bears similarity with the often used simple form '$\xi\boldsymbol{L}\cdot\boldsymbol{S}$', however, is more general.

Looking at (3.38) we see that the step from a SR Hamiltonian to a SR+SO one is realized by replacing an originally *RLs*-diagonal C matrix with a matrix $\overline{C}$ that stays *Rl*-diagonal but intermixes the magnetic (*m*) and spin (*s*) quantum numbers.

In all our calculations we selfconsistently calculate electronic structure in the SR approximation and at the end of this calculation we determine the ξ parameters.† Then, we use thus determined parameters in our transport code, which as if works with the modified Hamiltonian (3.38) (but utilizes the Green-function language).

† It would be possible to modify the selfconsistent code to take the SO interaction also into account at each step, but this would require using larger matrices leading to a certain slow-down. We didn't take this route.

4

Electronic Transport

4.1. General Grounds

Discovery of the electron by J. J. Thomson in 1897 suggested a mechanism behind electrical and thermal conduction in metals: transport of electrons. The first theory of conductivity, by P. Drude in 1900, was designed in analogy with the successful kinetic theory of gases. A metal was pictured as a gas of conduction electrons that move and scatter against a background of heavy immobile ions (this is, however, different from the molecules of an ordinary gas, which only collide with themselves). An excellent discussion of the model is given in (Ashc76). Although the theory is fully classical, both Ohm's law (known from 1826) and the Hall effect (discovered in 1879) can be described, using just a few material-specific quantities, such as the electron density n and the electronic scattering rate $1/\tau$. The Drude-theory expression for the frequency-dependent (ω) conductivity σ of a material in an electric field E reads*

$$\boldsymbol{j}(\omega) = \sigma(\omega)\boldsymbol{E}(\omega), \qquad \sigma(\omega) = \frac{1}{1-i\omega\tau}\frac{nq^2\tau}{m}, \tag{4.1}$$

with m and q being the mass and charge of the carriers, respectively. Quite reasonably, the most valued Drude-theory predictions are those independent of τ. For instance, the high-frequency ($\omega\tau \gg 1$) limit of the last expression, $\sigma(\omega) \approx i\frac{nq^2}{m\omega}$, leads to a dielectric constant $\varepsilon(\omega) = 1-(\omega_p/\omega)^2$, with $\omega_p^2 = \frac{nq^2}{\varepsilon_0 m}$, and predicts the existence of the plasma frequency ω_p, above which the material becomes transparent, in a rather good agreement with experiments at least for alkali metals (see table 1.5 in Ashc76). Also, the expected Hall coefficient, $R_H = 1/nq$, is τ-independent and again match the experimental findings in certain cases. On the other hand, some other conclusions drawn from the Drude theory, mainly concerning thermal behaviour,† are, due to the used Maxwell–Boltzmann statistics (the only known at the time), wrong by orders of magnitude. This was resolved during the years 1926–28 by W. Pauli and A. Sommerfeld, who replaced the statistics of the Drude model with the Fermi–Dirac one.‡ However important this step was for clarification of the heat-capacity mystery besetting the Drude theory, a noteworthy point is that the statistics does not enter the calculation of electrical conductivity in the Drude (and thus nor in the Sommerfeld) theory, and the two theories give the same result (4.1).

---

* Here we assume that the current and the electric field are parallel, which can be, of course, generalized. Further, we use the usual complex notation which is suitable for expressing phase relations.

† The Drude theory predicts $3/2\,k$ (i.e. temperature independent) contribution per electron to the heat capacity, however, measured heat capacity at low temperatures drops to zero.

‡ The advent of quantum mechanics (1923–25, most notably through the work of L. de Broglie, E. Schrödinger and W. Heisenberg) lead to the formulation of the exclusion principle for electrons by W. Pauli (1925), which was then extended by E. Fermi (1926) to give the general formula for statistics of noninteracting electrons. Fermi also noticed the analogy with the Bose-Einstein statistics (1924).

The Sommerfeld free-electron model still makes many predictions unambiguously contradicted by observation. For a longer list, see (chapter 3 of Ashc76). We only point out the simple fact that not all elements are metals, explanation of which is already beyond the theory's powers. This is another moment where quantum mechanics is needed (the first time it provided the Fermi–Dirac statistics). In 1928 F. Bloch showed that electrons can freely propagate through a perfectly periodic potential (created e.g. by ions sitting on crystal lattice sites) and formulated the concept of allowed energy bands and gaps between them.¶ Subsequently, A. H. Wilson in 1930 laid down the foundations for the classification of crystals into metals, semiconductors, and insulators.

Interestingly, in spite of the necessity to employ quantum mechanics to treat the effect of the lattice potential, it is possible to sustain the Drude–Sommerfeld classical picture of particle-like carriers a bit further. We can use quantum mechanics only to provide a relevant particle dispersion law $E = E_n(\boldsymbol{k})$ (n is the band index and $\boldsymbol{k}$ the carrier crystal-momentum), and replace with it the quadratic dependence of the Sommerfeld (free-electron) model. This combined treatment, in which some aspects are coped with quantum-mechanically while others are treated almost classically is called the semiclassical approach. Figure 4.1 illustrates the situation, showing also essential conditions under which such an approach can be used.

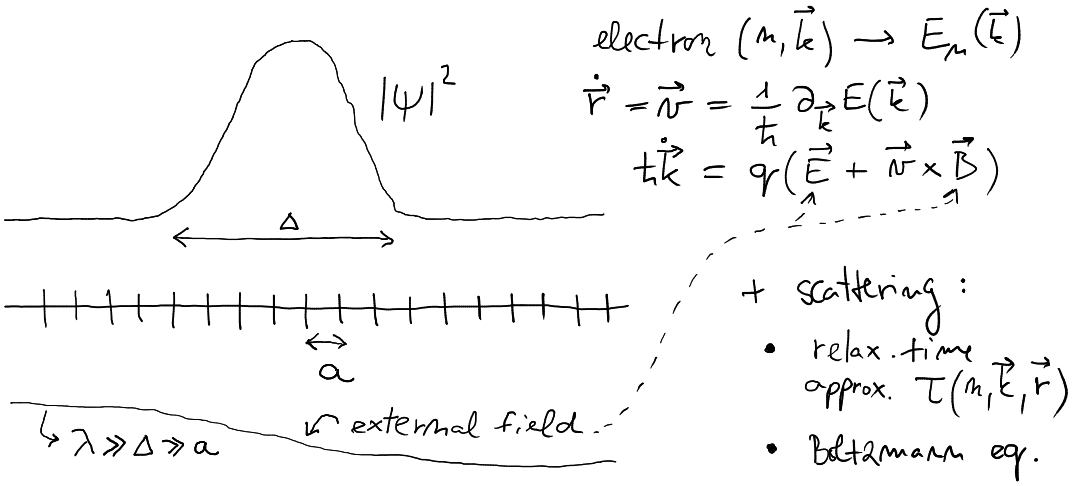


Figure 4.1: Semiclassical view on the movement of an electron in a crystal. An electron is described by a wave packet of Bloch states. If the spread of the wave packet is to be small in the *k*-space with regard to the Brillouin zone (only then labelling an electron with a crystal momentum $\boldsymbol{k}$ is meaningful), the electron must be spread in real space over many primitive cells. The semiclassical model describes response of such electrons to externally applied electric and magnetic fields that vary slowly over the dimension of an electron wave packet. Then the classical equations of motion are justifiable. (Note that the lattice potential varies quickly and therefore must be dealt with on the quantum level.) Scattering by any disorder can be allowed for by e.g. a relaxation-time approximation or a more sophisticated Boltzmann kinetic equation.

The semiclassical approach will soon show insufficient for our goals. Still a few points are noteworthy. Thanks to Bloch, we know that a perfectly periodic crystal is transparent to electrons, and thus has infinite conductance.* Finite conductance of a material must thus be caused by some deviation from this periodicity (scattering by lattice disorder, e.g. vacancies, impurities, at finite temperature phonons) or by other mechanisms that can cause electron scattering (e.g. electron-electron interaction). On the simplest level these effects can still be to some extent

¶ In the Appendix B we briefly note a common misconception about Bloch's theorem.

* This statement should be understood in this way: If an electron propagates through a crystal, it will continue to do so even if the electric field is zero. We thus have a nonzero current at zero external field, obtaining formally the infinite conductance through the relation $\boldsymbol{j} = \sigma \boldsymbol{E}$. We emphasize this view, since if one tried to study an electron movement in an ideally clean crystal under the influence of a finite external electric field, one would find out that the electron's movement is actually periodic in space and the macroscopic current would be zero.

phenomenologically described by the above introduced parameter τ, however now with a rethought meaning (originally the ion lattice was believed to be the main source of scattering). But having now a richer theory (from the dispersion law $E_n(\boldsymbol{k})$ we can infer that an electron movement may be different in different directions) this single relaxation parameter can be generalized to depend on $\boldsymbol{k}$. Although better, such a description still suffers from the actual non-existence of a way to determine $\tau(\boldsymbol{k})$ as well as from being blindly a one-electron theory in the sense that an electron scattering is in no way influenced by the state of other electrons (not even on the level of statistics; collision output is not dependent on the situation before the collision; only the tendency towards local equilibrium is enforced, with speed $\propto 1/\tau(\boldsymbol{k})$). Thus another step was made leading finally to the use of the famous Boltzmann kinetic equation,‡ with a quantum-mechanically determined collision term. This theory for example can (as opposed to the relaxation-time theory even with arbitrary $\tau(\boldsymbol{k})$ dependence) account for observed deviations from the Wiedermann–Franz law.§

Despite the successes of the above theories, it eventually showed up that there are situations in which the semiclassical view breaks down and the full quantum-mechanical treatment is inevitable; usually this happens when a system is small enough. The language of the above theories heavily relies on the particle-like behaviour of electrons, while quantum-mechanically they also may display wave-like characteristics. There are experiments showing double-slit-like interference of electrons in small ring geometries, other experiments showing departure from the classical Hall-voltage linear dependence on magnetic field when the field is strong (formation of Landau levels), yet others revealing exotic localization phenomena, or apparently random dependence of the resistance of a given sample on magnetic field, reproducible in a given sample although differing from sample to sample (another example of quantum interference), etc. There is a review describing some of these effects, (Wash92).

The decision about what description would be adequate for a particular system is usually based on several characteristic scales, see (Datt97, Chapter 1). In our situation we will deal with structures for which we assume that a completely coherent transport is possible through the whole system, with all scattering events (within the sample) being elastic. This already almost fully disqualifies the semiclassical theories and forces us to use a full quantum-mechanical treatment, based essentially on electron wavefunctions or Green functions, i.e. on objects capable of embracing interference (unlike simpler distribution functions of e.g. Boltzmann theories).

The main sample characteristics we calculate in this thesis is its conductance, *C*, (or equivalently resistance, *R*) *at zero bias*. (This information, of course, does not allow us to discuss any $I-V$ characteristics, but still describes the sample at least in the limit of zero applied voltage.) We will moreover limit ourselves to considering transport only within a single-electron picture, and will assume that the single-particle states in the Kohn–Sham formulation actually describe these independent single electrons. (There can be no hope such a simplistic model can catch all the essential physics, however, in analogy to standard band-structure calculations, in which the Kohn–Sham single-particle eigenvalues and eigenstates can be successfully interpreted as elementary excitations, we apply the same procedure to electronic transport.) We note at this place, that single-particle description is inappropriate when many-body effects become important, such as in systems featuring the Kondo effect or Coulomb blockade. For our systems we believe the single-electron picture is appropriate.

To calculate conductance, several theoretical approaches were invented. Among the first not semiclassical to be used were so-called linear-response theories, based essentially on the work of Kubo, (Kubo57) and Greenwood, (Gree58), which treat the current as a response of a quantum system to the applied field and determine the linear coefficient of their dependence; see also a nice article by Lax (Lax58). On the other hand, in the recent twenty years, in a certain sense

‡ The equation itself had already been around for some time (L. Boltzmann committed suicide in 1906). But the correct collision term needs to be evaluated by the quantum-mechanical scattering theory.

§ The Wiedermann–Franz law proposes that the ratio of thermal (κ) to electrical (σ) conductivity is proportional to temperature (*T*). In other words: $\frac{\kappa}{\sigma T}$ is a constant called a Lorentz number.

complementary viewpoint of Landauer, (Land57, Land87, Land88), and expanded later by Buttiker (Büt86, Büt88), has been gaining on popularity. Perhaps a nice introduction to the Landauer-like picture is given in (Imry99) and (Wort06), and, of course, the two books by Datta, (Datt97, Datt07). There is also a third, more advanced and general, method, related to both the linear response and the Landauer−Büttiker formalism, which uses the so-called Non-Equilibrium Green Functions (NEGFs). The formalism of NEGFs seems to provide a rigorous framework for the development of quantum device models, and provides a bridge between fully coherent transport models (transmission formalism of Landauer) and fully incoherent transport models (based on the Boltzmann equation); also, the method is not limited to the zero-bias limit. This last method was actually used, in essence, already by Caroli *et al.* in the 1970s, (Caro71). Lastly, we think definitely worth mentioning is also the work by Todorov, (Todo02), who uses still a slightly different, so-called tight-binding, transport formalism.

The method of choice in this thesis, described in detail below, will be a mixture of the Landauer transmission formalism and the formalism of NEGFs, basically inspired by and learnt from the mentioned books by Datta. Before heading off in this direction, though, we briefly comment on the linear response approach.

4.2. Linear Response

Linear response uses a first-order perturbation theory to calculate effect of a weak electric field on the density matrix of a system, from whose change the response in the current is determined. We provide derivation of some of the related expressions in Appendix A. In spite of regularly providing relevant results, the linear response theory, in the way it is usually formulated, contains certain awkward aspects. As, e.g., Imry and Landauer put it, (Imry99), linear response is usually derived within the Schrödinger-equation framework, which is ‘a theory of conservative systems, does not allow for dissipation, and some supplementary handwaving is needed to calculate a dissipative effect such as conductance’. Nonetheless, linear response probably can be put on firm ground using a stochastic theory of many body systems, (Haya06). Also importantly, a connection between linear response theories and the Landauer−Buttiker formalism was shown to exist, starting in work by (Fish81), and then improved on in (Ston88, Bara89, Mavr04).

Our Appendix A provides equations which form a starting point in Appendix A of (Cré01). The authors of the latter derive many other linear-response equation forms used in literature.

4.3. Landauer Approach to Transport

‘Response’ theories generally consider an applied electric field as a causative agent and the resulting current as the response to it. In this view one typically applies a voltage source across the specimen and measures the (total) established current. To relate the voltage and the field, often one assumes that the field produced by the source is homogeneous within the specimen (i.e., $E = U/L$, U being the source voltage and L the specimen’s length). While this may certainly be almost valid (at least when E is averaged over distance comparable to the typical inter-scatterers distance) for devices with constant density of scatterers, it is evidently wrong for, e.g., a clean device with a single scatterer somewhere at its centre. In situations like the latter (where it’s difficult to a priori deduce the field profile), it is perhaps more suitable to look at the experiment from another angle, as proposed by R. Landauer in 1957, (Land57).

Landauer’s view is that transport *through* a specimen is a result of carriers flow incident on the specimen *boundaries*. The voltage (or, equivalently, field) distribution within the sample then results from the self-consistent pile-up of the carriers: If there are no scatterers, there is no charge pile-up, no field, and finally no resistance. On the other hand, introduction of scatterers results in charge gradients within the device (there is a higher concentration of carriers in front of as compared to behind a scatterer), producing fields opposing further charge pile-up, finally establishing some non-trivial equilibrium.

Landauer’s picture is essentially a scattering problem. In its simplest form, see Fig. 4.2, we have a source of electrons, a metallic reservoir, which via a perfect lead supplies electrons that then

impinge on the sample. Some electrons get reflected back and re-appear in the original reservoir, others make their way to the other side of the specimen, leave via another perfect lead, and are absorbed by a second reservoir. In equilibrium the electrons move in both directions in equal amounts. To get a non-zero total current the reservoirs may be biased with a voltage source. Since the wiring from the battery is contacted onto the reservoirs, the reservoirs are regularly also called simply contacts. The one with excess of electrons is called a source, the other a drain.

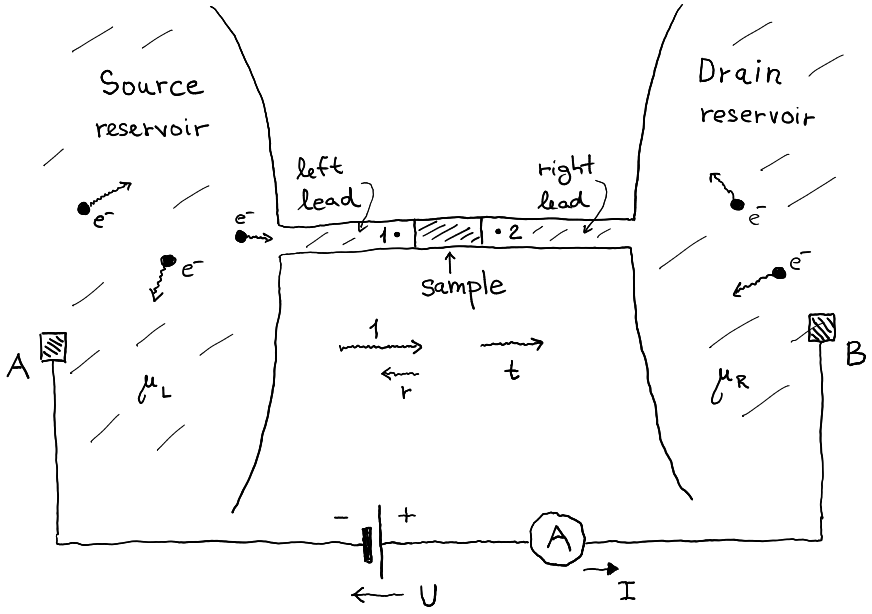


Figure 4.2: A schema of a simple resistance-measurement experiment. A battery biases the system, maintaining the electrochemical potential of the left reservoir, μ_L, higher than of the right reservoir, μ_R, $\mu_L = \mu_R - qU$ (q is a carrier's charge, i.e., it is a negative number here). The battery, wires to the reservoirs and reservoirs themselves are macroscopic (sizes $\gtrsim$ millimetres), the leads and the sample dimensions are much smaller, easily of submicron level, and typically counted in 10s of a nanometre.

Under the assumption that the leads are 'perfect', in the sense that any electron moving via them towards any reservoir enters the reservoir without being reflected back (such contacts are termed reflectionless), and also that there is negligible scattering in the leads, calculation of the total conductance C—of the whole set-up between points A and B, $C = I/U$—will only need (as we discuss later; see also section 2.2 in (Datt97)) characteristics of the sample itself and of the sample's coupling to the leads; thus a description of the situation between points 1 and 2. Notably, no further information about the (big and complicated) contacts is needed. This observation significantly reduces the necessary computational labour. To deploy it and be able to restrict our calculation only to the sample while correctly incorporating effects of the sample's coupling to its outside, the so-called embedding formalism, briefly described below, is suitable.

If we only had a from-everything disconnected sample, we could, within the one-particle approximation (which allows us to ascribe a separate wavefunction to each electron) express the time evolution of any sample's electron using the evolution operator U, $\psi(t) = U(t,t_0)\,\psi(t_0)$, and such evolution would be necessarily unitary—no electron can disappear from the sample or appear there if it has not been there already.

If we connect the sample to any other system, but keep looking just at the sample alone, two new phenomena occur: particles can escape from the sample, and other (if present) can enter it.

Focusing first on a particle originally in the sample, the part of its wavefunction in the sample diminishes with time, reflecting the particle's ability to leave the sample. The evolution, if the observation is limited to the sample's volume, is thus not generally unitary. The possibility to lose particles (in our case into the leads) will below be accounted for via use of the so-called (lead)

selfenergies, Σ_i, one for each lead.

On the other hand, as some particle is entering the sample, from the viewpoint of the sample it seems as if there was a particle 'being created' in the system. This effect will in our formalism be implemented via the so-called source(s).

Consider the case of a single electron flying in the left lead towards the sample. At the beginning its wavefunction in the sample is negligible, $\psi \approx 0$.‡ As soon as the particle starts entering the sample, a source localized in the sample (usually close to the left interface) is as if turned on and provides the 'transfer' of the particle across the interface into the sample. The part of the particle already in the system is then taken care of by the sample's propagator. Once all of what can enter the sample (not all of the particle's amplitude gets in, there is also reflection) has been 'transferred in', the source as if turns off. All what happens afterwards is conducted by the propagator: particle can travel to and fro within the sample, loosing its amplitude into the leads, eventually being most probably somewhere outside of the sample.

The process as just described sounds complicated, nonetheless natural. It's nothing but a description of a particle scattering against a (possibly complicated) barrier represented by the sample, expressed in the time domain, plus some extra complications brought about by our aim to watch the situation from within the sample only.

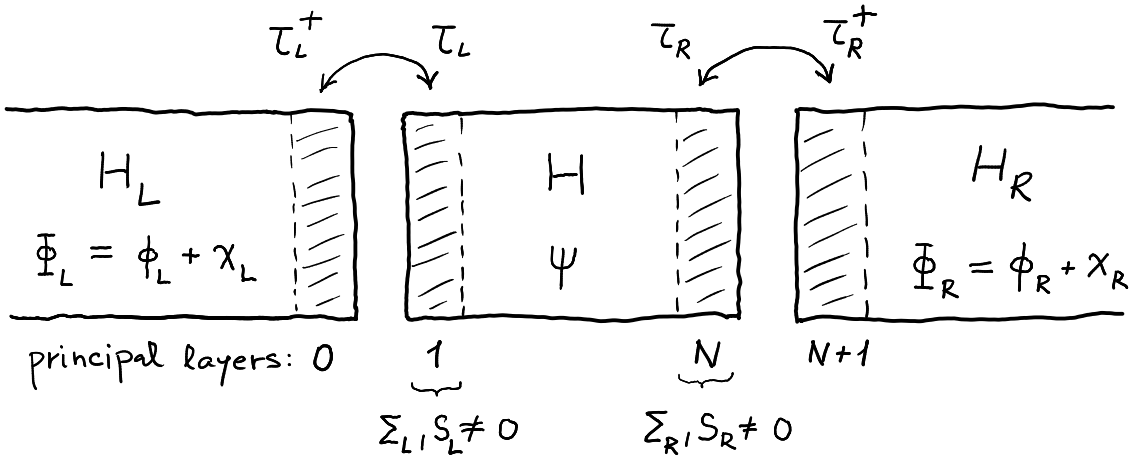


Figure 4.3: Conceptual division of the whole system into the left lead (index L), the sample (no index) and the right lead (index R). The parts are coupled together by the τ matrices with a limited real-space range (shaded areas) determining the size of the principal layers; for both leads τ directs into the sample, τ^+ out of the sample. Also shown are the areas of non-zero selfenergies $\Sigma_{L,R}$ and sources $S_{L,R}$ originating from the presence of the leads; their action is limited to the 1st and the N^{th} (i.e. the last) sample's principal layer.

Let us write down the mathematical description, following Fig. 4.3. The whole system consists of the left lead, the sample and the right lead, respectively described by Hamiltonians H_L, H, and H_R when separate. When coupled together by $\tau_{L,R}$ the time evolution of an electronic wavefunction is given by the Schrödinger equation

$$i\partial_t \begin{bmatrix} \Phi_L \\ \psi \\ \Phi_R \end{bmatrix} = \begin{bmatrix} H_L & \tau_L^+ & 0 \\ \tau_L & H & \tau_R \\ 0 & \tau_R^+ & H_R \end{bmatrix} \begin{bmatrix} \Phi_L \\ \psi \\ \Phi_R \end{bmatrix}, \tag{4.2}$$

‡ At this point one can see that it is not straightforwardly possible to get by only with some 'restricted' evolution operator (a propagator), but we really need sources. The usual equation $\psi(t) = G^r(t,t_0)\,\psi(t_0)$, $t > t_0$, does not hold here: without the sources we would, starting from 0, always obtain 0. This is because the restricted propagator is not aware of the outside–world particles and can reliably only propagate what is 'already in'. On the other hand, since the propagator can be informed about the mechanisms via which the system loses electrons—this 'static' knowledge is encoded in the selfenergies Σ_i which enter the calculation of the modified propagator—the propagator itself is sufficient to describe how

where the part of the wavefunction in the left lead, the sample, and the right lead has been denoted by Φ_L, ψ, and Φ_R, respectively.

In writing the above equation we have assumed there is not direct interaction between the leads. As was discussed in the section on electronic-structure calculation, in reality, as well as in our codes, the system consists of a sequence of atomic layers, which may be in the tight-binding representation mentally grouped into principal layers with the interaction then limited to the neigbouring principal layers at furthest. In the following we are based on such a tight-binding picture.

Looking at the equation for ψ,

$$(i\partial_t - H)\psi = \tau_L \Phi_L + \tau_R \Phi_R \equiv \tilde{S} \tag{4.3}$$

we see we might consider $\tilde{S}$ as a source for ψ.¶ However, such a source is not quite useful since $\Phi_{L,R}$ still depend on what is happening in the sample, i.e. on ψ, as is clear from their equations, e.g.

$$i\partial_t \Phi_L = H_L \Phi_L + \tau_L^+ \psi. \tag{4.4}$$

We would prefer to not have such an interdependency, we would like that the source depend only on the reservoirs plus leads. And this can be achieved. First, for simplicity, let us consider a situation around just one (say the left) lead, and divide the wavefunction in the lead as

$$\Phi_L = \phi_L + \chi_L, \tag{4.5}$$

with the ϕ_L part being defined to evolve according to H_L, i.e., as if the sample has not been connected,

$$i\partial_t \phi_L = H_L \phi_L. \tag{4.6}$$

Then χ_L can be considered as a 'correction for the presence of the connected H', or 'the part that rises because of the connected sample'. The equation (4.4) then means

$$(i\partial_t - H_L)\chi_L = \tau_L^+ \psi, \tag{4.7}$$

which can be solved for χ_L using the Green-function method. A general solution can be written as

$$\chi_L(t) = f_L(t) + \int \mathrm{d}t'\, G_L(t-t')\tau_L^+ \psi(t'), \tag{4.8}$$

where G_L is any (so-called Green) function satisfying

$$(i\partial_t - H_L) G_L(t) = \delta(t), \tag{4.9}$$

and $f(t)$ is a solution of the homogeneous equation, $(i\partial_t - H_L) f_L(t) = 0$.† Although any G_L solving (4.9) can be used, two particular Green functions have a privileged position: the retarded, G_L^r, and the advanced, G_L^a, Green function, defined by $G_L^r(t<0)=0$ and $G_L^a(t>0)=0$. Their advantage over the others is their simple use when boundary conditions are known at some time in the past or in the future. In our case we know (prepare) the situation in the past and are interested in the further development. This favours the use of the retarded form,

---

the system loses particles.

¶ Note that in the source there are terms containing τ_L and τ_R (and no $\tau_{L,R}^+$). Thus $\tau_{L,R}$ can be thought of as the agents supplying the wavefunction *into* the sample; similarly, $\tau_{L,R}^+$ will be seen to function in the opposite direction (see Fig. 4.3). Also note that for this to work we need to have τ_L below the diagonal and τ_R above the diagonal in (4.2).

† We may want to show that χ_L can really be written using any (matrix) Green function solving the equation (4.9). Let us thus choose any two such functions, G_1 and G_2, and assume $\chi_L(t) = f_L(t) + \int \mathrm{d}t' G_1(t-t')s(t')$, where $s(t') = \tau_L^+ \psi(t')$. Evidently, from (4.9), a difference of any two Green functions is a function satisfying the homogeneous equation; for $G_{1,2}$ let us denote this difference by g, $G_1 = G_2 + g$. Then

$$\chi_L(t) = f_L(t) + \int \mathrm{d}t'\, (G_2+g)(t-t')s(t') = f_L(t) + \int \mathrm{d}t'\, g(t-t')s(t') + \int \mathrm{d}t'\, G_2(t-t')s(t'),$$

and we only need to show that $h_L(t) \equiv \int \mathrm{d}t' g(t-t')s(t')$ is a solution of the homogeneous equation:

$$(i\partial_t - H_L)h_L(t) = \int \mathrm{d}t'\, (i\partial_t - H_L) g(t-t')s(t') = 0,$$

since g solves the homogeneous equation, $(i\partial_t - H_L)g(t-t') = 0$, for any t'.

$$\chi_L(t) \;=\; f_L(t) \;+\; \int_{-\infty}^{t} \mathrm{d}t'\; G_L^r(t-t')\,\tau_L^+\,\psi(t'), \tag{4.10}$$

where we have explicitly noted the limited upper bound of the integral originating from the retarded-function condition. In our scattering picture we know that before the time the electron reaches the sample, t_0, the wavefunction inside the sample is zero, $\psi(t<t_0)=0$. Thus the integral is for all $t<t_0$ zero.† Also, until t_0 the 'correction' χ_L of the wavefunction in the lead due to the presence of the sample is zero (the sample has not yet interacted with the electron), $\chi_L(t<t_0)=0$. This fixes $f_L(t)$ to be a zero function, since otherwise it could not be zero for $t<t_0$: it evolves according to H_L, i.e. unitarily, and preserves its norm. Hence

$$\chi_L(t) \;=\; \int \mathrm{d}t'\; G_L(t-t')\,\tau_L^+\,\psi(t'), \tag{4.11}$$

where we have also dropped the superscript *r* from the Green function; in the following, *all Green functions without an explicit 'r' or 'a' superscript are retarded.* Finally, Fourier-transforming the last result (assuming it can be done, see below) we get

$$\chi_L(\omega) \;=\; G_L(\omega)\,\tau_L^+\,\psi(\omega). \tag{4.12}$$

Equation (4.11, 4.12) expresses χ_L by means of the sample's part of the wavefunction.

Evaluation of G_L (and analogically of G_R) follows from (4.9). One way of thinking is: Anywhere but for $t=0$ the Green function solves the homogeneous equation. We know that the evolution operator $U_L(t)=e^{-iH_L t}$ also satisfies the homogeneous equation. Further $U_L(t)\to 1$ as $t\to 0$. Hence $G_L(t)\equiv G_L^r(t)=-iU_L(t)\,\theta(t)$. [Similarly, $G_L^a(t)=iU_L(t)\,\theta(-t)$.] This close relation of *G*s to *U* explains why Green functions are also called propagators. For the use in the derived equations we further need the Fourier transform of $G_{L,R}(t)$. Apparently, however, such a transform, e.g. $G_L(\omega)=-i\int_0^\infty \mathrm{d}t\; e^{-i(H_L-\omega)t}$, does not exist. By replacing ω with a complex variable *z* one sees the integral nonetheless exists for any *z* in the upper half-plane, i.e. for $\mathrm{Im}z>0$, with the result

$$G_L(z;\,\mathrm{Im}z>0) \;=\; (z-H_L)^{-1}. \tag{4.13}$$

This suggests (in a sense this is a definition)

$$G_{L,R}(\omega) \;=\; \lim_{\varepsilon\to 0^+} (\omega+i\varepsilon-H_{L,R})^{-1} \;\equiv\; (\omega+i0-H_{L,R})^{-1}, \tag{4.14}$$

which is to be understood in the distributive sense, i.e., $G_{L,R}(\omega)$, when actually used, is to be integrated over some interval of ω, generally together with some other (well-behaved) function of ω, finally taking the indicated limit.‡* [Similarly, $G_{L,R}^a(\omega)=(\omega-i0-H_{L,R})^{-1}=G_{L,R}^{r+}(\omega)=G_{L,R}^+(\omega)$.]

Using (4.11) and (4.5) in (4.3) we get (considering explicitly only the left lead)

$$(i\partial_t - H)\,\psi(t) \;=\; \tau_L\left[\phi_L(t) \;+\; \int G_L(t-t')\,\tau_L^+\,\psi(t')\right], \tag{4.15}$$

which, when Fourier-transformed, is

$$(\omega - H)\,\psi(\omega) \;=\; \tau_L\left[\phi_L(\omega) \;+\; G_L(\omega)\,\tau_L^+\,\psi(\omega)\right], \tag{4.16}$$

---

† If they were a perfectionist, they would say the argument is spoilt a bit by the fact that in a non-relativistic theory any perturbation has an immediate effect everywhere in space since there is no speed limit. However, this point really goes against the non-relativistic theory, not against the presented idea.

‡ Using the residue theorem it is straightforward to show that $G_{L,R}(t)\equiv -i\theta(t)\,e^{-iH_{L,R}t}=\lim_{\varepsilon\to 0^+}\int\frac{\mathrm{d}\omega}{2\pi}\;\frac{e^{-i\omega t}}{\omega+i\varepsilon-H_{L,R}}$.

* Perhaps more physically, we may think of the limiting procedure as of replacing the Green function in the time domain, $G_L(t)=-i\theta(t)\,e^{-iH_L t}$, with $G_L(t;\varepsilon)\equiv G_L(t)\,e^{-\varepsilon t}=-i\theta(t)\,e^{-i(H_L-i\varepsilon)t}$. Then $G_L(t;\varepsilon)$ has the transform for any positive ε, and for finite times $G_L(t)$ and $G_L(t;\varepsilon)$ can be made as close as one wishes by using a small enough ε. After all, any observation happens at a finite time. (Of course, $G_L(\omega;\varepsilon)$ differs for different εs, but the impact of this change within the finite time of observation can be made negligible.)

i.e.,

$$(\omega - H - \tau_L G_L \tau_L^+)\psi = \tau_L \phi_L, \tag{4.17}$$

or

$$(\omega - H - \Sigma_L)\psi = S_L, \tag{4.18}$$

where we have defined the left-lead (retarded; it depends on the retarded propagator G_L) selfenergy** Σ_L and the source S_L related to the left lead,

$$\Sigma_L \equiv \tau_L G_L \tau_L^+, \qquad S_L \equiv \tau_L \phi_L. \tag{4.19}$$

The nice thing about (4.17) is that the source on the RHS now only depends on the wavefunction ϕ_L, which is in no way modified by the presence of the sample.

When the right lead is considered along with the left one, the argumentation goes along exactly the same lines with the result

$$(\omega - H - \Sigma)\psi = S, \qquad \Sigma \equiv \Sigma_L + \Sigma_{R,} \quad S \equiv S_L + S_R, \tag{4.20}$$

where (c.f. (4.19))

$$\Sigma_R \equiv \tau_R G_R \tau_R^+, \qquad S_R \equiv \tau_R \phi_R. \tag{4.21}$$

All the objects in (4.20) are 'small' in that they are limited to the sample's volume. Although the selfenergies and sources depend, respectively, on the leads' Green functions and the leads' parts of the wavefunction, which are 'large' (large matrices and vectors if we use a discrete representation), $\tau^{(+)}$s extract only the needed information from them—only $[\tau_L]_{1,0}$, $[\tau_L^+]_{0,1}$, $[\tau_R]_{N,N+1}$, $[\tau_R^+]_{N+1,N}$ are non-zero in the principal-layer indices (by definition of principal layers). Hence only the following principal-layer indices of the selfenergies and sources are non-zero: $[\Sigma_L]_{1,1}$, $[S_L]_1$, $[\Sigma_R]_{N,N}$, $[S_R]_N$.

Finally, we really want to write (4.20) as an equation for ψ. If the inverse of $(\omega - H - \Sigma)$ exists, the expression is clear. Otherwise (and this possibly may happen) we need to add $i0$ to ω; this does not change the result when the plain inverse of $(\omega - H - \Sigma)$ exists, otherwise it i) guarantees the inverse existence and ii) ensures that thus obtained G is the retarded form of the propagator:¶§‖

** The name selfenergy is generally used when describing interaction of some particle with its surroundings. In the current context it is more of a misnomer. Nonetheless, since the equations occuring in both cases are basically the same, the name is used in the current context, too.

¶ We stress that G is the propagator for the central part (the sample) with the *leads connected*, i.e., not for an isolated sample. This might be slightly deceptive especially if one compares G to leads' $G_{L,R}$, (4.14), the latter being related to *isolated* leads. [Often, small (g) and capital (G) letters are used to explicitly distinguish 'connected' from 'isolated' cases, however, we will not do so here; we will only need a lead's Green function when the lead is isolated, and the sample's Green function when the leads are connected.]

§ The existence of the inverse as well as its retarded character can be inferred from another view on G. As G propagates the sample's part of the wavefunction, it must alternatively be obtainable also by restricting the 'whole-world' retarded propagator to the sample's space. I.e., $G = [G_w]_{\text{sample}}$, $G_w = (\omega + i0 - H_w)^{-1}$, where the last formula for G_w is justified by the same arguments as those that lead to (4.14), but now used for H_w describing the 'whole world', $H_w = H_L + \tau_L + \tau_L^+ + H + \tau_R + \tau_R^+ + H_R$, instead of just one lead. The 'whole-world' Hamiltonian H_w is hermitian, thus G_w exists (H_w can be diagonalized and adding $i0$ to the necessarily real numbers on the diagonal makes the matrix trivially invertible). We may use the general formula for calculating an 'interesting' sub-block $[M^{-1}]_{II}$ of the inverse of M

$$[M^{-1}]_{II} = (M_{II} - M_{IU}[M_{UU}]^{-1} M_{UI})^{-1}, \tag{4.22}$$

where the super-index I spans the matrix indices of the inverse that we are interested in, and U spans the uninteresting indices, and apply it to get the sample's part of G_w, I = sample, U = leads. The formula is only usable if $[M_{UU}]^{-1}$ exists. [E.g., although $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & a & 0 \\ 0 & a & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ has the inverse $\frac{1}{a}\begin{bmatrix} a & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & a \end{bmatrix}$ for any non-zero a, the block of the inverse with $I = \{1,2\}$ cannot be obtained from the formula because M_{UU}, $U = \{3,4\}$ does not have the inverse.] Fortunately, the inverse of the U-part being an inverse of a matrix with two independent blocks, $\omega + i0 - H_L - H_R$ exists by the same

$$\psi(\omega) = G(\omega)S(\omega), \qquad G(\omega) = (\omega + i0 - H - \Sigma)^{-1}. \tag{4.23}$$

The last equation, together with the expressions for S and Σ, (4.20, 4.19, 4.21), allows us to determine the sample's part of a one-electron wavefunction if we know the (by sample unperturbed) lead's part of the same electron. (Of course, an electron arrives just from one lead, never coherently from both; for an electron coming from the left, ϕ_R is identically zero, and vice versa.)

Having equations to describe propagation of one electron, we now want to take into consideration all the electrons. To this end we employ the 1-particle density matrices. First, for evaluation of 1-particle quantities such a matrix is sufficient. Second, for non-interacting electrons in equilibrium we know the matrix (in the basis of 1-particle energy eigenvectors it is a diagonal matrix of the related Fermi factors). Third, and importantly, in the density matrix formalism if two non-coherent subsystems are to be described at once, the two matrices simply add. This last aspect demonstrates itself already in the case of just one lead, where the electrons at different levels are to be taken incoherent, and will also be used below to account for electrons coming to the sample from different leads.

The task is to evaluate the 1-particle density matrix ρ pertaining to the sample. We consider each electron energy (ω) level of each lead separately, calculate the sample's part ψ arising from filling of the lead's level, create from it a contribution to the sample's $\rho(t)$, i.e. $\psi(t)\psi^+(t)$, and add all such contributions, weighted with probability of each ω-level filling, together.†‡ Since we are considering leads' energy eigenstates, the contribution to $\rho(t)$ from each such eigenstate will be independent of time, and hence ρ itself will be constant in time (we are describing a steady state).¶ For an electron level i with energy ω_L^i in the left lead the contribution is (apart from the filling factor)

$$\psi^i \psi^{i+} = G(\omega_L^i)\tau_L \phi_L^i \phi_L^{i+} \tau_L^+ G(\omega_L^i)^+. \tag{4.24}$$

(Consult footnote ¶ if unsure about the notation and/or meaning.) Considering all the left-lead levels and including the filling factors (note that ρ_l below is a contribution to the *central*-part density matrix, and is not the left-lead density matrix)

$$\rho_l = \sum_i f_L(\omega_L^i) G(\omega_L^i)\tau_L \phi_L^i \phi_L^{i+} \tau_L^+ G(\omega_L^i)^+ =$$

---

argument as was given for the existence of G_w, and represents the retarded propagators $G_{L,R}$. We thus arrive at the formula $G=(\omega+i0-H-\tau_L G_L \tau_L^+ - \tau_R G_R \tau_R^+)^{-1}$, which is exactly the G in (4.23), but now with the $+i0$ justified and proven existence. [One may wonder why we did not use this derivation of G straight away. The reason was to not only obtain the propagator, but also show the ontology of the sources $S_{L,R}$.]

‖ Despite the same look of the $+i0$ terms in the formulæ for $G_{L,R}$, (4.14), and G, (4.23), one should assess their different importance. In the case of the leads, the $i0$ is necessary for any ω from the spectrum of $H_{L,R}$. For the sample, on the other hand, once we have added the selfenergies to account for the leads, it may well happen that for all ω the inverse $(\omega - H - \Sigma)^{-1}$ exists without any added $i0$. This is because the leads' Σs already contain some finite imaginary contributions, effectively causing that $H+\Sigma$ has fewer (in the extreme case none) real eigenvalues. Physically it means that majority of the states originally necessarily localized in the sample when it was isolated are, after the leads have been connected, able to escape from the sample. Perhaps an illustrative example is an infinite 1D tight-binding chain with one site, representing the sample, having a set of different energy levels. When calculating G for such a sample, we do not have to add $i0$ as long as ω lies within the energy band of the leads. Outside the band the $i0$ is necessary, the sample there does not effectively couple to the leads and behaves much like isolated.

† We note that compared to the original presented picture where we discussed a localized electron wavefunction impinging on the sample (since this way of thinking feels natural), we now separately consider each (lead's) energy eigenstate, i.e., potentially quite a delocalized state. The reason for doing so is that statistical thermodynamics enforces effective decoherence just between these energy eigenstates. (See section 2.2.1 of (Vent08).)

‡ A subtlety is to realize that we do not have to be anxious about exclusion-principle blocking when adding all the density contributions together as long as the transport can be considered coherent. See, e.g., the related note in section 9.2 of (Datt07).

¶ A lead's energy eigenstate with energy equal to ω^i has necessarily the form $\phi_{\omega^i}(\boldsymbol{r},t)=\phi_{\omega^i}(\boldsymbol{r})e^{-i\omega^i t}$. Hence its ω-representation, as used for example in (4.23), is $\phi_{\omega^i}(\boldsymbol{r},\omega)=\phi_{\omega^i}(\boldsymbol{r})2\pi\delta(\omega-\omega^i)$. Using (4.23) we get $\psi_{\omega^i}(\boldsymbol{r},\omega)=G(\omega)\tau\phi_{\omega^i}(\boldsymbol{r})2\pi\delta(\omega-\omega^i)$, which back in the time representation is $\psi_{\omega^i}(\boldsymbol{r},t)=G(\omega^i)\tau\phi_{\omega^i}(\boldsymbol{r},t)$. Then the contribution to $\rho(t)$, if the level is filled, is $\psi_{\omega^i}(\boldsymbol{r},t)\psi_{\omega^i}(\boldsymbol{r},t)^+=G(\omega^i)\tau\phi_{\omega^i}(\boldsymbol{r},t)\phi_{\omega^i}(\boldsymbol{r},t)^+\tau^+G(\omega^i)^+$, which is clearly a time independent expression, i.e., $\rho_{\omega^i}=G(\omega^i)\tau\phi_{\omega^i}(\boldsymbol{r})\phi_{\omega^i}(\boldsymbol{r})^+\tau^+G(\omega^i)^+$; in short, $\rho_{\omega^i}=\psi_{\omega^i}\psi_{\omega^i}^+ = =G(\omega^i)\tau\phi_{\omega^i}\phi_{\omega^i}^+\tau^+G(\omega^i)^+$.

$$= \int \frac{d\omega}{2\pi}\, f_L(\omega) G(\omega) \tau_L \left[2\pi \sum_i \delta(\omega-\omega_L^i)\phi_L^i \phi_L^{i+} \right] \tau_L^+ G(\omega)^+ =$$

$$= \int \frac{d\omega}{2\pi}\, f_L(\omega) G(\omega) \tau_L A_L(\omega) \tau_L^+ G(\omega)^+ . \tag{4.25}$$

Here $f_L(\omega) \equiv f_0(\omega-\mu_L)$, μ_L is the chemical potential of the left lead, f_0 is the Fermi–Dirac distribution function, $f_0(x)=(1+e^{\beta x})^{-1}$, $\beta = 1/kT$, and we have defined the so-called spectral function of the left lead A_L,*

$$A_L(\omega) \equiv 2\pi \sum_i \delta(\omega-\omega_L^i)\phi_L^i \phi_L^{i+} = 2\pi\delta(\omega-H_L) = i(G_L - G_L^+), \tag{4.27}$$

where the last equality follows from the (distribution-like) identity

$$\frac{1}{x \pm i0} = \mathrm{P}\frac{1}{x} \mp i\pi\delta(x) \tag{4.28}$$

and (4.14). The term $\tau_L A_L \tau_L^+$ appearing in (4.25) can be rewritten as follows

$$\tau_L A_L \tau_L^+ = i\tau_L (G_L - G_L^+)\tau_L^+ = i(\tau_L G_L \tau_L^+ - \tau_L G_L^+ \tau_L^+)$$
$$= i(\Sigma_L - \Sigma_L^+) \equiv \Gamma_L , \tag{4.29}$$

using (4.27), (4.19), and defining the so-called broadening matrix Γ_L. Then (4.25) reads

$$\rho_l = \int \frac{d\omega}{2\pi}\, G(\omega) f_L(\omega) \Gamma_L(\omega) G(\omega)^+ . \tag{4.30}$$

Analogically the contribution to the sample's density matrix from the right lead is

$$\rho_r = \int \frac{d\omega}{2\pi}\, G(\omega) f_R(\omega) \Gamma_R(\omega) G(\omega)^+ , \tag{4.31}$$

having defined f_R, A_R, Γ_R in analogy with the left definitions. Hence

$$\rho = \rho_l + \rho_r = \int \frac{d\omega}{2\pi}\, G[\Gamma_L f_L + \Gamma_R f_R] G^+ = \int \frac{d\omega}{2\pi}\, G^{\mathrm{n}}, \tag{4.32}$$

where§

$$G^{\mathrm{n}} \equiv G[\Gamma_L f_L + \Gamma_R f_R] G^+ , \tag{4.33}$$

or

$$G^{\mathrm{n}} = G\Sigma^{\mathrm{in}} G^+ , \qquad \Sigma^{\mathrm{in}} \equiv (\Sigma_L^{\mathrm{in}}) + [\Sigma_R^{\mathrm{in}}] \equiv (\Gamma_L f_L) + [\Gamma_R f_R]. \tag{4.34}$$

Finally, we rewrite

$$G^{\mathrm{n}} = A_l f_L + A_r f_R , \qquad A_l \equiv G\Gamma_L G^+ , \quad A_r \equiv G\Gamma_R G^+ , \tag{4.35}$$

and note that

---

* It is instructive to realize that using thus defined spectral function one may express the lead's (equilibrium) density matrix

$$\rho_L = \sum_i f_L(\omega_L^i)\phi_L^i \phi_L^{i+} = \int d\omega\, f_L(\omega) \sum_i \delta(\omega-\omega_L^i)\phi_L^i \phi_L^{i+} = \int \frac{d\omega}{2\pi}\, f_L(\omega) A_L(\omega). \tag{4.26}$$

From this we may interpret the spectral function as an energy-resolved object describing the energy levels, in the sense of potential contributions to the 1-particle density matrix. These levels can then be used (filled), and in equilibrium this filling is according the Fermi–Dirac distribution.

§ The used notation follows (Datt07, Datt00). G^{n} stands for the electron correlation function, as opposed to G^{p} (not used here) which is the hole correlation function. These functions are directly related to more frequently used lesser and greater correlation functions of non-equilibrium Green-function formalism (Kada62): $G^{\mathrm{n}} = -iG^<$, $G^{\mathrm{p}} = iG^>$. Similarly, the so-called in-scattering selfenergy Σ^{in} and out-scattering selfenergy Σ^{out} (not used here) relate to more common lesser and greater selfenergies, $\Sigma^{\mathrm{in}} = -i\Sigma^<$ and $\Sigma^{\mathrm{out}} = i\Sigma^>$. Hence (4.34) can be also written as $G^< = G^r \Sigma^< G^a$ (and is closely related to the so-called Generalized Kadanoff–Baym (GKB) equation).

$$A_l + A_r = G(\Gamma_L + \Gamma_R)G^+ = G\Gamma G^+ = G^+\Gamma G = A, \tag{4.36}$$

where

$$A \equiv i(G - G^+), \tag{4.37}$$

$$\Gamma \equiv i(\Sigma - \Sigma^+) = i(\Sigma_L + \Sigma_R - \Sigma_L^+ - \Sigma_R^+) = \Gamma_L + \Gamma_R. \tag{4.38}$$

Regarding the last two equalities in (4.36): From $G=(\omega - H - \Sigma)^{-1}$ one sees $G^{-1} - (G^+)^{-1} = -\Sigma + \Sigma^+$, which equals $i\Gamma$. Multiplying by G from the left and by G^+ from the right, and the other way round, yields the presented.

Equations (4.35, 4.36) show that G^{n} can be thought of as arising from filling of two spectral functions $A_{l,r}$ (according to f_L and f_R, respectively), which moreover add up to the sample's total spectral function A.¶ [In equilibrium, $f_L = f_R \equiv f$, and (4.35) expectedly reduces to $G^{\mathrm{n}} = Af$. One should compare (4.32) to (4.26) to see that in the former G^{n} stands as if at the place of $A_L f_L$ of the latter. When calculating ρ_L of an *isolated* lead, (4.26), the lead is assumed to be in equilibrium and hence the equilibrium-like filling of A_L. On the contrary, the sample with the leads *connected* is potentially (for $\mu_L \neq \mu_R$) out of equilibrium and the situation is described by the more complicated G^{n}.]

Having derived the expression for the sample's (generally non-equilibrium) density matrix, (4.32, 4.35), we will now proceed to calculate the conductance of the sample.

With knowledge of a 1-particle density matrix we may calculate the expectation value of any one-body operator O as

$$\begin{aligned} \overline{O} &= \mathrm{Tr}\, O\rho = \int \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{Tr}\, OG^{\mathrm{n}} = \int \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{Tr}\, OG(\Gamma_L f_L + \Gamma_R f_R)G^+ = \\ &= \int \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{Tr}\, \{(\Gamma_L f_L + \Gamma_R f_R)G^+ OG\}. \end{aligned} \tag{4.39}$$

We now introduce an operator X (generalized coordinate) defined to be zero for a few first principal layers of the sample, and defined to be one for a few last principal layers of the sample, with some not really important profile around the sample's centre, where the value changes from zero to one. The current operator can then be expressed as a time derivative of this operator,

$$I = -i[X,H]. \tag{4.40}$$

Using I as O in (4.39) we come to a term like

$$G^+ IG = -iG^+(XH - HX)G. \tag{4.41}$$

From $G=(\omega - H - \Sigma)^{-1}$ we see that $H = \omega - \Sigma - G^{-1}$, hence $HG = (\omega - \Sigma)G - 1$ and $G^+ H = G^+(\omega - \Sigma^+) - 1$. Then

$$G^+ IG = i(G^+ X - XG) + iG^+(X\Sigma - \Sigma^+ X)G. \tag{4.42}$$

The operator X has been defined so that wherever Σ is non-zero, which is only in the first and last principal layer, the operator is constant. This means that $[X,\Sigma] = [X,\Sigma^+] = 0$. Furthemore, X has zero (equal-to-one) elements in the first (last) principal layer, to which all non-zero elements of Σ_L (Σ_R) are confined. Hence $X\Sigma_L^{(+)} = \Sigma_L^{(+)} X = 0$, and $X\Sigma_R = \Sigma_R X = \Sigma_R$, $X\Sigma_R^+ = \Sigma_R^+ X = \Sigma_R^+$; from these we immediately derive $\Gamma_L X = X\Gamma_L = 0$ and $\Gamma_R X = X\Gamma_R = \Gamma_R$. The second term on the RHS of the last equation is then

$$iG^+(X\Sigma - \Sigma^+ X)G = iG^+(\Sigma - \Sigma^+)XG = iG^+(\Sigma_R - \Sigma_R^+)XG = G^+\Gamma_R G. \tag{4.43}$$

¶ One should be careful to distinguish $A_{L,R}$, which are the spectral functions of the *leads*, and $A_{l,r}$, which are two parts of the *sample's* spectral function. The indices l,r are in the latter case used to indicate according to what Fermi–Dirac distribution the two parts of A are to be filled.

Equation (4.39) for $\bar{I}$ is then

$$\bar{I} = \int \frac{d\omega}{2\pi} \underset{A}{\mathrm{Tr}\,(\Gamma_L f_L + \Gamma_R f_R)} \underset{B}{\Big[} \underset{C}{i(G^+ X - XG)} + \underset{D}{G^+ \Gamma_R G} \Big]. \tag{4.44}$$

From the above-mentioned then

$$\mathrm{Tr\ AC} = 0, \tag{4.45}$$

$$\mathrm{Tr\ AD} = f_L \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G, \tag{4.46}$$

$$\mathrm{Tr\ BC} = f_R \mathrm{Tr}\, i(G^+ \Gamma_R - \Gamma_R G) = f_R \mathrm{Tr}\, i\Gamma_R (G^+ - G) = -f_R \mathrm{Tr}\, \Gamma_R A, \tag{4.47}$$

$$\begin{aligned} \mathrm{Tr\ BD} &= f_R \mathrm{Tr}\, \Gamma_R G^+ \Gamma_R G = f_R \mathrm{Tr}\, (\Gamma - \Gamma_L) G^+ \Gamma_R G = \\ &= -f_R \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G + f_R \mathrm{Tr}\, \Gamma_R G \Gamma G^+ = \\ &= -f_R \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G + f_R \mathrm{Tr}\, \Gamma_R A, \end{aligned} \tag{4.48}$$

where on the last line we used (4.36). Summing all together,

$$\bar{I} = \int \frac{d\omega}{2\pi} (f_L - f_R) \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G. \tag{4.49}$$

Finally, from (4.49) we may easily derive an expression for the linear-response conductance, i.e., the conductance of the system when it is infinitesimally close to equilibrium. The conductance, C, is a derivative of the (charge) current with respect to the bias voltage $V = (\mu_L - \mu_R)/q$, q is the charge of an electron,

$$C = q \frac{d\bar{I}}{dV}\bigg|_{f_L = f_R} = q \int \frac{d\omega}{2\pi} \frac{d(f_L - f_R)}{dV} \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G. \tag{4.50}$$

Recalling that $f_{L,R}(\omega) = f_0(\omega - \mu_{L,R})$ and denoting the equilibrium chemical potential by μ,

$$\begin{aligned} \delta(f_L - f_R) &= f_0'(\omega - \mu)(-\delta\mu_L + \delta\mu_R) = -f_0'(\omega - \mu)\delta(\mu_L - \mu_R) \\ &= -f_0'(\omega - \mu) q \delta V. \end{aligned} \tag{4.51}$$

Thus

$$C = q^2 \int \frac{d\omega}{2\pi} [-f_0'(\omega - \mu)]\ \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G. \tag{4.52}$$

At zero temperature, to which our further calculations are confined, the derivative of the Fermi–Dirac distribution is a (minus) delta function, $-f_0'(\omega - \mu) = \delta(\omega - \mu)$, and the conductance is a 'Fermi-surface property' (all quantities should be evaluated at $\omega = \mu$) with the value

$$C_{T=0} = q^2\ \mathrm{Tr}\, \Gamma_L G^+ \Gamma_R G. \tag{4.53}$$

Thanks to 'localization' of Γs, using principal-layer indices the last expression can be rewritten as

$$C_{T=0} = q^2\ \mathrm{tr}\, [\Gamma_L]_1 [G^+]_{1N} [\Gamma_R]_N [G]_{N1}. \tag{4.54}$$

This result is, in the essence, used by our codes to calculate systems' conductancies. In reality, we use a related formula which instead of the (so-called physical) Green functions Gs uses the auxiliary Green functions gs, (3.37). We will not provide more details here and refer the reader to (Kudr00, Carv06).

PART II — Illustrative Systems

In this second part of the thesis we present several examples of what impact may the spin-orbit (SO) interaction have on conductance of layered systems, with current flowing perpendicular to the planes (the CPP geometry). We investigate, on the one hand, (anisotropic) magnetoresistance ((A)MR) of metallic systems of fairly light elements, Co, Cu, Ni, and, on the other hand, systems with a tunnelling barrier, made of MgO or GaAs, where the tunnelling anisotropic magnetoresistance (TAMR) is of our interest. In both these classes, purely metallic or with a barrier, we display examples in which the SO interaction has no marked influence, and other in which a neglect of it would lead to quite erroneous conclusions.

In all our calculations we assume *purely coherent transport* of electrons throughout the system.†

† This actually does not quite preclude a diffusive-like transport. For a complicated (pseudo-) random structure the in essence coherent transport may look like diffusive on a global scale.

5

Metallic Systems

It was shown on the *ab-initio* calculation level by Banhart et al. (Banh96, Banh97) that in ferromagnetic Fe-Ni and Co-Ni *bulk* alloys the spin-orbit (SO) interaction plays a crucial role in the alloys' electronic-transport properties. The authors lucidly showed that allowing for the SO interaction leads, e.g., to an increase in isotropic residual resistivity by a factor of 50 in the case of the $Co_{0.3}Ni_{0.7}$ alloy as compared to a two-current model calculation, which, although incorporating the Darwin and the mass-velocity enhancement relativistic effects, ignored the SO interaction altogether (i.e., as if the scalar-relativistic approximation (SRA) were adopted).

Inspired by the bulk results we decided to test the effect of the SO interaction on conducting properties of *layered* systems. We investigate fcc systems composed of Co, Ni and Cu layers, with current flowing perpendicular to the planes (the CPP geometry). We start with pure-material systems, then look at simple interfaces and proceed to more complicated thicker multi-interface structures, moving from perfectly periodic ones to structures with randomly fluctuating widths of individual layers. The primary objective of doing so is to mimic the conditions under which the character of transport could change from ballistic on one side to diffusive on the other. A highly disordered layered structure might share some properties with random bulk alloys, at least in the direction across the layers, and one can anticipate that the pronounced SO-interaction effects seen in bulk systems could have some parallel in layered structures too. This part of our work was also inspired by that of J. Mathon, (Math96, Math97).

All the calculations presented in this chapter use 2024 $\boldsymbol{k}_{\parallel}$ vectors per surface (2D) Brillouin zone (BZ) for the electronic-structure calculation, and 10^4 $\boldsymbol{k}_{\parallel}$ vectors for the calculation of conductances. The small (ideally infinitesimal) imaginary quantities used in calculation of leads' retarded selfenergies was chosen to be 10^{-6} Ry.

5.1. Simple (Co, Cu, Ni)-based Magnetic Multilayers

Pure-material (Sharvin) Conductances

Conductances per a surface atom of pure Co, Ni, Cu fcc layered systems with current flowing along the *z*-axis oriented either along the 001 or the 111 crystallographic direction (i.e., by individual atomic layers we mean layers with the Miller indices 001 and 111, respectively)* and with magnetization $\boldsymbol{M} \parallel z$ are summarized in Tab. 5.1. We observe that conductance in the 001 direction is generally higher than in the 111 direction‡ and that switching-on of the SO interaction diminishes the conductance a little for magnetic systems (maximum is 3% for 001 Ni), while, quite

* Hereafter we denote such layered systems simply by fcc 001 and fcc 111.

‡ One should, however, be careful. The presented results are given per a surface atom. Thus geometry also plays a role when comparison in units using some fixed area, e.g. $\Omega^{-1}\mathrm{m}^{-2}$, is to be done (namely, in the case of 001 and 111 interfaces, if the conductances per a (same) unit area were the same, the numerical values per a surface atom would be in a ratio $C_{001}/C_{111} = 2/\sqrt{3} \approx 1.15$). The claim given in the main text is, nonetheless, true, even when such a geometrical factor is included.

expectedly, has no effect in the case of Cu.

$\boldsymbol{M} \parallel z$ C	fcc 001		fcc 111	
	SR	SO	SR	SO
Co (2.62)	2.67	2.64	2.17	2.16
Ni (2.60)	2.98	2.89	2.41	2.36
Cu (2.67)	1.84	1.84	1.59	1.59

Table 5.1: Conductances (per a surface atom) for pure Co, Ni, Cu fcc systems in units of (e^2/h) when the SO interaction is not (SR) or is (SO) included, current is flowing along the z-direction set to the 001 or 111 crystallographic direction, and Co and Ni systems' magnetization is along the z-axis. Numbers in the brackets are the used Wigner-Seitz radii (in the Rydberg atomic units).

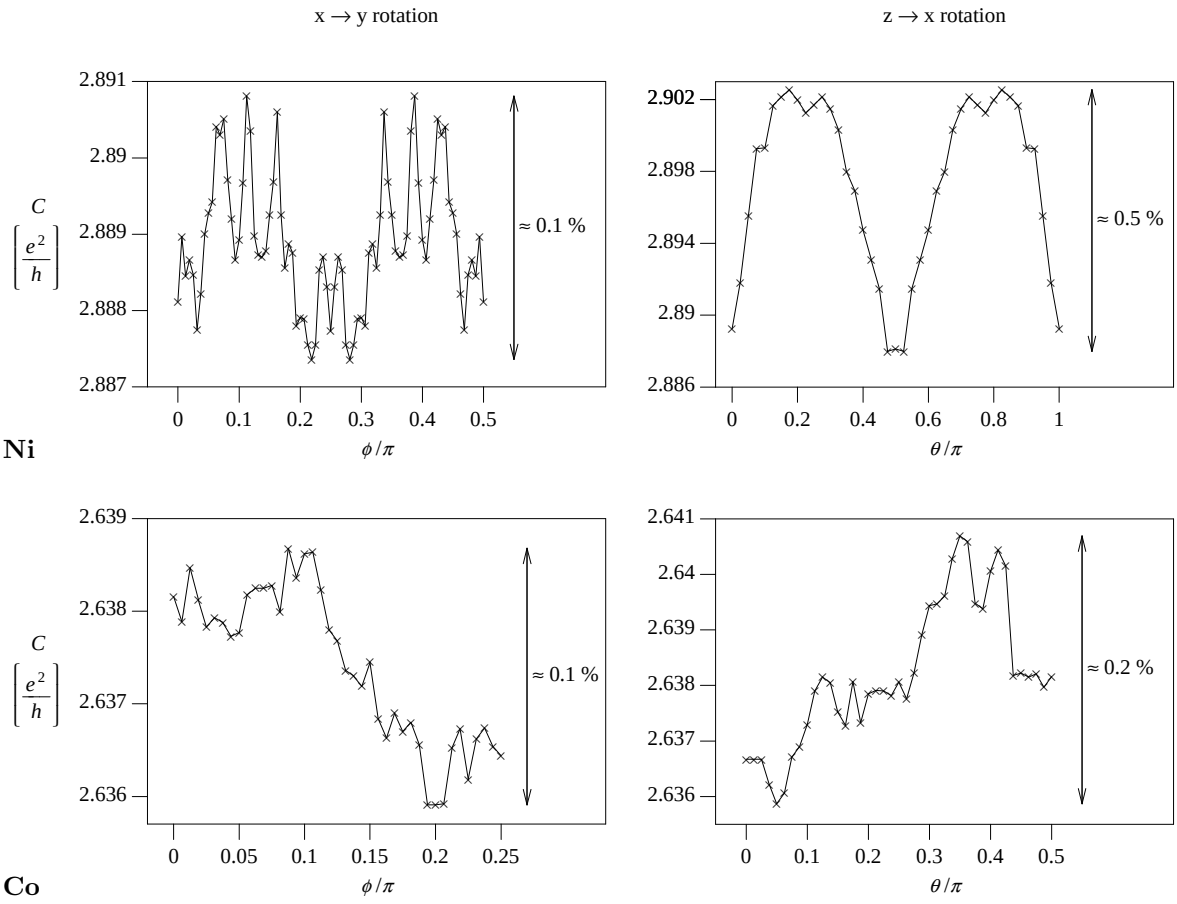


Figure 5.1: Dependence of the conductance (per a surface atom) of Ni and Co fcc 001 layered systems on magnetization direction, the AMR effect; on the left side $\theta = \pi/2$, on the right side $\phi = 0$. Used Wigner-Seitz radii are those used in Tab. 5.1.

Figure 5.1 depicts the dependence of fcc 001 Ni and Co Sharvin conductances on the magnetization direction, with $x \to y$ and $z \to x$ rotations considered; the direction is given by spherical angles θ, ϕ. Of course, the SO interaction must be on, otherwise no dependence would exist. For in-plane magnetization ($\theta = \pi/2$), the expected C_{4v} symmetry is (partially) demonstrated for the case of Ni and fully established if one realizes that changing a sign of $\boldsymbol{M}$ (or equivalently the effective magnetic field $\boldsymbol{B}_{\text{eff}}$) does not change the resulting total conductance.† We see that the

† Magnetic field enters our current theory via the term $\propto \boldsymbol{B}_{\text{eff}} \cdot \boldsymbol{S}$, hence a sign change in $\boldsymbol{B}_{\text{eff}}$ accompanied with a swap of

anisotropic magnetoresistance (AMR) for pure Ni and Co materials is rather small; the strongest dependence makes just a 0.5% change.

Single-interface systems

We now move on to systems with one interface, i.e., Co/Ni, Co/Cu, Ni/Cu, in both 001 and 111 geometries, and with magnetization kept along the *z*-axis (perpendicular to the planes). Such systems are modelled as composed of the left lead, several left-lead-material layers, several right-lead-material layers, and the right lead. In the present situation 4 atomic layers of each material are used in the central region. The individual materials have slightly different natural lattice parameters, however, we need have, in our model, the same 2D periodicity for all the layers. Thus we choose here the average Wigner-Seitz (WS) radius of the two materials.

Similarly to the case with pure materials above we calculate conductance of each system and compare the values with the SO interaction off and on. This is summarized in Table 5.2. Beside the total conductance of a system, e.g. $C_{\mathrm{Co/Ni}}$, presented are also i) Sharvin conductances of the pure materials with the relevant (changed) WS radii, ii) total resistance of the system, iii) resistance of the system's *interface*, defined thus: $R^{A/B}_{\mathrm{interface}} \equiv 1/C_{A/B} - \frac{1}{2}(1/C_A + 1/C_B)$. The latter to a certain degree describes the impact of the interface existence on the total resistivity. For the systems under study we may read off that an interface contributes about the same resistance as is the contact resistance (fundamentally caused by the connection of a system to reservoirs; or, from a slightly different perspective, by a limited number of transverse modes, see (Datt97)).

Co↑/Ni↑ (2.61)								
		C_{Co}	C_{Ni}	$C_{\mathrm{Co/Ni}}$	$R_{\mathrm{Co/Ni}}$	$\delta C[\%]$	$R^{\mathrm{Co/Ni}}_{\mathrm{interface}}$	$\delta R_{\mathrm{interface}}[\%]$
001	SR	2.653	2.964	1.701	0.588	−1.7	0.231	+1.7
	SO	2.626	2.899	1.672	0.598		0.235	
111	SR	2.163	2.405	1.552	0.644	−1.4	0.205	+2.0
	SO	2.153	2.360	1.531	0.653		0.209	
Co/Cu (2.64)								
		C_{Co}	C_{Cu}	$C_{\mathrm{Co/Cu}}$	$R_{\mathrm{Co/Cu}}$	$\delta C[\%]$	$R^{\mathrm{Co/Cu}}_{\mathrm{interface}}$	$\delta R_{\mathrm{interface}}[\%]$
001	SR	2.668	1.842	1.255	0.797	−1.4	0.338	+3.0
	SO	2.638	1.842	1.237	0.809		0.348	
111	SR	2.173	1.591	1.103	0.906	−0.4	0.362	+0.6
	SO	2.159	1.590	1.099	0.910		0.364	
Ni/Cu (2.64)								
		C_{Ni}	C_{Cu}	$C_{\mathrm{Ni/Cu}}$	$R_{\mathrm{Ni/Cu}}$	$\delta C[\%]$	$R^{\mathrm{Ni/Cu}}_{\mathrm{interface}}$	$\delta R_{\mathrm{interface}}[\%]$
001	SR	2.978	1.842	1.468	0.681	−1.1	0.242	+0.8
	SO	2.891	1.842	1.452	0.689		0.244	
111	SR	2.414	1.591	1.216	0.823	−1.2	0.302	+1.7
	SO	2.363	1.590	1.201	0.832		0.307	

Table 5.2: Single-interface systems' conductances. But for an added interface the notation in the table is the same as in Tab. 5.1.

In any case, the impact of the SO interaction in these simple systems stays small, on a one-percent level.

up and down spin channels leads to the same result.

Two-interface systems

We quickly go through systems with two internal interfaces. We first opt for the central part to be from a nonmagnetic Cu and the leads from magnetic Ni and Co. In such systems the Giant Magnetoresistance (GMR) effect can be studied, for which the leads' magnetizations are either oriented parallel (P) or anti-parallel (AP). The results, with the notation established above, are given in Tab. 5.3. We use the giant magnetoresistance ratio in its 'optimistic' form,

$$GMR \equiv 100\,\frac{C_{\rm P} - C_{\rm AP}}{C_{\rm AP}} \;=\; 100\,(\frac{C_{\rm P}}{C_{\rm AP}} \;-\; 1) \;=\; 100\,(\frac{R_{\rm AP}}{R_{\rm P}} \;-\; 1)\,. \qquad (5.1)$$

For all these systems we use 4 atomic planes of Cu in the central region, surrounded from both sides by 4 buffer layers of the lead material.

(2.64, $\boldsymbol{M} \parallel z$) C and GMR	fcc 001				fcc 111			
4 intermediate Cu layers	SR		SO		SR		SO	
Co↑ / Cu / Co↑	1.20	46.3	1.18	45.7	1.01	29.5	1.01	31.2
Co↑ / Cu / Co↓	0.82		0.80		0.78		0.77	
Ni↑ / Cu / Ni↑	1.42	9.2	1.40	11.1	1.15	12.7	1.13	14.1
Ni↑ / Cu / Ni↓	1.30		1.26		1.02		0.99	
Co↑ / Cu / Ni↑	1.18	11.3	1.15	10.6	1.02	13.3	1.00	12.4
Co↑ / Cu / Ni↓	1.06		1.04		0.90		0.89	

Table 5.3: Two-interface systems with Cu in the central region used for studying the influence of the SO interaction on the GMR effect. Shown are the systems' conductances and the respective *GMR* ratios.

Apparently, the SO interaction modifies the *GMR* ratios by about 1% and this may be both a decrease or an increase. (Conductances generally decrease when SO is switched on; the *GMR* ratio is, however, a more complicated object whose value may increase.)

To gain some feeling about behaviour of anisotropic magnetoresistance when a system gets more complicated (we already showed results for pure materials in Fig. 5.1), we present in Fig. 5.2 some illustrative results for two-interface Co/4Ni/Co, Ni/4Cu/Ni and Co/4Cu/Co 001 systems. WS radii for all the systems were chosen equal to 2.64, and, as before, we use 4 atomic lead-material buffer layers on either side of the central material. Comparing the results with the results for AMR of pure materials, we see the angular dependence is now smoother. This can perhaps be understood in the sense that when combining different materials, their complicated Fermi-surface structures mutually effectively average out in the resultant conductance. The AMR is now a bit stronger than in the pure-material study, nonetheless still stays rather weak.

5.2. Thicker (Co, Cu, Ni) Magnetic Multilayers with Growing Structural Randomness

So far we have only found weak effects of the SO interaction. To move in the direction where our systems would more resemble bulk alloys (in which we know the SO interaction can manifest itself visibly) we have tried and created more complex layered structures.

J. Mathon (Math96, Math97) investigated CPP-geometry GMR of Co/Cu 001 and Fe/Cr 001 superlattices with fluctuating layer thicknesses embedded between Cu leads. He reported that, depending on the extent of fluctuation, the transport character changes from ballistic (when no thickness fluctuations occur, i.e., the superlattice is perfectly periodic) through ohmic (when moderate fluctuations are introduced) to an Anderson-localization regime (for large enough fluctuations).

Following Mathon we too have had a look at Co/Cu 001 superlattice systems with an increasing amount of disorder. For each system we calculate both the system's resistance and the GMR ratio, comparing the situation with all Co layers magnetized in the same (parallel) sense (P) and the situation when any next Co section is magnetized opposite to the previous one (AP order).

Later, we investigate Co/Ni 001 superlattices, which have turned out to be more interesting as

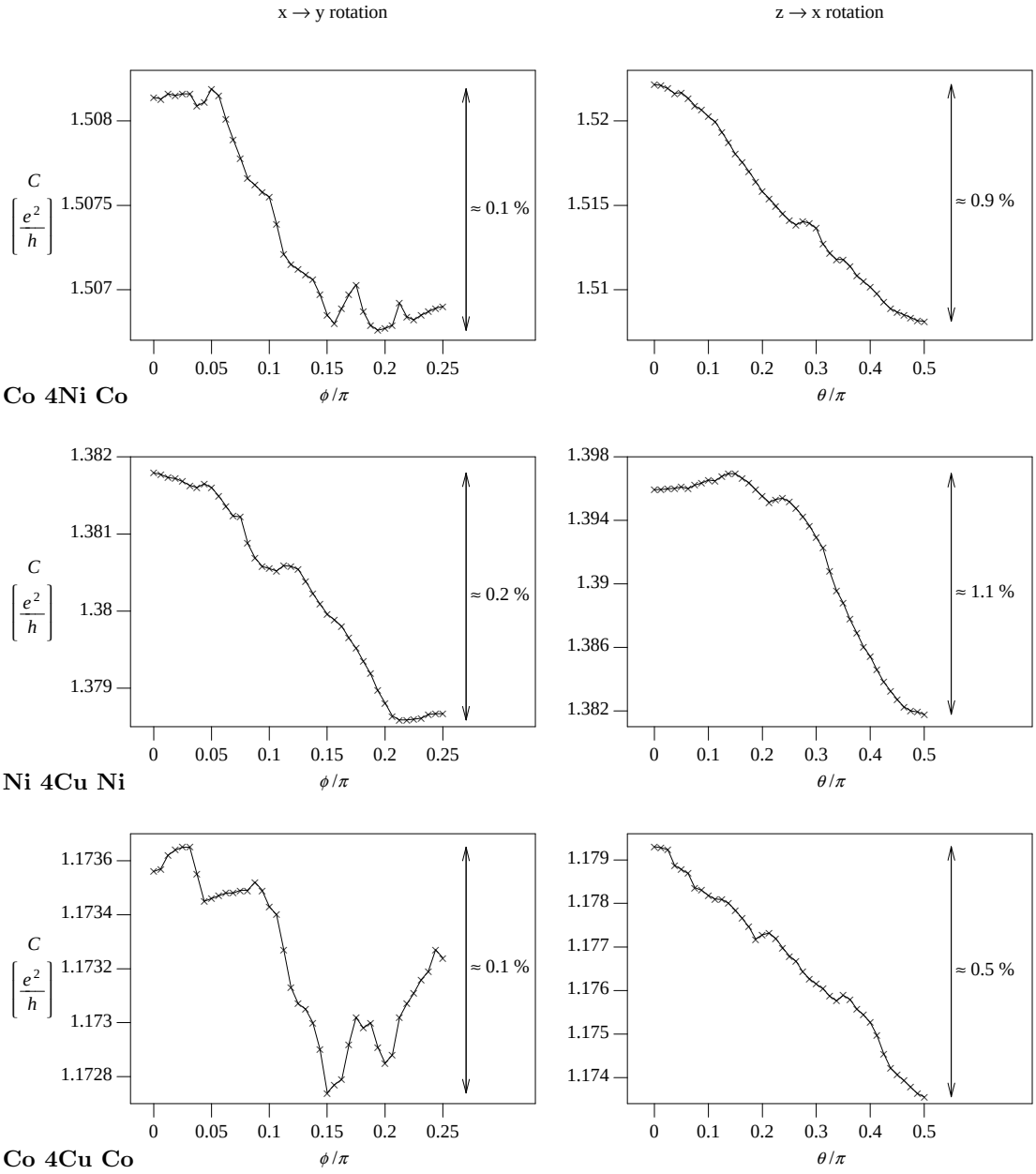


Figure 5.2: Anisotropic magnetoresistance, shown as a change in conductance (per a surface atom), of three chosen systems with two interfaces. On the left, the $x \to y$ rotation ($\theta = \pi/2$) of magnetization is depicted, on the right, there is the $z \to x$ rotation ($\phi = 0$). The corresponding SR (constant) results are $C^{\mathrm{SR}}_{\mathrm{Co/4Ni/Co}} = 1.56$, $C^{\mathrm{SR}}_{\mathrm{Ni/4Cu/Ni}} = 1.42$, and $C^{\mathrm{SR}}_{\mathrm{Co/4Cu/Co}} = 1.20$. WS radii of all atoms were chosen to be equal to 2.64.

far as the impact of the SO interaction is concerned. In these systems, due to the lack of a nonmagnetic spacer, we cannot straightforwardly study the GMR effect; instead, we first just look at systems' total resistivities, then study the AMR effect, and as a final step we create yet a bigger system with a nonmagnetic spacer in its middle and demonstrate a substantial influence of the SO interaction on the GMR ratio.

To be ever able to consider extended supperlattices with even hundreds of atomic layers and possibly fluctuating thicknesses of one-material sections, a few programs have been written which create necessary input files for our main codes. Basically, with a selfconsistent electronic-structure program, we precalculate electronic structures for all interfaces that may appear in the superlattice, and from these we build up the (not fully but close to selfconsistent) electronic structure for the whole superlattice. The input specifying a structure may, e.g., look like '2Cu 1-2[Ni 3-5Ag 2[Cr

2Au]] Mg', though for the systems considered in this thesis it is always a much simpler prescription. (Regarding the example one may hopefully understand that one concrete instance of the specified structure may be '2Cu Ni 3Ag Cr 2Au Cr 2Au Ni 5Ag Cr 2Au Cr 2Au Mg'.) When generating a structure we always take care to ensure there are about 5 lead-material buffer layers on either side of the central structure.

Co/Cu 001 superlattices

Figure 5.3 depicts results for the Co/Cu 001 superlattices. All the considered structures are connected to Cu leads, and for the sake of brevity we do not explicitly write this information out, nor we specify the buffer-material layers. The WS radii are all set to 2.64.

When there is no disorder, the n[5Co 5Cu] case, a saturation in resistance, and hence in the GMR ratio, is clearly visible after about 20 repetitions of the 5Co 5Cu substructure (the GMR ratio there is about 200%). This is an expected behaviour since we actually have a periodic structure (with the 5Co 5Cu period) and Bloch's unhindered propagation is possible.

When a small amount of disorder is introduced, the n[8-9Co 5-6Cu] case, the picture changes in that no saturation seems to exist in the considered range of n. The resistance in the AP configuration grows linearly (ohmic-like) with n, while the resistance in the P configuration grows only very slowly, thanks to a good matching of the spin-up Co electronic structure to the Cu structure. The GMR ratio then also grows linearly with n in the considered range. The picture, to the extent we tested, is stable: when another instance of the same (probabilistic) prescription is taken, the resistance values do not change much; in the figure there are two instances drawn for thicknesses $n = 10, 15, 20, 25, 30$. We note that although we have not found signs of GMR saturation, if the P and AP resistances keep growing linearly with the thickness, eventually the saturation has to appear, as is clear from the defining expression for GMR.

Even greater disorder, modelled by the structure n[2-6Co 5-8Cu], leads to higher resistance values and GMR ratios. The dependency on n seems to deviate from linearity, being stronger. Again, in the figure we show results of at least two instances for each thickness, for $n = 30$ there are even five. The variance of the results thus seem to be reasonably limited.

In the last row of Fig. 5.3 we tried to check for what happens if instead of using a smooth selfconsistent electronic structure derived from precalculated interfaces we use bulk electronic-structure parameters for each atom. It seems the character of the figure is very similar to the previous 'smooth' case, however, the numerical values are smaller. This is slightly surprising to the author (who expected right the opposite in the case of resistances).

Concerning the SO-interaction impact, we still see nothing really serious. The GMR ratios turn out to be smaller when the interaction is included, by about 40% in the most disordered case and largest considered thickness.

The numbers used to specify the intervals of one-material thicknesses in the just discussed systems were chosen to be the same as Mathon used in (Math97), since we wanted to compare his and our results. The results show similarities, but i) we find slower saturation in the ordered case with slightly higher saturated GMR (his is around 1.5), and more seriously ii) we do not observe exponential GMR growth in the most disordered, n[2-6Co 5-8Cu], case. Mathon quite convincingly argues—and his results confirm—that the system should enter the Anderson-localization regime. Unfortunately, and so far for an unknown reason, we have not seen this happen. Actually, the last row in Fig. 5.3 with abrupt (sharp) changes in atoms' electronic structure was an attempt to come closer to Mathon's model. However, apparently, no other behaviour was induced by this change.

Co/Ni 001 superlattices

Figure 5.4 shows resistance dependence on thickness for Co/Ni 001 superlattices, with growing disorder in the systems. All these systems are connected to Co leads and WS radii of all the atoms were set to 2.61. Since whole systems are now magnetic, we do not speak about parallel or anti-parallel configuration, instead, any part of a system is magnetized in the same direction, here

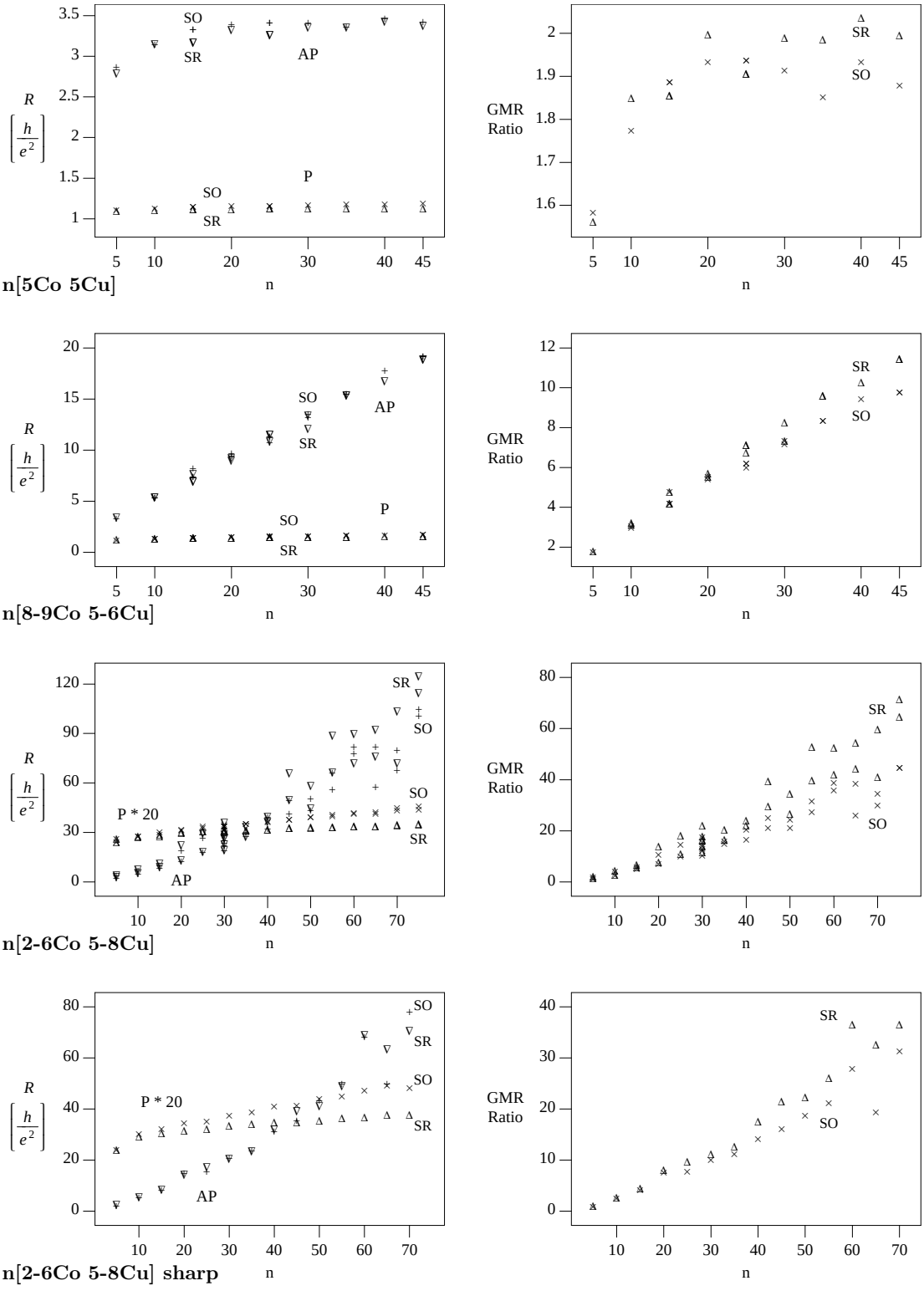


Figure 5.3: Resistances (per a surface atom) and GMR ratios (now *not* multiplied by 100) of Co/Cu 001 superlattices as a function of the system's thickness specified by the number 'n' of repetitions of a basic substructure. All the systems have attached Cu leads with at least 5 Cu atomic buffer layers on each side, and the WS radius has been chosen 2.64 for all atoms. The triangular symbols (∇, Δ) refer to a scalar-relativistic (SR) calculation, the cross-like symbols (+,×) to a calculation taking the SO interaction into account. For resistances, Δ, × symbols refer to the P configuration, while ∇, + to the AP one; for GMR ratios, only distinction between SR and SO calculation is conveyed by the symbols.

along the z axis, i.e., perpendicular to the 001 layers. From our perspective, the most interesting feature of the figure is, finally, a visible departure of SR and SO results. For the most disordered case of the n[1-5Co 6-10Ni] prescription and thickness $n=80$ we see the resistance with SO interaction included is three times higher compared to the SR calculation. Furthemore, from the rather different slopes we may infer the difference would grow for larger thicknesses.

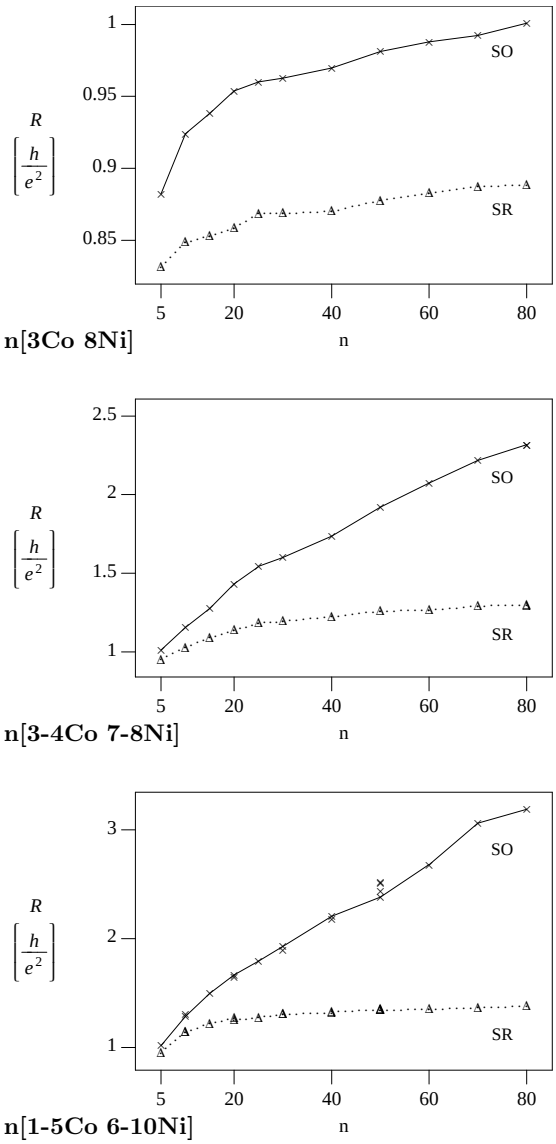


Figure 5.4: Resistances (per a surface atom) of Co/Ni 001 systems with Co leads as a function of structure thickness and growing disorder. For the most disordered case, two instances of the probabilistic prescription were considered for $n=10,20,30,40$ and four instances for $n=50$ to estimate variance of the results.

In Figure 5.5 we present our results for anisotropic magnetoresistance (AMR) for thickness $n=40$ and different disorders. As AMR is conditioned by the presence of the SO interaction, the figure is mainly meant to demonstrate how the anisotropy grows with disorder. For disordered systems we show results for three different instances of each probabilistic prescription to at least grossly estimate the variance of results we should expect. An interesting feature of these disordered structures is the existence of a resistance minimum in the $z\rightarrow x$ rotation at $\theta\approx 0.3\pi$, for which, however, we do not provide any explanation.*

* We have found a very similar picture also for a disordered Ni-Fe system, particularly 40[6-10Ni 1-2Fe] connected to Cu leads.

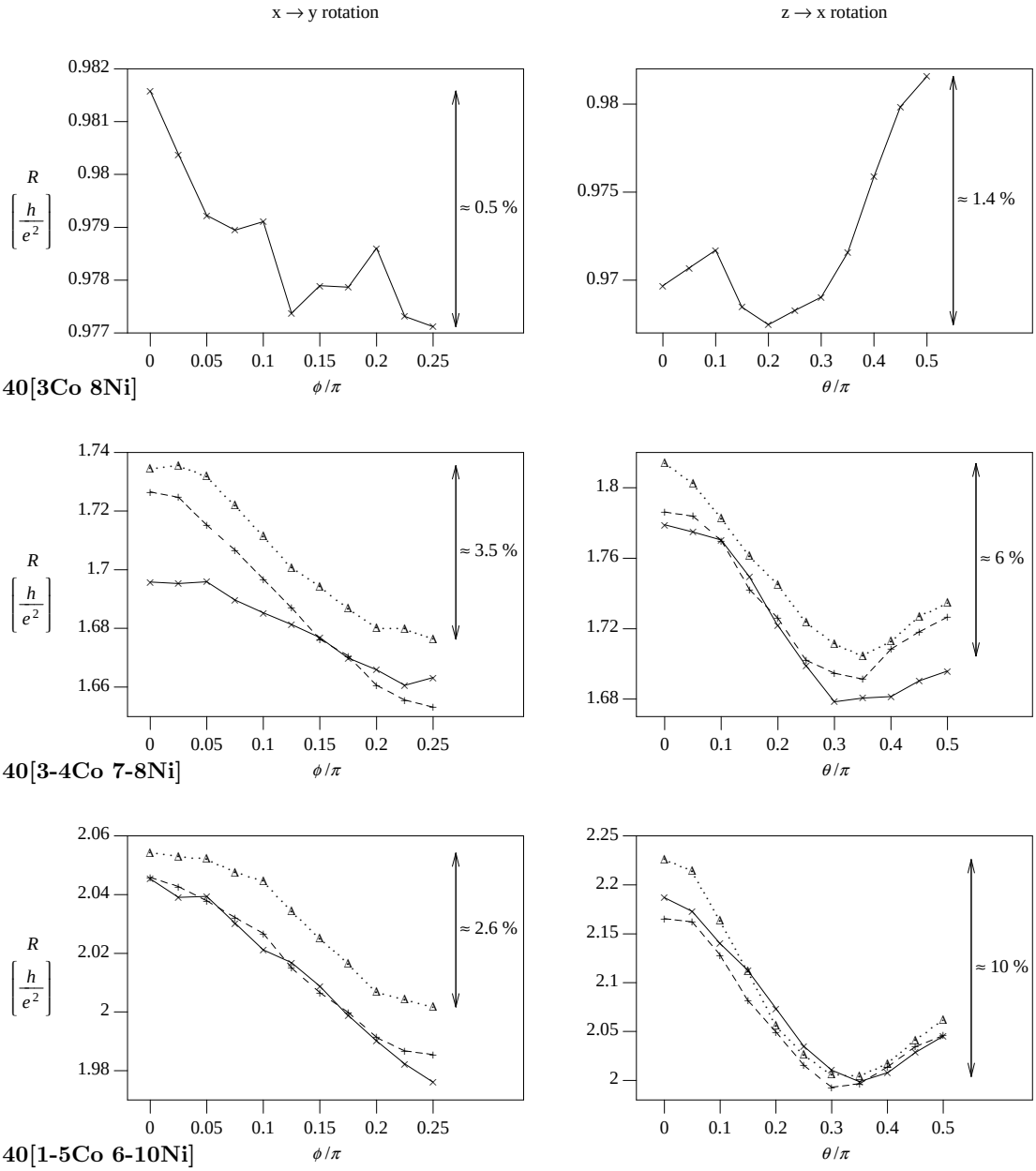


Figure 5.5: Anisotropic magnetoresistance of Co/Ni 001 superlattices with Co leads. For disordered system we consider three instances, distinguished by different symbols. Indicated is the span of the Δ-instance resistance.

Encouraged by these results we have finally built up a structure with two magnetic parts composed of disordered Co/Ni systems just studied and separated with ten atomic layers of Cu to allow study of GMR. We considered such a system connected both to Co and Cu leads with very similar results, in Figure 5.6 Co-leads results are shown. As is clear from the GMR-ratio results for the case with greater disorder, *neglect of the SO interaction in this system leads to a completely erroneous, by two orders of magnitude wrong GMR-ratio estimate*.

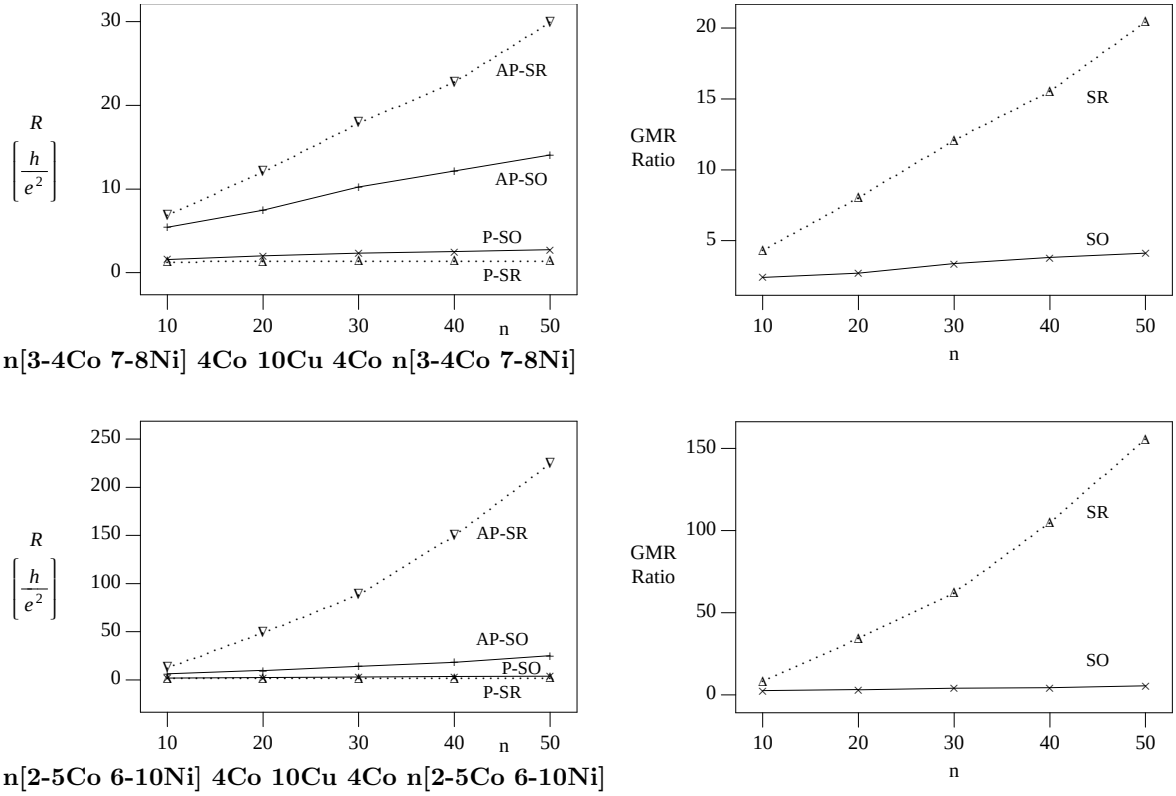


Figure 5.6: Resistances and GMR ratios for disordered layered systems in which the SO interaction seems to have a remarkable effect. The indicated systems were connected to Co leads.

6

Systems with a Tunnelling Barrier

6.1. Fe/MgO/Ag system

The first layered system with a tunnelling barrier in which we ever tried the influence of the spin-orbit interaction on transport was a Fe/5MgO/Ag 001 multilayer. The reason for choosing this particular system was its close resemblance to a well-studied Fe/MgO/Fe system, see e.g. (Math01, Butl01, Heil08, Tsym07, Khan08, Heil06), and the simplifying feature that it has only one ferromagnetic electrode. We knew that at a Fe/MgO interface a localized-to-surface state exists in the Fe minority channel, (Tius04), and that such states located at opposite sides of the barrier in a Fe/MgO/Fe structure may form bonding and antibonding hybrids. The localized interface states, even if only weakly coupled to the leads' propagating states, transform into interface resonances and lead to sharp spikes (so-called hot-spots) in the $\boldsymbol{k}_{\parallel}$-resolved conductance, (Bela05). (Similarly, (Wunn02) discussed the existence of hot-spots in Co/vacuum/Co tunnel junctions.) We wanted to avoid complications arising from having two resonances at (originally) exactly the same energies (we calculate conductance at zero bias) and opted for Ag as a replacement for one Fe lead. Silver's natural fcc lattice constant 4.09 Å is favourable to match that of MgO, 4.21 Å (3% misfit, even smaller than between MgO and Fe, which is about 3.5%). Furthemore, we really wanted to concentrate on the tunnelling anisotropic magnetoresistance (TAMR) effect, for which only one ferromagnetic part is needed, (Goul04, Chan07a).

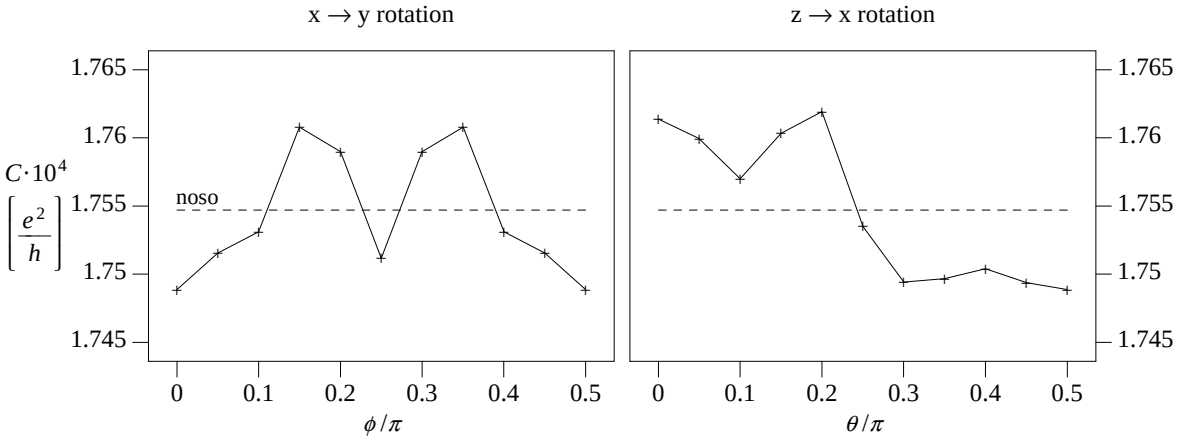


Figure 6.1: Calculated conductances, per a surface atom, of the Fe/5MgO/Ag system. On the left the rotation of Fe magnetization is changed (in-plane) from the *x*-direction to the *y*-direction. On the right the rotation from the (out-of-plane) *z*-direction to the *x*-direction is considered. Labelled 'noso' is the calculated conductance when the SO interaction is disregarded.

We consider both the interfaces to be ideal, the geometry for the Fe/MgO interface was taken from (Meye02), and the geometry for the MgO/Ag interface from (Giov02). Specifically: Transverse distances (the 'in-plane lattice constant') are fixed by that of bcc Fe, 2.86 Å, the Fe 001 layers keep their natural separation 1.43 Å but for the last Fe layer, which is 1.69 Å away from the last-but-one Fe layer. The first MgO layer is 2.35 Å away from the last Fe layer, the distances between individual MgO layers are 2.24 Å, 2.17 Å, 2.17 Å, 2.24 Å. The first Ag layer is 2.4 Å away from the last MgO layer, and the distance between all other Ag layers is 2.03 Å. At the Fe/MgO interface the O atoms are above the Fe atoms and we put an empty sphere 1.9 Å away from the interface Mg atom, towards the last Fe layer. At the MgO/Ag interface the Ag atoms lie above the O atoms, and again we use an empty sphere at the distance 2.2 Å from the Mg atom towards the Ag layer. Note that both the cubic MgO and Ag fcc 001 systems are rotated by 45° with respect to the Fe bcc 001 system.

In the electronic structure calculation the LMTO valence basis consisted of *s*-, *p*-, *d*- like orbitals. The low-lying O-2*s* orbitals were treated as core orbitals. As mentioned in the geometry description, empty spheres were used to ensure better space filling.

The results of conductance calculations are summarized in Fig. 6.1. (Our *x* and *y* axes run along Fe (100) and (010) directions, respectively, the *z* axis is perpendicular to the layers, i.e. in the Fe (001) direction.) Circa 2×10^4 $\boldsymbol{k}_{\parallel}$ points in the full 2D Brillouin zone were used. In the figure we see a nice demonstration of the C_{4v} in-plane symmetry. However, the observed TAMR effect is rather small, about 1%.

With the smallness of the observed TAMR results we have then decided to abandon the system and rather investigate systems with heavier atoms. With hindsight (having the experience from investigation of the Fe/GaAs/Ag system below), however, the author now thinks we should investigate the system a bit more, especially try to increase the number of MgO layers as well as the number of $\boldsymbol{k}_{\parallel}$ vectors and check more carefully the behaviour of the conductance near the $\bar{\Gamma}$ point. [The main inspiration for the author's work was (Butl01) and 5 MgO layers seemed to be i) enough to be in a tunnelling regime (see Fig. 16 in the reference), and ii) few enough that the structure could be grown—for thicker MgO layers the misfit dislocations form to partially relieve the 3.5% compressive in-plane strain that arises due to the larger MgO lattice constant than that of Fe. However, in spite of this fact, systems with larger barriers (of 8, 12 MgO layers) can be found in literature, and can be, if not practically then at least theoretically, interesting.]

6.2. Fe/GaAs/Ag system

(Some of the results of this section have been published in (Sýk12a), now 'in press', and majority of the detailed study below has been submitted for publication, (Sýk12b)).

Introduction

The TAMR effect observed in conventional Fe/GaAs/Au tunnel junctions (Mose07) proved to be rather weak which prevents its direct use. Similarly, a complementary phenomenon, namely, the current (or voltage) induced spin-transfer torques (Miro10), which might be employed for magnetization switching in the FM/I/NM devices, requires a large TAMR effect as a necessary prerequisite for efficient spintronic devices (Manc11). Recent attempts to enhance the TAMR values included, e.g., modification of the surface structure of the semiconductor layer during the growth process (Uemu11) or the use of an antiferromagnetic metal instead of the ferromagnetic one in the magnetic electrode (Park11). The latter approach yields strongly enhanced TAMR values, which, however, could be observed only at low temperatures, despite the much higher Néel temperature of the antiferromagnetic material.

On the theoretical side, a number of various topics were addressed in the framework of phenomenological models. These approaches discussed the role of the anisotropic density of states of the FM electrode (Goul04), the symmetry properties and the interplay of the Rashba and Dresselhaus contributions to the SO interaction, the effect of external magnetic fields and applied bias voltages (Mose07, Mato09, etc. *Ab initio* calculations of the TAMR were carried out for tunnel

junctions with Fe and Cu electrodes separated by vacuum (Chan07a) and GaAs (Chan07b) as well as for the Fe/GaAs/Au trilayers (Eber10). The first-principles studies proved that interface states (resonances) formed at the FM/I interface play an important role in the TAMR phenomenon. (FM denotes a ferromagnetic metallic electrode, I denotes a non-magnetic insulating barrier, finally NM will stand for a non-magnetic metallic electrode.)

All existing theoretical studies ascribe the TAMR primarily to electronic properties of the FM electrode, the tunnel barrier and their interface, whereas the NM electrode and the I/NM interface are considered to be of secondary importance. In the following we show by means of first-principles calculations for the Fe/GaAs/Ag system and by the study of a simple tight-binding (TB) model Hamiltonian in the next section, that the standard picture of the TAMR is not generally valid. In particular, we find that interface resonances at the I/NM interface—when hybridized with those at the FM/I interface—can yield high TAMR values. Moreover, we predict a non-monotonic dependence of the TAMR value on the thickness of the tunnelling barrier and discuss its physical origin.

Model and methods

The structural model of the Fe/GaAs/Ag(001) tunnel junction represents a simple generalization of the model for the Fe/GaAs/Cu trilayers, (Chan07b). The slab of the zinc-blende (zb) structure of GaAs is attached epitaxially to semiinfinite leads of the body-centred cubic (bcc) Fe and the face-centred cubic (fcc) Ag; the atomic planes are the (001) planes of all three structures. The structure of the Fe/GaAs interface has been assumed without any reconstructions and layer relaxations on the basis of the ideal ratio of the lattice parameters $a_{\mathrm{zb}}/a_{\mathrm{bcc}} = 2$, that is satisfied with a good accuracy for GaAs and Fe. The structure of the GaAs/Ag interface was also assumed without reconstructions; it employed the ideal ratio $a_{\mathrm{zb}}/a_{\mathrm{fcc}} = \sqrt{2}$ and the mutual rotation of both bulk structures by $\pi/4$ around the common (001) axis. The interplanar distance between the adjacent atomic planes of the GaAs and Ag parts was set equal to the arithmetic average of the distances in the bulk GaAs ($a_{\mathrm{zb}}/4$) and the bulk Ag ($a_{\mathrm{fcc}}/2$). The *z* axis of the coordinate system is perpendicular to the atomic planes, while the *x* and *y* axes coincide respectively with the [100] and [010] directions of the fcc lattice, i.e., they point along the $[1,\pm 1,0]$ directions of the bcc and zb lattices. The GaAs slab contains *n* As-atomic (001) planes with As-termination on both sides; the investigated systems are thus abbreviated as Fe/As(GaAs)$_{n-1}$/Ag(001).

In the electronic-structure calculation empty spheres were used for an efficient treatment of the open zb structure. The low-lying Ga-3*d* orbitals were treated as core orbitals so that the valence basis comprised the Ga-4*d* orbitals, (Ture06). This choice leads to a good description of the bulk bandstructure of the GaAs concerning both the band gap (around 1.2 eV) and the valence bandwidth (around 6.8 eV), which agree quite well with measured data, (Chia80).

For thicknesses of the GaAs barriers relevant in experiments ($n > 20$), the tunnelling current is carried mainly by the states with $\boldsymbol{k}_{\parallel}$-vectors from a small central region of the whole 2D BZ. A sufficiently dense mesh of sampling points (corresponding to 5.4×10^{5} $\boldsymbol{k}_{\parallel}$-vectors in the full 2D BZ) has been used to get reliable values of the CPP conductances.

Electronic structure of Ag/GaAs/Ag(001) systems

In recent theoretical studies, the Fe/GaAs(001) system has been attached to a hypothetical bcc(001) Cu electrode since the bulk bcc Cu has a free-electron-like bandstructure and the GaAs/Cu interface has a featureless transmission function (Chan07a, Chan07b). This setup is advantageous for investigations of the role of the Fe/GaAs(001) interface state (resonance) lying in the minority-spin channel at the Fermi energy. This interface resonance gives rise, e.g., to a reversal of spin polarization of the tunnelling current with applied voltage (Chan07b) and it can lead to a pronounced TAMR effect (Chan07a). Motivated by these facts, we have focused on electronic properties of the non-magnetic GaAs/Ag(001) interface prior to the study of the Fe/GaAs/Ag junctions.

We have thus studied systems Ag/GaAs/Ag(001) with different thicknesses and both terminations (Ga or As) of the GaAs barrier. These studies were done first without the SO interaction.

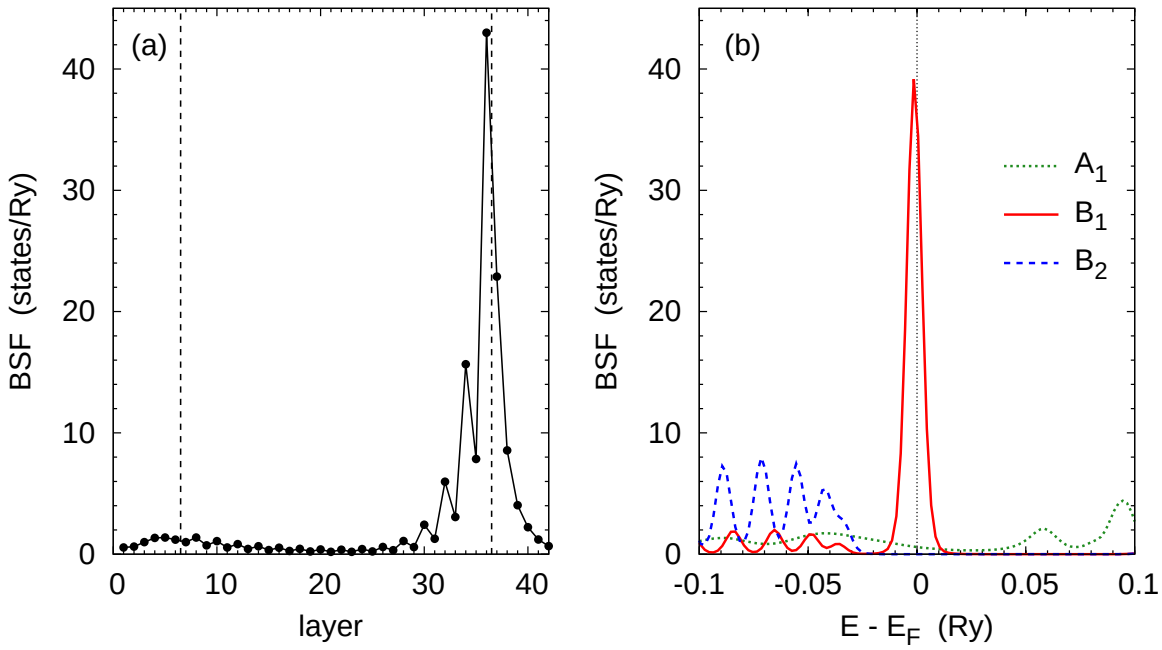


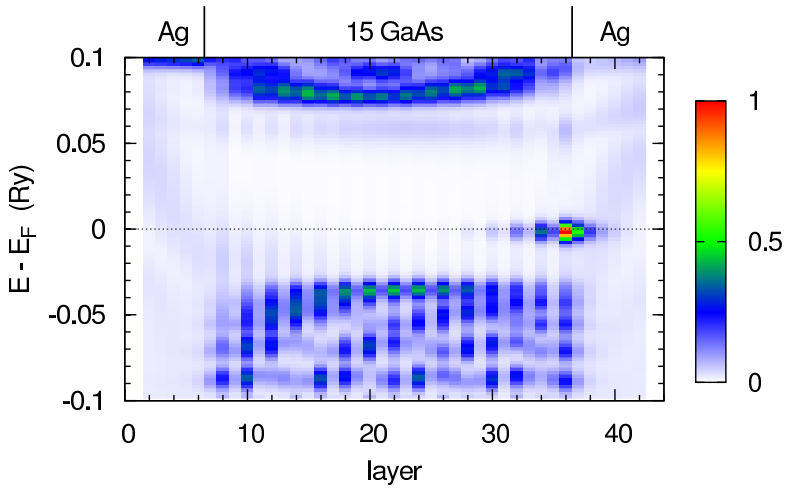
Figure 6.2: The Bloch spectral functions (BSF) of the system Ag/$(GaAs)_{15}$/Ag(001) for $\boldsymbol{k}_{\parallel} = \bar{\Gamma}$: (a) The layer-resolved BSF for the energy E located 1.5 mRy below the Fermi level. The dashed vertical lines denote the left (Ag/Ga) and right (As/Ag) interfaces. (b) The BSFs of the As interface atom as functions of energy, resolved according to the irreducible representations of the point group C_{2v}. The vertical line denotes the position of the Fermi level of fcc Ag.

Since the tunnelling current is mostly carried by states with $\boldsymbol{k}_{\parallel}$-vectors in the vicinity of the $\bar{\Gamma}$ point, we paid special attention to $\boldsymbol{k}_{\parallel} = \bar{\Gamma}$. Figure 6.2(a) displays the layer-resolved Bloch spectral function of the system Ag/$(GaAs)_{15}$/Ag(001) for an energy slightly below the Fermi energy ($E - E_F = -1.5\,\mathrm{mRy}$). One can see a clear indication of an interface state at the GaAs/Ag interface, i.e., at the As-terminated GaAs barrier. In more colours the same can also be seen in Figure 6.3(a). The amplitude of the interface state is maximal in the As layer adjacent to the Ag electrode. A similar interface state was found in the Au/GaAs/Au(001) system, located however 10 mRy below the Fermi energy, Figure 6.3(b), whereas in the Cu/GaAs/Cu(001) with bcc Cu, no such state appears, Figure 6.3(c), in agreement with previous studies (Chan07b). It should be noted that no similar interface state was found at the Ga-terminated boundary of the GaAs barrier in a wide energy interval around the Fermi energy (inside the band gap of GaAs), irrespective of the electrode metal (Cu, Ag, Au).

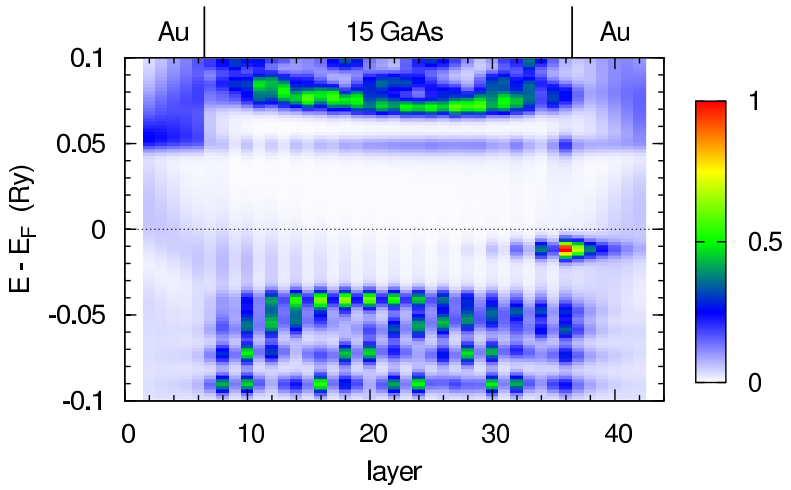
The origin of the interface state can be understood from the Bloch spectral functions of the boundary As site resolved with respect to the symmetry given by the point group of the interface, namely, the C_{2v} group. This group has four one-dimensional irreducible representations: A_1, A_2, B_1, and B_2, (Hein60), of which only the A_1 (subspace spanned by orbitals s, p_z, d_{z^2} and $d_{x^2-y^2}$) is compatible with the symmetry of propagating states of the Ag(001) electrode at the Fermi level. As can be seen in Figure 6.2(b), the interface state is entirely of the symmetry B_1 (subspace spanned by orbitals p_y and d_{yz}), which is incompatible with the propagating states available in the Ag electrode. This incompatibility is an important factor, since the semiinfinite metallic electrode acts essentially like a vacuum half-space in the formation of the interface state.

In the Landauer picture of the ballistic transport, interface states do not contribute directly to the system conductance, since the latter is given solely by the transmission coefficients between the

(a)



(b)



(c)

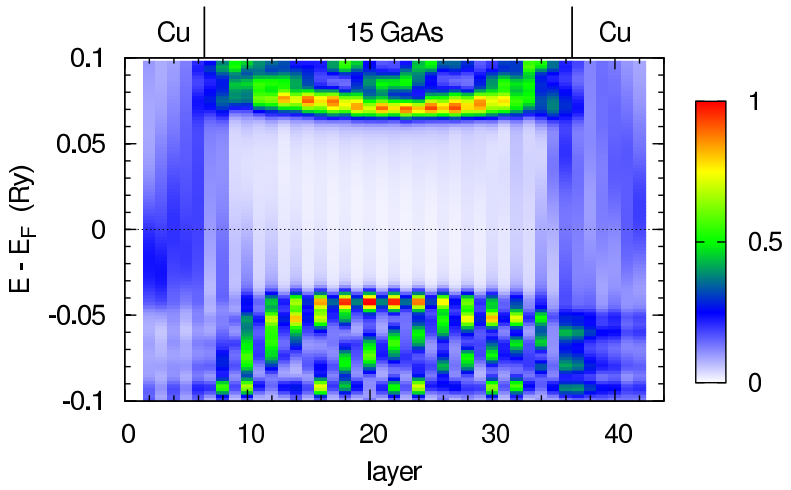


Figure 6.3: Layer-resolved Bloch spectral functions for $\boldsymbol{k}_{\parallel} = \bar{\Gamma}$ and systems with a $(\mathrm{GaAs})_{15}$ barrier connected to different leads as discussed in the main text. Drawn to illustrate the presence or lack of interface states. (The colour axes are for orientation only. They are normalized so that the maximum BSF in each figure is equal to one.)

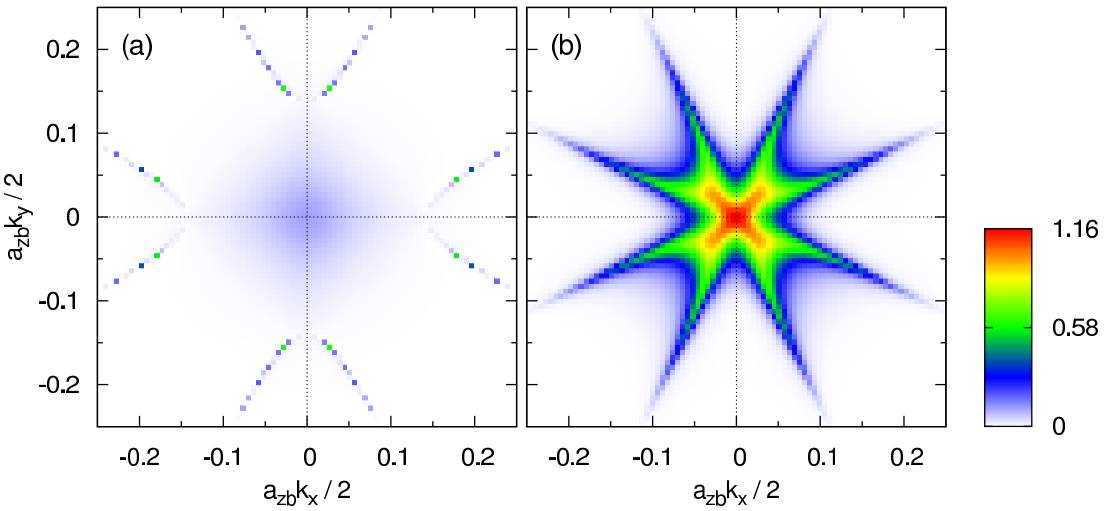


Figure 6.4: The $\boldsymbol{k}_{\parallel}$-resolved transmission of the Ag/As(GaAs)$_{14}$/Ag(001) system at the Fermi energy: (a) without the SO interaction and (b) with the SO interaction. The coloured scale of the $T(\boldsymbol{k}_{\parallel})$ for both panels is shown on the right.

propagating channels of the leads, (Datt97). However, if an interface state is coupled, e.g., by a weak interaction to the propagating states, it can become a resonance with a non-negligible effect on the conductance. In the present case, the SO interaction provides such a coupling of the B_1-like interface state to the A_1-like propagating state, which follows from an analysis of the double group C_{2v} and its irreducible representations: all spin-orbitals belong to the single additional two-dimensional representation of this double group, (Kost57). The influence of the interface resonances on tunnelling is especially strong in symmetric junctions with identical electrodes, where both resonances become hybridized across the tunnel barrier which enhances the conductance appreciably, (Wunn02). This phenomenon is illustrated in Figure 6.4, where the $\boldsymbol{k}_{\parallel}$-resolved transmissions $T(\boldsymbol{k}_{\parallel})$ are compared for the symmetric junction Ag/As(GaAs)$_{14}$/Ag(001) treated without and with SO coupling. The pronounced enhancement of the $T(\boldsymbol{k}_{\parallel})$ in vicinity of the $\bar{\Gamma}$ point is clearly visible; the total conductance of the junction increases by one order of magnitude due to the SO interaction. This result indicates importance of the GaAs/Ag interface for the transport behaviour of the Fe/GaAs/Ag system.

Conductances and TAMR of Fe/GaAs/Ag junctions

Figure 6.5 shows the local density of states (DOS) of the central GaAs layer (two neighbouring atomic planes) inside the Fe/As(GaAs)$_{10}$/Ag(001) junction. The shape of the DOS is bulk-like, with the band gap clearly formed around the Fermi energy and with negligible spin polarization. These features prove that the junction is in a tunnelling regime with metal-induced gap states significantly suppressed for this and higher barrier thicknesses. The tunnelling regime is also manifested by an exponential decay of the conductance with the increasing GaAs thickness n, plotted in Figure 6.6(a) for three orientations of the iron magnetic moment, i.e., along the x, y, and z axis. For a given n, the conductance is obviously sensitive to the magnetization direction which leads to the TAMR effect. The orientational dependence was studied in detail for magnetization directions in the atomic planes, i.e., as functions of the angle ϕ with the fixed value of $\theta = \pi/2$. The resulting angular dependences, shown in Figure 6.7 for $n = 23$ and $n = 33$, reflect the two-fold rotation symmetry (point group C_{2v}) of the system, in full agreement with previous calculations (Eber10) and measurements (Mose07) performed for similar Fe/GaAs/Au(001) junctions.

The in-plane TAMR, defined from the angular dependence of the conductance $C(\theta,\phi)$ as $\mathrm{TAMR} = [C(\pi/2,\pi/2) - C(\pi/2,0)]/C(\pi/2,0)$, is presented in Figure 6.6(b) as a function of the barrier thickness n. The calculated TAMR effect is quite large, exceeding 10%, in reasonable agreement with the values calculated for the Fe/GaAs/Au system, (Eber10), but about two orders of

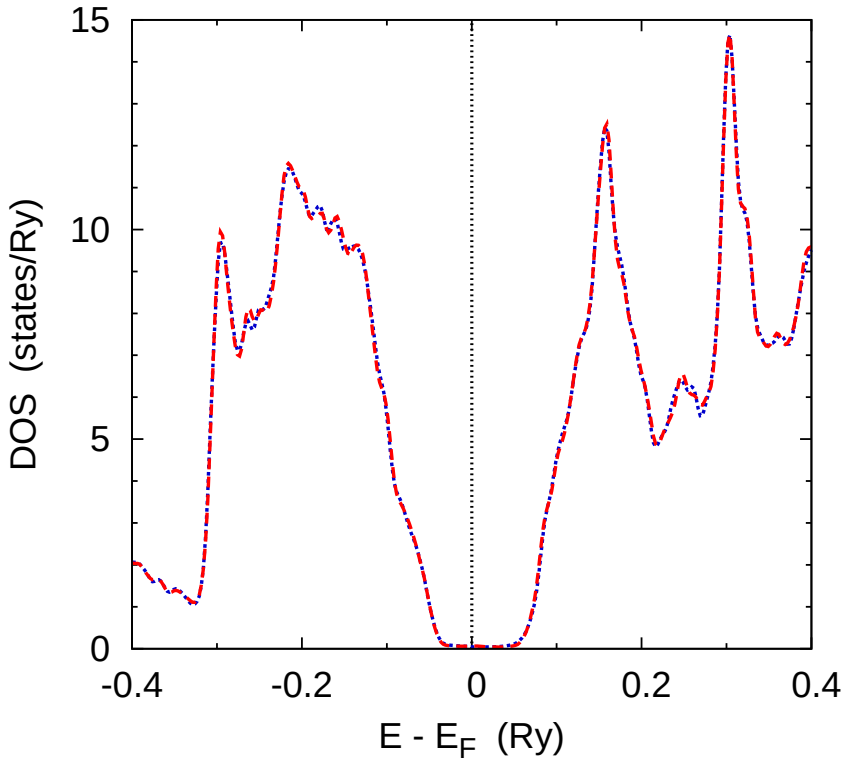


Figure 6.5: The density of states in two neighbouring Ga- and As-atomic planes in the middle of the $Fe/As(GaAs)_{10}/Ag(001)$ system for the majority spin ($\cdots$, blue) and the minority spin ($---$, red). The vertical line denotes the position of the Fermi level.

magnitude stronger than the experimentally observed TAMR values (Mose07). Moreover, the calculated thickness dependence is non-monotonic with a maximum obtained around $n = 27$ corresponding to the GaAs thickness of about 7.5 nm. In order to identify possible reasons for the calculated high TAMR values and the non-monotonic thickness dependence, additional analysis is needed.

position	TAMR (%)
—	41.2
inside Fe	39.4
at Fe/As	3.7
inside GaAs	41.4
at As/Ag	4.7
inside Ag	32.2
at Fe/As and As/Ag	3.6

Table 6.1: Dependence of the in-plane TAMR on the position of two atomic planes with disorder in the $Fe/As(GaAs)_{26}/Ag(001)$ junction. The first row corresponds to the ideal system.

In general, discrepancy between the calculated and measured transport properties of epitaxial magnetic multilayers can often be ascribed to imperfect atomic structure at the interfaces. The correct treatment of structure defects on an *ab initio* level employs either supercell techniques (Xia01, Drch02) or effective medium approaches (Carv06) combined with a particular microscopic model of the structure imperfection. In order to get a rough insight into the sensitivity of the TAMR to the quality of interfaces, we adopted here a simplified approach. We have simulated chemical disorder in the system by a finite imaginary part $\varepsilon > 0$ of the energy arguments $z = E_F \pm i\,\varepsilon$

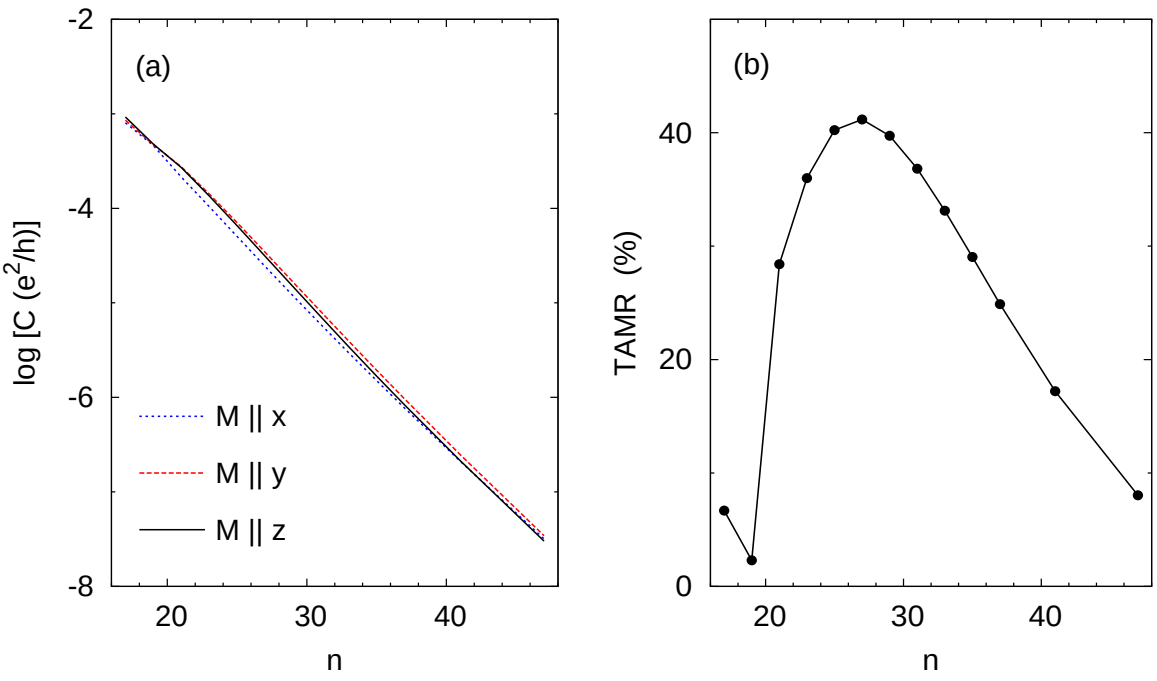


Figure 6.6: Dependence of the transport properties of the tunnel junctions Fe/As$(GaAs)_{n-1}$/Ag(001) on the thickness n: (a) the conductances for the Fe magnetization pointing along the x, y and z axis, and (b) the corresponding in-plane TAMR.

of the potential functions in the TB-LMTO Kubo-Landauer formalism (Kudr00, Carv06). This modification was used only in a few selected atomic planes of the whole system: in two neighbouring planes located at a single interface (Fe/As or As/Ag), at both interfaces, inside the GaAs barrier, or inside the metallic electrodes. The value of $\varepsilon = 5\,\mathrm{mRy}$ was used in all cases. The results for $n = 27$ (the thickness corresponding to the maximum TAMR effect in the perfect junctions) are collected in table 6.1. One can see that the disorder deep inside each part (Fe, GaAs, Ag) of the junction has only a minor influence on the resulting TAMR. However, the interface disorder reduces the TAMR value quite significantly, which proves that at least a part of the difference between the large theoretical values of the TAMR and the much weaker observed effect is due to the interface roughness. Moreover, both interfaces influence the TAMR effect to a similar extent, see table 6.1, which indicates that they are of equal importance for the calculated trend of the TAMR (Figure 6.6(b)).

Further insight into the obtained results follows from the $\boldsymbol{k}_{\parallel}$-resolved transmissions $T(\boldsymbol{k}_{\parallel})$. Figure 6.8 shows this quantity for three barrier thicknesses, $n = 19$, $n = 27$, and $n = 47$, and for the magnetization directions along the x and y axis. Only a small region around the $\bar{\Gamma}$ point is included in the figure while the rest of the entire 2D BZ (defined by $a_{\mathrm{bcc}}\,|k_{x,y}| \leq \pi/\sqrt{2} \approx 2.22$) is unimportant for the tunnelling. The maximum TAMR ($n = 27$) corresponds to a few hot spots in the $T(\boldsymbol{k}_{\parallel})$ plots, see Figure 6.8(c, d); their positions and contributions to the total conductance are strongly sensitive to the direction of the iron magnetic moment. The lower TAMR values for thinner barriers are due to non-negligible contributions of bigger 2D regions to the total conductances, see the case of $n = 19$ in Figure 6.8(a, b). On the other hand, the lower TAMR values for very thick barriers ($n > 40$) are due to the same hot spots with non-zero $\boldsymbol{k}_{\parallel}$-vectors as found for intermediate thicknesses ($n \approx 27$) accompanied by another pronounced local maximum of the $T(\boldsymbol{k}_{\parallel})$ in the $\bar{\Gamma}$ point, see the case of $n = 47$ in Figure 6.8(e, f). The contribution of the latter maximum to the total conductance is little sensitive to the magnetization orientation which explains the reduction of the

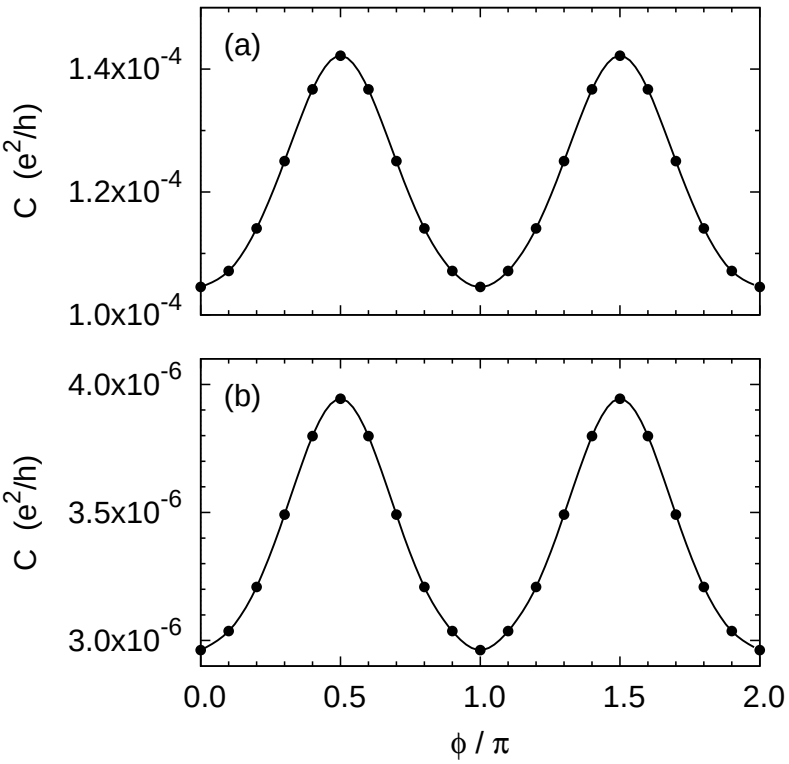


Figure 6.7: The conductance $C(\pi/2,\phi)$ of the systems Fe/As(GaAs)$_{n-1}$/Ag(001) as a function of the angle ϕ: (a) for $n=23$ and (b) for $n=33$.

TAMR effect.

The presence of the hot spots in the $\boldsymbol{k}_{\parallel}$-resolved transmissions is undoubtedly an important factor contributing to the high TAMR and to its non-monotonic dependence on the GaAs thickness. Similar hot spots appeared in various magnetic tunnel junctions (Wunn02, Dede02, Bela05, Tsym07) and were shown to be a direct consequence of a hybridization of two interface resonances across the barrier (Wunn02). In the present case, an interplay of these hybridization-induced hot spots with the contribution of the $\bar{\Gamma}$ point represents a new situation, relevant especially for large thicknesses of barriers with a direct band gap, such as MgO or GaAs (Dede02).

6.3. Hybridized interface resonances in a tight-binding model

In order to assess the role of the presence and hybridization of the interface resonances on the TAMR, we have formulated a simple TB model and investigated its properties. The atoms are placed in positions of a simple cubic lattice with the lattice parameter a; the atomic planes are the (001) planes. The active part of the FM/I/NM junction comprises $N+2$ atomic planes labelled by an index p, $p=0,1,\ldots,N+1$. The plane $p=0$ corresponds to a FM layer, the plane $p=N+1$ corresponds to a NM layer, and the tunnel barrier is represented by the planes $p=1,2,\ldots,N$. We assume a single orbital per site and spin and a spin-independent nearest-neighbour hopping between the orbitals. The hopping elements are different for pairs of atoms in neighbouring atomic planes (hopping t) and inside the same atomic plane (hopping $\tilde{t}$). The atomic energy levels of the tunnel barrier and of the NM layer are spin-independent while those of the FM layer are exchange-split. The direction of the exchange splitting is given by an in-plane unit vector $\boldsymbol{n}=(\cos\phi,\sin\phi,0)$. The FM layer is also influenced by a Rashba-type SO splitting derived from the canonical form $H^{\mathrm{SO}}\propto(\boldsymbol{p}\times\boldsymbol{\sigma})\cdot\mathbf{v}$, where the $\boldsymbol{p}=(p_x,p_y,p_z)$ denotes the momentum operator, the $\boldsymbol{\sigma}=(\sigma_x,\sigma_y,\sigma_z)$ are the Pauli spin matrices and the unit vector $\mathbf{v}=(0,0,1)$ is normal to the atomic planes. We assume full 2D translational symmetry so that the Hamiltonian of the system can be written after

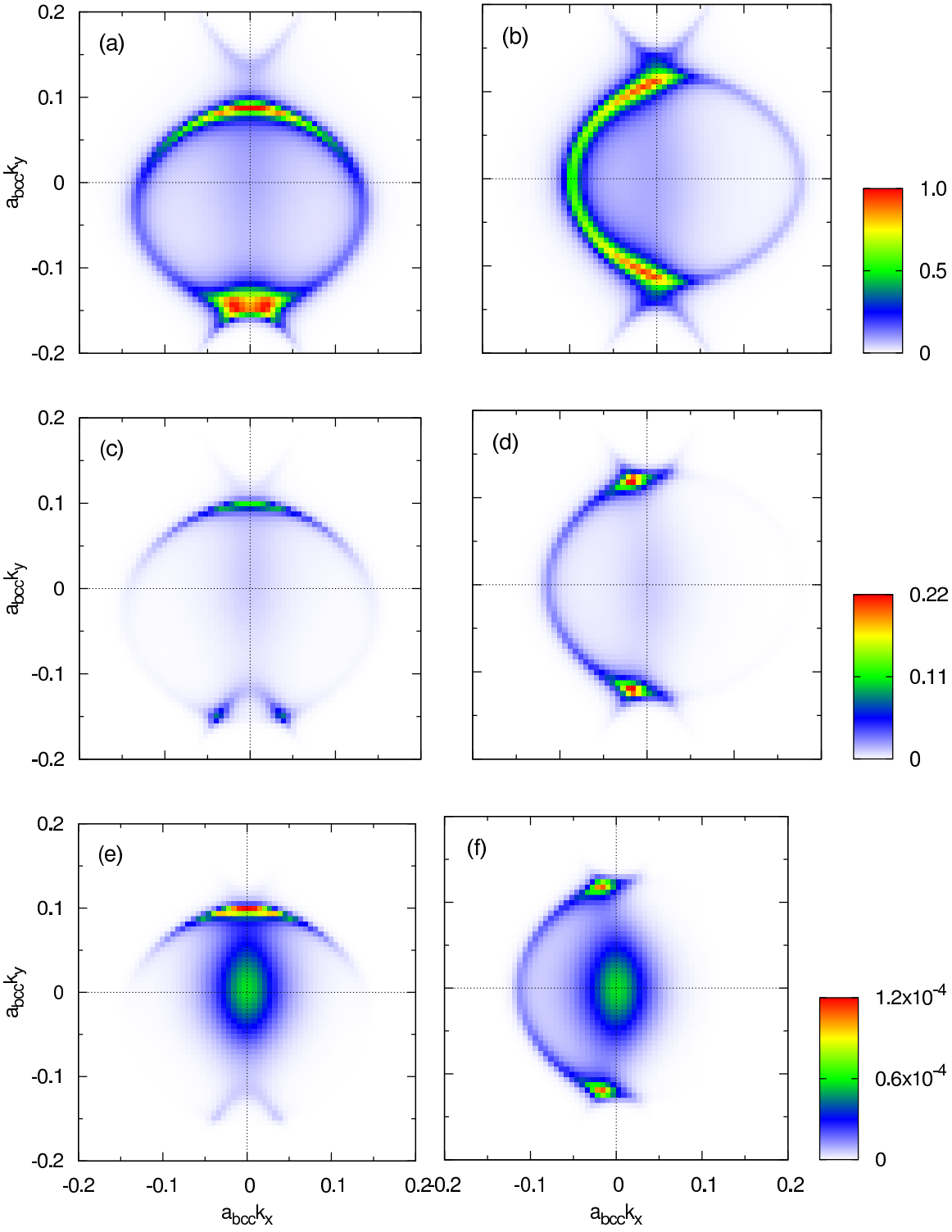


Figure 6.8: The $\boldsymbol{k}_{\parallel}$-resolved transmissions of the Fe/As(GaAs)$_{n-1}$/Ag(001) systems at the Fermi energy for $\theta=\pi/2$: (a) $n=19$, $\phi=0$, (b) $n=19$, $\phi=\pi/2$, (c) $n=27$, $\phi=0$, (d) $n=27$, $\phi=\pi/2$, (e) $n=47$, $\phi=0$, and (f) $n=47$, $\phi=\pi/2$. The coloured scales of the $T(\boldsymbol{k}_{\parallel})$, shown on the right, refer to both plots with the same n.

the 2D lattice Fourier transformation as:

$$H_{ps,p's'}(\boldsymbol{k}_{\parallel}) = \delta_{pp'}\left[h_{ss}^{(p)} - 2\tilde{t}[\cos(ak_x)+\cos(ak_y)]\delta_{ss'}\right] - \\ - \delta_{|p-p'|,1}\delta_{ss'}t + \delta_{p,0}\delta_{p',0}H_{ss'}^{SO}(\boldsymbol{k}_{\parallel}), \qquad (6.1)$$

where the s and s' are spin indices, $s,s'=\uparrow,\downarrow$, which refer to the global (fixed) spin quantization

axis along the (001) direction. The term $h_{ss'}^{(p)}$ comprises all on-site interactions in the p-th atomic plane and the last term describes the Rashba-type SO interaction in the FM layer ($p=0$). For the NM layer ($p=N+1$), the on-site term is given by $h_{ss'}^{(N+1)}=\varepsilon_{\rm NM}\delta_{ss'}$, and for the barrier layers ($p=1,2,\ldots,N$), it is given similarly as $h_{ss'}^{(p)}=\varepsilon_{\rm B}\delta_{ss'}$, where the parameters $\varepsilon_{\rm NM}$ and $\varepsilon_{\rm B}$ denote, respectively, the energy levels of the NM and barrier layers. The explicit form of the FM on-site term ($p=0$) is given by

$$h_{\uparrow,\uparrow}^{(0)} = h_{\downarrow,\downarrow}^{(0)} = \frac{\varepsilon_\uparrow+\varepsilon_\downarrow}{2}, \qquad h_{\uparrow,\downarrow}^{(0)} = h_{\downarrow,\uparrow}^{(0)*} = \frac{\varepsilon_\uparrow-\varepsilon_\downarrow}{2}\exp(-i\phi), \tag{6.2}$$

where the parameters $\varepsilon_\uparrow$ and $\varepsilon_\downarrow$ denote the exchange-split energy levels. The last term in (6.1) is given by

$$\begin{aligned} H_{\uparrow,\uparrow}^{\rm SO}(\boldsymbol{k}_\parallel) &= H_{\downarrow,\downarrow}^{\rm SO}(\boldsymbol{k}_\parallel) = 0, \\ H_{\uparrow,\downarrow}^{\rm SO}(\boldsymbol{k}_\parallel) &= H_{\downarrow,\uparrow}^{\rm SO}(\boldsymbol{k}_\parallel)^* = -\alpha[\sin(ak_y) + i\sin(ak_x)], \end{aligned} \tag{6.3}$$

where the parameter α scales the Rashba-like SO interaction. The effect of the semiinfinite FM and NM leads has been simplified by adding energy- and $\boldsymbol{k}_\parallel$-independent selfenergy operators to the on-site interactions of the NM and FM layers. These (retarded) selfenergies are given by

$$\begin{aligned} \Sigma_{ss'}^{\rm NM} &= -\frac{i}{2}\gamma_{\rm NM}\delta_{ss'}, & \Sigma_{\uparrow,\uparrow}^{\rm FM} &= \Sigma_{\downarrow,\downarrow}^{\rm FM} = -\frac{i}{4}(\gamma_\uparrow+\gamma_\downarrow), \\ \Sigma_{\uparrow,\downarrow}^{\rm FM} &= -\frac{i}{4}(\gamma_\uparrow-\gamma_\downarrow)\exp(-i\phi), & \Sigma_{\downarrow,\uparrow}^{\rm FM} &= -\frac{i}{4}(\gamma_\uparrow-\gamma_\downarrow)\exp(i\phi), \end{aligned} \tag{6.4}$$

where the parameters $\gamma_{\rm NM}$, $\gamma_\uparrow$ and $\gamma_\downarrow$ define the widths of the respective energy levels (resonance widths in the local spin reference system).

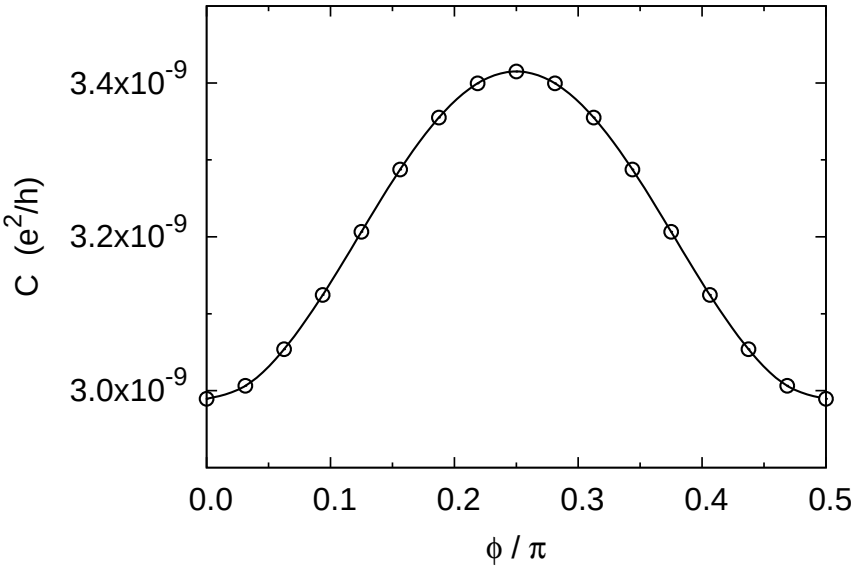


Figure 6.9: Dependence of the conductance of the FM/I/NM model on the angle ϕ for the barrier thickness $N=30$ and for the case 2.

The simplicity of the model allows one to choose easily its parameters in order to achieve the presence or absence of the resonance at either interface and for each spin channel. Here we confine ourselves to the case of one spin channel (spin-$\downarrow$) out of the resonance and with no propagating states in the FM lead; the latter condition is obtained by setting $\gamma_\downarrow=0$. We have considered four cases of the model. The first case, denoted as case 2, corresponds to the presence of two resonances: one at the FM/I interface in the spin-$\uparrow$ channel, the other at the I/NM interface. The particular values of the model parameters are: $t=0.48$, $\tilde{t}=0.03$, $\varepsilon_\uparrow=0.36$, $\varepsilon_\downarrow=-0.1$, $\varepsilon_{\rm B}=1.1$,

$\varepsilon_{NM} = 0.4$, $\gamma_{\uparrow} = 0.009$, $\gamma_{NM} = 0.007$, and $\alpha = 0.03$. Note that a small asymmetry has been intentionally introduced in the parameters $\varepsilon_{\uparrow}/\varepsilon_{NM}$ and $\gamma_{\uparrow}/\gamma_{NM}$ in order to simulate different properties of the FM and NM electrodes. The Fermi energy is set to zero, $E_F = 0$, which is located slightly below the bottom of the spectrum of the tunnel barrier, $V_B = \varepsilon_B - 2t - 4\tilde{t} = 0.02$. The second case, denoted as case 1F, corresponds to the case of one resonance, located at the FM/I interface for the spin-$\uparrow$ channel; its parameters coincide with the case 2 apart from the value of $\varepsilon_{NM} = 0.6$. The third case, denoted as case 1N, describes the situation with one resonance, present at the I/NM interface. This case is obtained from the case 2 by setting the value of $\varepsilon_{\uparrow} = 0.6$. The last case, denoted as case 0, refers to the absence of any resonance; its parameters are obtained from the case 2 by changing its two parameters, namely, $\varepsilon_{\uparrow} = 0.6$ and $\varepsilon_{NM} = 0.64$.

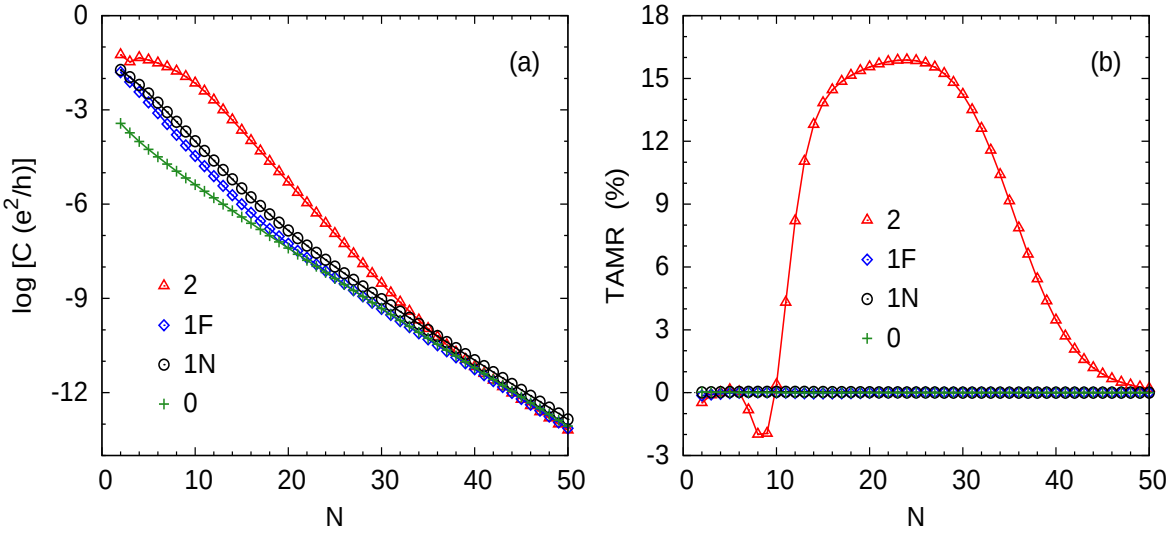


Figure 6.10: Dependence of the transport properties of the FM/I/NM model on the barrier thickness N for the four cases (for details, see text): (a) the conductance for $\phi = 0$, (b) the in-plane TAMR.

The angular dependence of the conductance $C(\phi)$ is plotted in Figure 6.9 for the case 2 and the barrier thickness $N = 30$. It is seen that the calculated dependence reflects the fourfold symmetry of the system (point group C_{4v}). The in-plane TAMR is thus defined as the ratio $[C(\pi/4) - C(0)]/C(0)$. The dependence of the conductances on the barrier thickness for all four cases and for $\phi = 0$ is shown in Figure 6.10(a). One can see clearly the effect of the resonances, pronounced especially for smaller thicknesses: the straight line for the case 0 is slightly modified by the presence of a single resonance (cases 1F and 1N), whereas the hybridization of two resonances is manifested by a strong enhancement of the conductance (case 2). The corresponding thickness dependences of the TAMR are depicted in Figure 6.10(b). It is seen that a sizeable TAMR effect is obtained only for the case 2 while the single resonances (on either side of the barrier) lead essentially to negligible TAMR values (cases 1F and 1N), similarly to the case 0. The hybridized interface resonances yield also a non-trivial dependence of the TAMR on the barrier thickness: small initial values for $N \leq 10$ are followed by a steep increase to a broad maximum for $20 \leq N \leq 30$ which is replaced by a final decrease for $N \geq 40$. This trend is qualitatively similar to that obtained for the Fe/GaAs/Ag system, see Figure 6.6(b).

The different regimes of the thickness dependence of the TAMR can be related to the corresponding $\boldsymbol{k}_{\parallel}$-resolved transmissions shown in Figure 6.11 for the case 2 and three values of N. For $N = 10$, the total conductances arise from contributions of substantial parts of the whole 2D BZ which leads to a modest TAMR effect for thin tunnelling barriers. For $N = 20$, the dominating contribution to the tunnelling is due to a narrow region (a hot spot), the position of which depends on the angle ϕ. The sensitivity of this sharp single local maximum to the angle ϕ gives rise to enhanced TAMR values for intermediate barrier thicknesses. For $N = 40$, the hot spots survive but

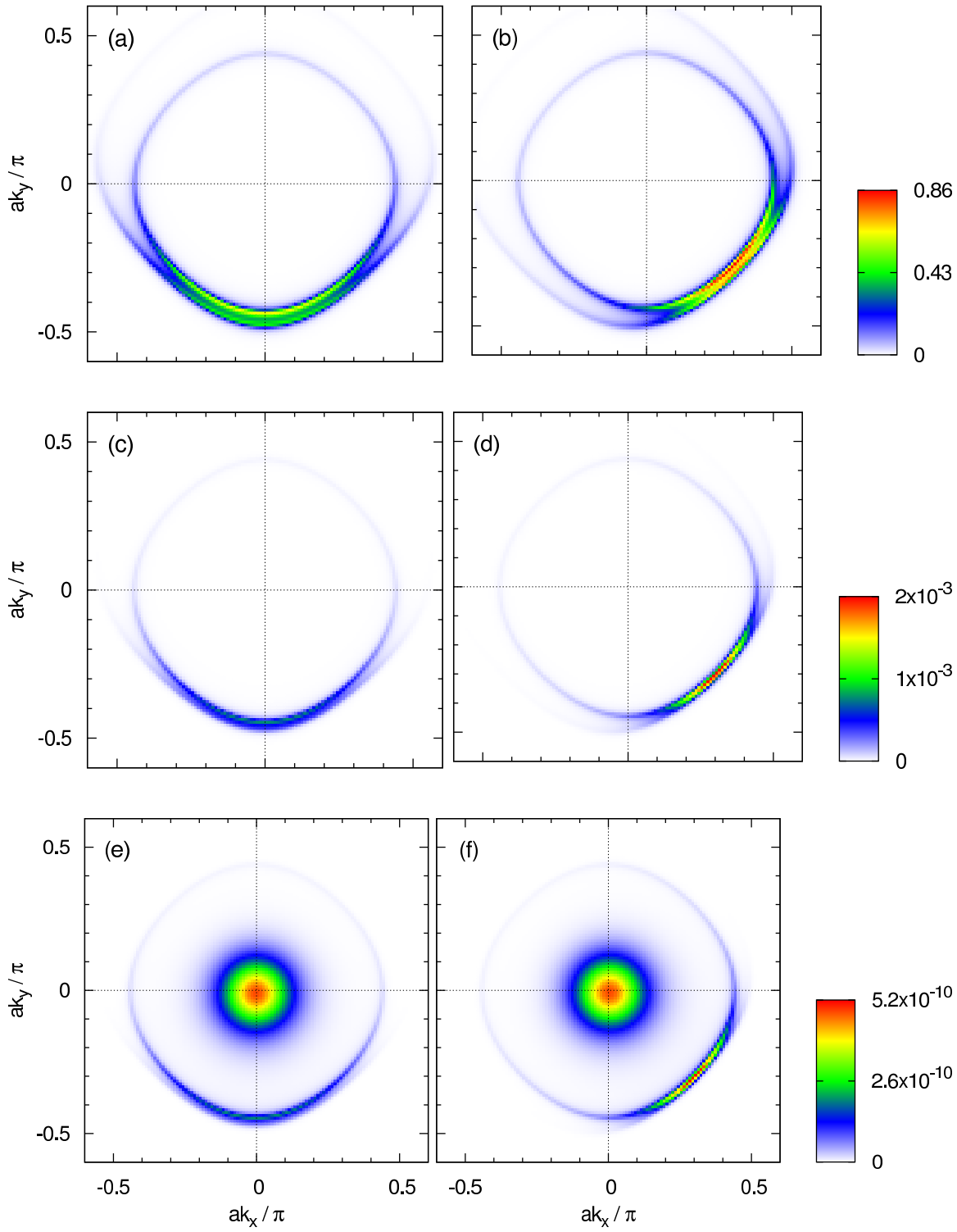


Figure 6.11: The $\boldsymbol{k}_{\parallel}$-resolved transmissions of the FM/I/NM model in the case 2: (a) $N=10$, $\phi=0$, (b) $N=10$, $\phi=\pi/4$, (c) $N=20$, $\phi=0$, (d) $N=20$, $\phi=\pi/4$, (e) $N=40$, $\phi=0$, and (f) $N=40$, $\phi=\pi/4$. The coloured scales of the $T(\boldsymbol{k}_{\parallel})$, shown on the right, refer to both plots with the same N.

are accompanied by a pronounced peak in the very centre of the 2D BZ, which is reflected by reduced TAMR values for very large insulator thicknesses N. This reduction is a simple consequence of vanishing of the Rashba term $H_{ss}^{SO}(\boldsymbol{k}_{\parallel})$ in the limit of $k_x \to 0$, $k_y \to 0$, see (6.3). The obtained changes in the $\boldsymbol{k}_{\parallel}$-resolved transmissions are in close analogy to the first-principles results, which corroborates the conclusions drawn in the previous section.

Conclusions

In the thesis' theoretical part we reviewed the DFT theory and the LMTO method of electronic structure calculation. We explained the background of a fully relastivistic theory and indicated a possible extension of the originally non-relativistic LMTO formalism to include relativistic effects, suitable mainly for dealing with light elements. Finally, we derived a conductance formula relevant for layered structures.

In the second, practical part of the thesis we applied the prepared formalism both to metallic systems and systems with a tunnelling barrier. Namely, we studied to what extent the SO interaction influences systems' conductance.

For metallic systems we showed that the influence of the SO interaction can be made convincingly large (comparison is always made with the scalar-relativistic results) when the systems are highly disordered (see Fig. 5.6).

Scalar-relativistic results for disordered metallic systems confirm some conclusions drawn by J. Mathon, (Math97), however our systems never enter the Anderson localization mode Mathon describes. Reasons why we do not observe this transition would deserve further investigation.

Finally, in the most disordered metallic system we observe a nonmonotonic AMR dependence when the magnetization is rotated from the out-of-plane (z) direction into the in-plane (x) direction, see the last pane of Fig. 5.5. The explanation for the minimum at about $\theta = 0.3\pi$ is yet unknown.

In the case of systems with a tunnelling barrier we have not found any serious impact of the SO interaction on the Fe/5Mgo/Ag system. On the other hand, for the Fe/GaAs/Ag system we demonstrated that interface resonances appearing on the two interfaces can hybridize and then strongly influence the TAMR. For intermediate thicknesses of the tunnel barrier this hybridization can lead to sizable TAMR values. Similar results were then obtained also in a simple tight-binding model of the FM/I/NM junction. The hybridized interface resonances can thus be added to the list of existing origins of the TAMR: the anisotropic density of states of the FM electrode (Goul04), the interference effects of the Rashba and Dresselhaus contributions to the SO interaction (Mose07, Mato09), and the interface states at the FM/I interface (Chan07a). This new mechanism—if realized experimentally in a special junction—might also be employed to enhance the TAMR effect for applications in spintronics.

Regarding the Fe/GaAs/Ag system, a more thorough study of the impact of possible imperfect atomic structure at the interfaces (we only used the crudest estimate by using a finite imaginary part of the energy arguments) would be desirable.

Further open problems related to all the presented results, such as, e.g., the effect of a finite bias, external magnetic fields or elevated temperatures, remain a task for future studies.

APPENDICES

A

Linear Response

A.1. General Relations

In the Schrödinger picture a system is fully described by a density matrix ρ whose time evolution is given by the Liouville equation

$$i\partial_t \rho = [H, \rho], \tag{A.1}$$

where H denotes the (possibly time-dependent if there are external varying fields) Hamiltonian of the system.

Let us suppose the Hamiltonian can be divided into two parts

$$H = H_0 + H'(t) = H_0 + QF(t), \qquad F(t) \in \mathrm{R}, \tag{A.2}$$

where H_0 describes the 'isolated' system, and H' represents some external perturbation on which we want to know the system's linear response. Concerning the perturbation, here we assume it is of a simple kind so as to be expressible as a product of some 'coupling' operator Q and a real function $F(t)$ which contains all external field time dependency.

We assume the system was free of perturbation in the distant past, $H'(t) = 0$ for $t = -\infty$, and further that it was then in the thermodynamic equilibrium,* i.e. with

$$\rho_0 \equiv \rho(-\infty) = \frac{1}{Z} e^{-\beta H_0}. \tag{A.3}$$

We want to study an expectation value p of some operator P, given by

$$p \equiv \langle P \rangle \equiv \mathrm{Tr}\, \rho P, \tag{A.4}$$

As long as the P operator is not explicitly time-dependent (we assume this as this is the usual case for an observable in the Schrödinger picture) and there is no perturbation, the expectation value is constant, since ρ stays constant. When the perturbation is switched on, ρ gains a nontrivial time evolution and the expectation value p can start to change:

$$\delta p(t) \equiv p(t) - p(-\infty) = \mathrm{Tr}\, (\rho(t) - \rho_0) P \equiv \mathrm{Tr}\, \delta\rho(t) P. \tag{A.5}$$

At this point switching to the interaction picture† (labelled with $\hat{}$) leads to a certain

* We just believe that thanks to some possibly weak interaction with the environment the system had been carried to equilibrium before we turn on the perturbation. After we turn the perturbation on we do not take account of this interaction any more supposing that the effect of the external perturbing field dominates.

† For any operator O we can at any time t define its counterpart $\hat{O}$ in this way:

$$\hat{O} \equiv e^{iH_0 t} O e^{-iH_0 t}.$$

If we follow this mapping in time, we may add the time labels: $\hat{O}(t) \equiv \hat{O(t)} = e^{iH_0 t} O(t) e^{-iH_0 t}$. Knowing the time evolution of one operator (i.e in one *picture*) defines the time evolution of the other:

simplification as the 'background' constant part of the Hamiltonian, H_0, drops out of the density operator time evolution:

$$i\partial_t \hat{\rho}(t) = [\hat{H'}(t), \hat{\rho}(t)], \tag{A.6}$$

integrating

$$\hat{\rho}(t) = \hat{\rho}(-\infty) - i\int_{-\infty}^{t} d\tau\, [\hat{H'}(\tau), \hat{\rho}(\tau)] = \rho_0 - i\int_{-\infty}^{t} d\tau\, [\hat{H'}(\tau), \hat{\rho}(\tau)], \tag{A.7}$$

$$\delta\hat{\rho}(t) = -i\int_{-\infty}^{t} d\tau\, [\hat{H'}(\tau), \hat{\rho}(\tau)]. \tag{A.8}$$

Since apparently Tr AB = Tr $\hat{A}\hat{B}$, we immediately have

$$\delta p(t) = \mathrm{Tr}\, \delta\hat{\rho}(t)\hat{P}(t) = -i\mathrm{Tr}\int_{-\infty}^{t} d\tau\, [\hat{H'}(\tau), \hat{\rho}(\tau)]\hat{P}(t) = -i\int_{-\infty}^{t} d\tau\, \mathrm{Tr}\, [\hat{P}(t), \hat{H'}(\tau)]\hat{\rho}(\tau) =$$

$$= i\int_{-\infty}^{t} d\tau\, F(\tau)\mathrm{Tr}\, [\hat{Q}(\tau), \hat{P}(t)]\hat{\rho}(\tau).$$

This is still an exact result. The *linear* response (i.e. to the first order in H', thus F) now amounts to estimating $\hat{\rho}(t)$ with ρ_0 in the last expression (see (A.7)):

$$\delta p^1(t) = i\int_{-\infty}^{t} d\tau\, F(\tau)\mathrm{Tr}\, [\hat{Q}(\tau), \hat{P}(t)]\rho_0 = i\int_{-\infty}^{t} d\tau\, F(\tau) <[\hat{Q}(\tau), \hat{P}(t)]>_0. \tag{A.9}$$

Now, from the definition of the interaction picture one can easily show that $<[\hat{Q}(\tau), \hat{P}(t)]>_0$ depends only on the time difference* $t-\tau$ and we can define the *susceptibility* $\chi_{Q\to P} \equiv \chi_{PQ}$:

$$\boxed{\delta p^1(t) = \int_{-\infty}^{\infty} d\tau\, \chi_{Q\to P}(t-\tau)\, F(\tau), \quad \chi_{Q\to P}(t) = -i <[\hat{P}(t), \hat{Q}(0)]>_0 \theta(t),} \tag{A.11}$$

where the subscript reminds us that we study the response in P as a result of perturbation coupled to Q. Thus $\chi_{Q\to P}(t) = -i <[\hat{P}(t), \hat{Q}(0)]>_0 \theta(t)$ equals the so-called *retarded Green function* of Q and P operators: $G^R(t)_{Q\to P} \equiv G^R(t)_{PQ}$.

Since (A.11) is in the form of convolution, we can also easily Fourier-transform†:

$$\delta p^1(\omega) = F(\omega)\chi(\omega). \tag{A.12}$$

Specifically, for

$$i\partial_t \hat{O}(t) = [\hat{O}(t), H_0] + e^{iH_0 t}\left[i\partial_t O(t)\right] e^{-iH_0 t}.$$

Specifically, for a time independent operator X in the former (Schrödinger) picture, we get

$$i\partial_t \hat{X}(t) = [\hat{X}(t), H_0] = [\hat{X}(t), \hat{H_0}(t)]$$

in the latter (interaction) picture. Note that always $\hat{H_0}(t) = H_0$. On the other hand, for the density operator ρ, whose Schrödinger picture form evolves according to the Liouville equation we started these notes with, one finally gets the dependency shown in the text:

$$i\partial_t \hat{\rho}(t) = [\hat{H'}(t), \hat{\rho}(t)].$$

* Quite generally, for operators A and B that are constant in the Schrödinger picture (i.e. the usual case)

$$<\hat{A}(t_1+\tau)\hat{B}(t_2)>_0 = \mathrm{Tr}\, \rho_0 \hat{A}(t_1+\tau)\hat{B}(t_2) = \mathrm{Tr}\, \rho_0 e^{iH_0\tau} e^{iH_0 t_1} A e^{-iH_0 t_1} e^{-iH_0\tau} e^{iH_0 t_2} B e^{-iH_0 t_2} = \tag{A.10}$$

$$= \mathrm{Tr}\, \rho_0 \hat{A}(t_1)\hat{B}(t_2-\tau) = <\hat{A}(t_1)\hat{B}(t_2-\tau)>_0.$$

The validity for commutators immediately follows, too.

† We use the convention: $f(\omega) = \int_{-\infty}^{\infty} dt\, f(t) e^{i\omega t}$.

$$F(\tau)=\delta(\tau) \quad \text{we get} \quad \delta p^1(t)=\chi(t), \tag{A.13}$$

thus the susceptibility can be viewed upon as a response to the 'unit pulse' at $t=0$. Similarly, for‡

$$F(\tau)=F_0 e^{-i\omega\tau} \quad \text{we get} \quad \delta p^1(t)=F_0 e^{-i\omega t}\chi(\omega). \tag{A.14}$$

It should be understood, that linearity of the response allows dividing the perturbing field into any number of components, calculating the corresponding responses separately, and finally adding them back together to obtain the complete response. Namely, we can consider individual Fourier frequencies independently.

The equation (A.11) represents one form of so-called *Kubo linear response formulæ*. With a use of the following identity it is possible to rewrite it into another form without the commutator.

For any operator A we may write

$$\partial_\lambda (e^{\lambda H_0} A e^{-\lambda H_0}) \;=\; e^{\lambda H_0}[H_0,A]e^{-\lambda H_0}. \tag{A.15}$$

Integrating $\int \mathrm{d}\lambda$ from 0 to β we get

$$e^{\beta H_0} A e^{-\beta H_0} - A \;=\; \int_0^\beta \mathrm{d}\lambda\; e^{\lambda H_0}[H_0,A]e^{-\lambda H_0} \;=\; \int_0^\beta \mathrm{d}\lambda\; [H_0,\hat{A}(-i\lambda)] \;=$$
$$= \int_0^\beta \mathrm{d}\lambda\; (-i\dot{\hat{A}}(-i\lambda)),$$

where the $\dot{\bullet}$ is just an abbreviation for $-i[\bullet,H_0]$—if the A operator can be thought of as a Schrödinger time-independent operator, then the $\dot{\hat{A}}$ has the same meaning as $\partial_t \hat{A}$. Multiplying from the left with $\rho_0=\frac{1}{Z}e^{-\beta H_0}$ we obtain

$$[A,\rho_0] \;=\; -i\rho_0 \int_0^\beta \mathrm{d}\lambda\; \dot{\hat{A}}(-i\lambda). \tag{A.16}$$

Using this identity in (A.11):

$$\chi_{Q\to P}(t) \;=\; i<[\hat{Q}(0),\hat{P}(t)]>_0\; \theta(t) \;=\; -i\mathrm{Tr}\,[Q,\rho_0]\hat{P}(t)\theta(t) \;=$$
$$= -\mathrm{Tr}\,\rho_0 \int_0^\beta \mathrm{d}\lambda\; \dot{\hat{Q}}(-i\lambda)\hat{P}(t)\theta(t) \;=\; -\int_0^\beta \mathrm{d}\lambda\; <\dot{\hat{Q}}(-i\lambda)\hat{P}(t)>_0\; \theta(t), \tag{A.17}$$

or similarly

$$\chi_{Q\to P}(t) \;=\; i<[\hat{Q}(-t),\hat{P}(0)]>_0\; \theta(t) \;=\; i\mathrm{Tr}\,[P,\rho_0]\hat{Q}(-t)\theta(t) \;=$$
$$= \mathrm{Tr}\,\rho_0 \int_0^\beta \mathrm{d}\lambda\; \dot{\hat{P}}(-i\lambda)\hat{Q}(-t)\theta(t) \;=\; \int_0^\beta \mathrm{d}\lambda\; <\dot{\hat{P}}(-i\lambda)\hat{Q}(-t)>_0\; \theta(t). \tag{A.18}$$

‡ Here we must be more cautious and clarify what we mean by a complex perturbation. Actually, any physical perturbation is real, and we really imagine that only the real part of the complex field is acting. However, since we consider *linear* response, i.e. $\mathrm{response}(A+iB)=\mathrm{response}(A)+i\,\mathrm{response}(B)$, we can safely calculate the response on a complex perturbation and take the real part of the result in the end. Note that a response on a real field is necessarily real.

Summarizing, with added $\hbar$:

$$\boxed{\begin{aligned} \delta p^1(t) &= \int_{-\infty}^{\infty} d\tau\, F(\tau)\, \chi_{Q\to P}(t-\tau) \\ \chi_{Q\to P}(t) &= \frac{i}{\hbar} <[\hat{Q}(0),\hat{P}(t)]>_0 \theta(t) = \\ &= -\int_0^{\beta} d\lambda\, <\dot{\hat{Q}}(-i\hbar\lambda)\hat{P}(t)>_0 \theta(t) = \int_0^{\beta} d\lambda\, <\dot{\hat{P}}(-i\hbar\lambda)\hat{Q}(-t)>_0 \theta(t) \end{aligned}} \tag{A.19}$$

The forms without the commutator are more suitable e.g. for taking the classical limit... Note that using the identity (A.10) we may 'shuffle' the operators' arguments to and fro at will, thus getting for instance

$$\chi_{Q\to P}(t) = -\int_0^{\beta} d\lambda\, <\dot{\hat{Q}}(0)\hat{P}(t+i\hbar\lambda)>_0 \theta(t). \tag{A.20}$$

A.2. Application to Electrical Conductivity

For a single charged particle in an electric field the Hamiltonian yielding correct equation of motion is

$$H = \frac{\boldsymbol{p}^2}{2m} - q\boldsymbol{r}\cdot\boldsymbol{E}(t). \tag{A.21}$$

Thus for many particles labelled with i (and summation convention applied)

$$H'(t) = E_\mu(t)\ (-q_i r_{i\mu}). \tag{A.22}$$

The current linear response $j_\nu{}^1(t)$ corresponding to the current operator* $J_\nu = q_i \dot{r}_{i\nu}$ on the electric field E_μ is then

$$j_\nu{}^1(t) = \int_{-\infty}^{\infty} d\tau\, \chi_{\nu\mu}(t-\tau)E_\mu(\tau) \quad \text{or} \quad j_\nu{}^1(\omega) = \chi_{\nu\mu}(\omega)E_\mu(\omega), \tag{A.23}$$

where

$$\begin{aligned} \chi_{\nu\mu}(t) \equiv \chi_{\mu\to\nu}(t) &= i <[-q_i \hat{r}_{i\mu}(0)\ ,\ q_i \dot{\hat{r}}_{i\nu}(t)]>_0 \theta(t) = \\ &= \int_0^{\beta} d\lambda\, <q_i \dot{\hat{r}}_{i\mu}(-i\lambda)\ q_j \dot{\hat{r}}_{j\nu}(t)>_0 \theta(t) \overset{\dagger}{=} \int_0^{\beta} d\lambda\, <\hat{J}_\mu(-i\lambda)\ \hat{J}_\nu(t)>_0 \theta(t) \\ &= \int_0^{\beta} d\lambda\, <J_\mu\ \hat{J}_\nu(t+i\lambda)>_0 \theta(t). \end{aligned} \tag{A.24}$$

---

* For a time-independent Schrödinger operator O, by $\dot{O}$ we mean

$$\dot{O} \equiv \frac{1}{i}[O,H].$$

It is defined so that (using the Schrödinger equation for a state $\psi(t)$)

$$\partial_t <\psi(t)|\, O\, |\psi(t)> = <\psi(t)|\, \frac{1}{i}(OH-HO)\, |\psi(t)> = <\psi(t)|\, \frac{1}{i}[O,H]\, |\psi(t)> = <\psi(t)|\, \dot{O}\, |\psi(t)>.$$

Note the presence of the *full* Hamiltonian, which means $\dot{O}$ is actually a time-dependent operator (since H is time-dependent).

† In relation to the previous footnote we should be careful here and note that $\dot{\hat{O}}(t)$ and $\hat{\dot{O}}(t)$ are generally different things:

$$\dot{\hat{O}}(t) = -i[e^{iH_0 t}Oe^{-iH_0 t}, H_0] = -i[O,\hat{H}_0](t),$$

Finally, if we work with current densities instead, i.e. $\underline{j}_\nu{}^1 \equiv \frac{j_\nu{}^1}{\Omega}$, $\underline{J}_\nu \equiv \frac{J_\nu}{\Omega}$ and $\underline{\chi}_{\nu\mu} \equiv \frac{\chi_{\nu\mu}}{\Omega}$, where Ω is some elementary volume, we get:

$$\underline{j}_\nu{}^1(\omega) = \underline{\chi}_{\nu\mu}(\omega) E_\mu, \tag{A.25}$$

with

$$\underline{\chi}_{\nu\mu}(\omega) = \Omega \int_0^\infty \mathrm{d}t \int_0^\beta \mathrm{d}\lambda \, \langle \underline{J}_\mu \, \hat{\underline{J}}_\nu(t+i\lambda) \rangle_0 e^{i\omega t}. \tag{A.26}$$

A.3. Proof that Many-Body Linear Response is Identical to Single-Particle One in the Case of Non-Interacting Particles

$$\chi_{\nu\mu}(t) = -i \langle [\Pi_\mu(0), J_\nu(t)] \rangle_0 \theta(t) \tag{A.27}$$

$$\langle [\Pi_\mu(0), J_\nu(t)] \rangle_0 = \mathrm{Tr} \left[\rho_0 [\Pi_\mu(0), J_\nu(t)] \right]$$

$$= \mathrm{Tr} \left[\rho_0 [a_k^+ a_l, a_m^+ a_n] \right] (\pi_\mu)_{kl} (j_\nu)_{mn} e^{i\varepsilon_{mn} t}, \tag{A.28}$$

where $\varepsilon_{mn} = \varepsilon_m - \varepsilon_n$, and we used $e^{iH_0 t} a_k^+ e^{-iH_0 t} = a_k^+ e^{i\varepsilon_k t}$ and $e^{iH_0 t} a_l e^{-iH_0 t} = a_l e^{-i\varepsilon_k t}$ in

$$J_\nu(t) = e^{iH_0 t} J_\nu(0) e^{-iH_0 t} = e^{iH_0 t} a_k^+ e^{-iH_0 t} e^{iH_0 t} a_l e^{-iH_0 t} (j_\nu)_{kl} = a_k^+ a_l e^{i\varepsilon_{kl} t} (j_\nu)_{kl}. \tag{A.29}$$

We simplify the trace

$$\mathrm{Tr}\, \rho_0 [a_k^+ a_l, a_m^+ a_n] = \mathrm{Tr}\, \rho_0 \left[a_k^+ [a_l, a_m^+ a_n] + [a_k^+, a_m^+ a_n] a_l \right] =$$

$$= \mathrm{Tr}\, \rho_0 (a_k^+ a_n \delta_{lm} - a_m^+ a_l \delta_{kn}) \tag{A.30}$$

(valid for both fermions and bosons) and utilize that particles do not interact, i.e.

$$H_0 = \varepsilon_k a_k^+ a_k \quad \Rightarrow \quad \rho_0 = \frac{1}{Z} e^{-\beta \varepsilon_k a_k^+ a_k} \quad \Rightarrow \quad \mathrm{Tr}\, \rho_0 a_k^+ a_l = f_k \delta_{kl}, \tag{A.31}$$

leading to

$$\mathrm{Tr}\, \rho_0 [a_k^+ a_l, a_m^+ a_n] \overset{\text{non-int.}}{=} (f_k - f_m) \delta_{ml} \delta_{kn}. \tag{A.32}$$

Hence

$$\chi_{\nu\mu}(t) = i\theta(t)(f_k - f_l)(\pi_\mu)_{kl} (j_\nu)_{lk} e^{-i\varepsilon_{kl} t} = -i\theta(t)(f_k - f_l) \langle k | \pi_\mu(0) | l \rangle \langle l | j_\nu(t) | k \rangle =$$

$$= -i\theta(t) \langle [\pi_\mu(0), j_\nu(t)] \rangle_0^{(1 \text{ particle})}. \tag{A.33}$$

---

whilst

$$\dot{\hat{O}}(t) = -i[\hat{O}, \hat{H}](t) = -i[O, \hat{H}_0](t) - i[O, \hat{H}'](t) = \dot{\hat{O}}(t) - i[O, \hat{H}'](t).$$

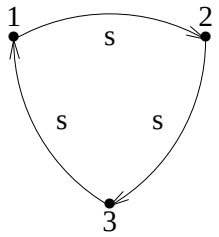
For the use in the linear response, however, the last term, being proportional to positive powers of the perturbing field $F(t)$ (i.e. here $E_\mu(t)$), is to be neglected.

B

Note on the Bloch theorem

Many textbooks state that a stationary state of an electron in a periodic potential must be of the Bloch form. Here we show on a simple explicit example that it is really not so.

Consider 3 equivalent sites, $j = 1, 2, 3$, with one orbital on each site and with the same hopping s between any two of them. Set the on-site energy to zero. The situation and governing (Schrödinger) equations are



$$\begin{aligned} s(\psi_1 + \psi_3) &= E\psi_2, \\ s(\psi_2 + \psi_1) &= E\psi_3, \\ s(\psi_3 + \psi_2) &= E\psi_1. \end{aligned} \tag{B.1}$$

The wavefunction† $\psi_j = e^{\pm(\mathrm{i}jk)}$ solves these equations with the energy $E = 2s\cos k$, if $e^{3\mathrm{i}k} = 1$. Thus k can be 0, $2/3\pi$, $4/3\pi$. (Such wavefunctions are of Bloch form.)

Now consider $\phi_j = e^{\mathrm{i}jk} + e^{-\mathrm{i}jk} = 2\cos kj$, specifically for $k = 2/3\pi$, i.e. $E = -s$: $\phi_1 = -1$, $\phi_2 = -1$, $\phi_3 = 2$. We see ϕ solves our equations (it is a combination of Bloch solutions pertaining to a single energy level) but *it is not of the Bloch form.* Note especially, that the probability of finding an electron on different sites is different (1/6, 1/6, 2/3).

Bloch's theorem does not state that any (stationary) solution must be of the Bloch form. It only claims that we can *choose* the basis vectors to be in this form (like ψ above). Only if an energy level is not degenerate then its wavefunction must be of the Bloch form.

† Normalization is of no importance to us now. The normalized-to-1 wavefunction would be $(1/3)^{½}\psi$; similarly later $(1/6)^{½}\phi$.

C

Dirac (spin-½) Particle in a Spherical Potential

Stationary, i.e. with sharp energy E, (4-component) states ψ of a Dirac particle of mass m in a spherically symmetric spin-independent potential $V(r)$ are solutions of the corresponding (time-independent) Dirac equation

$$H\psi = [c\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta mc^2 + V(r)]\psi = E\psi. \tag{C.1}$$

Here $\boldsymbol{\alpha}$, β are generally any 4×4 Hermitian¶ matrices for which any two satisfy $\{\bullet,\times\} = 2\delta_{\bullet\times}$ (the Dirac [or Clifford] algebra‡). Pauli's fundamental theorem§ shows that all realizations of the given algebra by 4×4 matrices are unitarily equivalent. Henceforth the so-called Dirac representation will be used, which chooses

$$\boldsymbol{\alpha} = \begin{bmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{bmatrix}, \qquad \beta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{C.2}$$

where $\boldsymbol{\sigma}$ denotes the Pauli matrices and ± 1 in the β matrix are 2×2 diagonal matrices.** Also, operators and their x-representation will be freely interchanged, hopefully not causing much confusion.

The above Hamiltonian—thanks to its spherical symmetry—commutes with the total-angular momentum operator $\boldsymbol{J} = \boldsymbol{L} + \tfrac{1}{2}\boldsymbol{\sigma}$, $\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{p}$ being the orbital angular momentum, as well as with the space-inversion operator $P = \eta_P \beta.\dagger,\|,*$ Since H, $\boldsymbol{J}^2$, J_z, P mutually commute, they have common

---

¶ Hermiticity is necessary for the Hamiltonian to be Hermitian.

‡ The algebra guarantees that a solution of the Dirac equation also solves the Klein-Gordon equation, $(\boldsymbol{p}^2 + m^2)\psi = (E - V)^2\psi$, a necessity imposed by relativity.

§ See e.g. (Saku67, appendix C).

** The Dirac representation is advantageous when considering the non-relativistic limit. In this representation the magnitudes of bispinor's upper and lower components ψ_A, ψ_B are in ratio of the order of v/c (upper being large for a positive-energy solution and vice-versa for a negative-energy one). However, the standard representation slightly conceals what is evident from the so-called Weyl representation, in which the Lorentz group generators for the bispinor representation (their 2-multiple given by $\Sigma^{\mu\nu} = i/2[\gamma^\mu,\gamma^\nu]$) are block-2×2 diagonal and thus showing that the bispinor representation is reducible: the so-called left- and right-hand spinors ψ_L, ψ_R transform without mixing. The relation between Weyl and Dirac bispinors is $\psi_{L,R} = 1/\sqrt{2}\,(\psi_A \pm \psi_B)$, $\psi_{A,B} = 1/\sqrt{2}\,(\psi_L \pm \psi_R)$.

Yet another fact resulting from choosing a concrete set of the γ-matrices is worth realizing. The choice determines the matrix representation of the spin operator $S_i = \tfrac{1}{2}\Sigma_i = \tfrac{1}{2}\,(\tfrac{1}{2}\varepsilon_{ijk}\Sigma^{jk})$, in the standard (but by 'chance' also in the Weyl) representation giving $\boldsymbol{S} = \tfrac{1}{2}\begin{bmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{bmatrix}$. Thus picking up a set of γ-matrices in fact implicitly means also choosing the bispinor basis. In the standard representation the S_z operator is diagonal.

Finally, combining the information just given, we might actually choose another 'standard-like' representation related to the standard one by a suitable unitary block-diagonal matrix, $\gamma^\mu \to \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}\gamma^\mu\begin{bmatrix} U^+ & 0 \\ 0 & U^+ \end{bmatrix}$, $UU^+ = 1$, which would preserve the division into large and small components, but would change the bispinor basis so that the $\boldsymbol{n}\cdot\boldsymbol{S}$ operator would, instead of S_z, be diagonal for any given unit (3D) vector $\boldsymbol{n}$. Such a unitary transformation (function of $\boldsymbol{n}$) is apparently a special case of the unitary transformation freedom in choosing of the γ-matrices.

† For a *free* particle, the commutation of the Hamiltonian with the total-angular-momentum operator (generator of

eigenstates. As shown in (Form00, ch. 3.1.3), common eigenstates of $\boldsymbol{J}^2$, J_z, P have to be of the form

$$\psi^l_{j\mu}(\boldsymbol{r}) \;=\; \begin{bmatrix} R_1(r)\,\Omega^l_{j\mu}(\boldsymbol{n}) \\ R_2(r)(\boldsymbol{\sigma}\cdot\boldsymbol{n})\,\Omega^l_{j\mu}(\boldsymbol{n}) \end{bmatrix}, \quad \text{where} \quad \begin{array}{l} j = 1/2,\ 3/2,\ 5/2,\ 7/2,\ \ldots \\ l \text{ is restricted by } j = l \pm \tfrac{1}{2} \\ \mu = -j, -j+1, \ldots, j-1, j \end{array} \quad ; \quad \boldsymbol{n} = \frac{\boldsymbol{r}}{r}, \tag{C.3}$$

with the corresponding eigenvalues $j(j+1)$, μ and $\eta_P(-)^l$, respectively. Appearing $\Omega^l_{j\mu}$ (generally with the same allowed values of j, l and μ as shown above) are the so-called spherical spinors, i.e. (2-component) eigenstates of $\boldsymbol{J}^2$, $\boldsymbol{L}^2$, ($\boldsymbol{S}^2 = \tfrac{3}{4}$), J_z with eigenvalues $j(j+1)$, $l(l+1)$, ($\tfrac{3}{4}$), μ, respectively. Finally R_1, R_2 are arbitrary radial functions. Thanks to the restriction on l, we may introduce a new quantum number κ and simplify our notation:

$$(jl) \to \begin{cases} \kappa = -(j+\tfrac{1}{2}) \ [\text{i.e.} < 0] & \text{if} \quad j = l + \tfrac{1}{2} \\ \kappa = +(j+\tfrac{1}{2}) \ [\text{i.e.} > 0] & \text{if} \quad j = l - \tfrac{1}{2} \end{cases} \quad \& \quad \kappa \to j = |\kappa| - \tfrac{1}{2}, \quad \begin{cases} l = -\kappa - 1 & \text{if } \kappa < 0 \\ l = \kappa & \text{if } \kappa > 0 \end{cases}, \tag{C.4}$$

so (jl) and κ carry the same information and we may further use only κ. Note that κ can have all integer values except zero. The bispinor (C.3) then reads

$$\psi_{\kappa\mu(}\boldsymbol{r}) \;=\; \begin{bmatrix} R_1(r)\,\Omega_{\kappa\mu}(\boldsymbol{n}) \\ R_2(r)(\boldsymbol{\sigma}\cdot\boldsymbol{n})\,\Omega_{\kappa\mu}(\boldsymbol{n}) \end{bmatrix}. \tag{C.5}$$

Requesting (C.5) to solve the Dirac equation (C.1) we get

$$[c(\boldsymbol{\sigma}\cdot\boldsymbol{p})(\boldsymbol{\sigma}\cdot\boldsymbol{n})R_2 + (mc^2 + V - E)R_1]\,\Omega_{\kappa\mu} \;=\; 0, \tag{C.6a}$$

$$[c(\boldsymbol{\sigma}\cdot\boldsymbol{p})R_1 + (-mc^2 + V - E)R_2(\boldsymbol{\sigma}\cdot\boldsymbol{n})]\,\Omega_{\kappa\mu} \;=\; 0. \tag{C.6b}$$

Using the identity†

$$(\boldsymbol{\sigma}\cdot\boldsymbol{a})(\boldsymbol{\sigma}\cdot\boldsymbol{b}) \;=\; \boldsymbol{a}\cdot\boldsymbol{b} + i(\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{\sigma}, \tag{C.7}$$

which is valid for any operators $\boldsymbol{a}$, $\boldsymbol{b}$ which commute with all σ matrices, one may show that

$$(\boldsymbol{\sigma}\cdot\boldsymbol{n})(\boldsymbol{\sigma}\cdot\boldsymbol{p}) \;=\; \frac{i}{r}[(\boldsymbol{\sigma}\cdot\boldsymbol{L}) - r\partial_r], \tag{C.8}$$

$$(\boldsymbol{\sigma}\cdot\boldsymbol{p})(\boldsymbol{\sigma}\cdot\boldsymbol{n}) \;=\; -\frac{i}{r}[(\boldsymbol{\sigma}\cdot\boldsymbol{L}) + r\partial_r + 2]. \tag{C.9}$$

Slightly more complicated is

$$(\boldsymbol{\sigma}\cdot\boldsymbol{p})\,f(r)\,\Omega_{\kappa\mu} \;=\; (\boldsymbol{\sigma}\cdot\boldsymbol{n})(\boldsymbol{\sigma}\cdot\boldsymbol{n})(\boldsymbol{\sigma}\cdot\boldsymbol{p})\,f(r)\,\Omega_{\kappa\mu} \;=\; (\boldsymbol{\sigma}\cdot\boldsymbol{n})\,\frac{i}{r}[(\boldsymbol{\sigma}\cdot\boldsymbol{L}) - r\partial_r]\,f(r)\,\Omega_{\kappa\mu}$$

---

rotations), $[H_0, \boldsymbol{J}] = 0$, as well as with the space-inversion operator, $[H, P] = 0$, is required by symmetry as a part of the fundamental *request* that the Dirac equation be covariant with respect to any Lorentz transformation (rotations, boosts, space and time inversions). This is a rather strong demand on a theory that needs to be relaxed a bit e.g. in the case of weak interactions, where neither space- nor time-inversion is a symmetry. However, there is a deeper reason why full Lorentz covariance should hold in electromagnetism and strong interactions; see (Wein95, ch. 12.5). Finally, note that adding a symmetric potential $V(r)$ evidently does not spoil the commutation.

‖ First, in our notation any operator that seems to not have enough components is implicitly meant to be promoted with unit matrices in the missing sectors; e.g. in the present case $\boldsymbol{J} = \begin{bmatrix} \boldsymbol{L}1_{2\times2} + \frac{1}{2}\boldsymbol{\sigma} & 0_{2\times2} \\ 0_{2\times2} & \boldsymbol{L}1_{2\times2} + \frac{1}{2}\boldsymbol{\sigma} \end{bmatrix}$. Second, it should be proved, despite the appealing form of the presented $\boldsymbol{J}$, that this really is the right total-angular-momentum operator (in the Dirac representation), i.e. that $\psi'(x) = \exp(i\Omega\boldsymbol{n}\cdot\boldsymbol{J})\,\psi(x)$ is the correspondence between wave functions of the same reality expressed in a coordinate system and a system rotated with respect to the former by an angle Ω around a direction $\boldsymbol{n}$. For that see e.g. (Form00, ch. 3.1.3).

* Unlike in the non-relativistic case, where $P\psi_{\text{non-rel}}(x) = \eta_P\,\psi_{\text{non-rel}}(-x)$, $\eta_P \in \mathrm{C}$, the space-inversion for a Dirac particle embodies the β matrix, $P\psi(x) = \eta_P\,\beta\psi(-x)$. ($\eta_P$ is the so-called intrinsic parity of a particle, $|\eta_P| = 1$.) The reason for this extra complexity lies in the requested Lorentz covariance of the free Dirac equation. See e.g. (Form00, ch. 3.1.1).

† Follows from $(\boldsymbol{\sigma}\cdot\boldsymbol{a})(\boldsymbol{\sigma}\cdot\boldsymbol{b}) = \sigma_i a_i \sigma_j b_j = a_i b_j \sigma_i \sigma_j$, and $\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k$. Many more identities can be found in (Szmy07).

$$= -\frac{i}{r}\left[(1+\kappa)f + rf'\right](\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu} \qquad \text{for any } f(r), \tag{C.10}$$

where we used (C.8) and*

$$(\boldsymbol{\sigma}\cdot\boldsymbol{L})\Omega_{\kappa\mu} = -(1+\kappa)\Omega_{\kappa\mu}. \tag{C.11}$$

Thus (C.6) can be recast into§

$$ic\left[R_2' + (1-\kappa)\frac{1}{r}R_2\right] = (V - E + mc^2)R_1, \tag{C.12a}$$

$$ic\left[R_1' + (1+\kappa)\frac{1}{r}R_1\right] = (V - E - mc^2)R_2, \tag{C.12b}$$

which can be further simplified using $P = rR_1$, $Q = icrR_2$,

$$Q' = \frac{\kappa}{r}Q - T(r)P, \tag{C.13a}$$

$$P' = -\frac{\kappa}{r}P + \omega(r)Q, \tag{C.13b}$$

where $\omega(r) = 1 + \dfrac{\varepsilon - V(r)}{2mc^2} = 1 + \dfrac{T(r)}{2mc^2}$ is the so-called mass-enhancement factor‡, $\varepsilon = E - mc^2$ is the non-relativistic energy and $T(r) = \varepsilon - V(r)$ is the kinetic energy.

Equations (C.13) represent conditions imposed on the radial functions under which the bispinor (C.3), (C.5) solves the time-independent Dirac equation (C.1) with energy E. For a general form of $V(r)$ (the case of DFT calculations) the equations must be solved numerically.

PS: The form of the bispinor (C.5)

We note that the bispinor (C.5) can also be written as

$$\psi_{\kappa\mu}(\boldsymbol{r}) = \begin{bmatrix} R_1(r)\Omega_{\kappa\mu}(\boldsymbol{n}) \\ -R_2(r)\Omega_{-\kappa\mu}(\boldsymbol{n}) \end{bmatrix} = \frac{1}{r}\begin{bmatrix} P(r)\Omega_{\kappa\mu}(\boldsymbol{n}) \\ \dfrac{iQ(r)}{c}\Omega_{-\kappa\mu}(\boldsymbol{n}) \end{bmatrix}. \tag{C.14}$$

The proof is based on the equality

$$(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu} = -\Omega_{-\kappa\mu}, \tag{C.15}$$

which can be proved as follows. First one realizes that $(\boldsymbol{\sigma}\cdot\boldsymbol{n})$ commutes with $\boldsymbol{J}$, $[(\boldsymbol{\sigma}\cdot\boldsymbol{n}), \boldsymbol{J}] = 0$. This follows from the fact that $\boldsymbol{\sigma}$ are matrix elements of a (pseudo-)vector operator and $\boldsymbol{n}$ behaves like a vector, too.† Their product then must be a scalar quantity, and this is equivalent to the zero commutator with $\boldsymbol{J}$. Thus $(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu}$ is an eigenstate of $\boldsymbol{J}^2$ and J_z with the same eigenvalues as the $\Omega_{\kappa\mu}$ alone. But then, since $\Omega_{\kappa\mu}$ and $\Omega_{-\kappa\mu}$ form a basis of the $(j = |\kappa| - ½, \mu)$ subspace,

$$(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu} = c_1(\kappa\mu)\Omega_{\kappa\mu} + c_2(\kappa\mu)\Omega_{-\kappa\mu}, \tag{C.16}$$

where c_1 and c_2 are some numbers, possibly dependent on κ and μ. Since $(\boldsymbol{\sigma}\cdot\boldsymbol{n})$ is a Hermitian matrix and $(\boldsymbol{\sigma}\cdot\boldsymbol{n})^2 = 1$, see (C.7), it is also unitary, $((\boldsymbol{\sigma}\cdot\boldsymbol{n})x, (\boldsymbol{\sigma}\cdot\boldsymbol{n})y) = (x, (\boldsymbol{\sigma}\cdot\boldsymbol{n})^2 y) = (x, y)$. The norm of the LHS is thus 1 and so must be of the RHS, giving $|c_1|^2 + |c_2|^2 = 1$ ($\Omega_{\kappa\mu}$ and $\Omega_{-\kappa\mu}$ are

---

* This follows from expressing $\boldsymbol{J}^2 = (\boldsymbol{L} + ½\boldsymbol{\sigma})^2 \Rightarrow \boldsymbol{\sigma}\cdot L = \boldsymbol{J}^2 - \boldsymbol{L}^2 - ¾$ and writing the eigenvalues in terms of κ.

§ In the process, we get rid of the trailing $\Omega_{\kappa\mu}$ or $(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu}$. This is fine: we have equations like $f(r)(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu}(\boldsymbol{n}) = 0 \;\forall r \forall \boldsymbol{n}$. Thus we only need one $\boldsymbol{n}$ for which $(\boldsymbol{\sigma}\cdot\boldsymbol{n})\Omega_{\kappa\mu}(\boldsymbol{n}) \neq 0$ to enforce that $f(r)$ must be identically zero. There surely are many.

‡ Note that ω differs from the relativistic γ-factor, $\gamma = (1 - v^2/c^2)^{-1/2}$. They are, however, related: $\omega = ½(\gamma + 1)$. The name of ω originates from the fact that $T = \dfrac{\boldsymbol{p}^2}{2m\omega}$.

† There are some notes about vector operators in PS2.

orthonormal). Further, the two terms on the RHS have opposite parities and only the second complies with the LHS, thus $c_1=0$, and

$$(\boldsymbol{\sigma}\cdot\boldsymbol{n})\,\Omega_{\kappa\mu} \;=\; c_2(\kappa)\,\Omega_{-\kappa\mu}, \qquad |c_2| \;=\; 1, \tag{C.17}$$

where, further, the μ-dependence of c_2 was ruled out using the Wigner-Eckart theorem.¶ As far as I know, the phase c_2 cannot be deduced only from say commutation laws, but the very definition of the spherical spinors must be utilized, taking account of various used (e.g. Condon-Shortley) phase conventions. Evaluating the definition for a particularly easy $\boldsymbol{n}=(0,0,1)$ (i.e. $\theta=0$)—knowing that the coefficient is $\boldsymbol{n}$-independent—finally gives us the wanted coefficient $c_2=-1$:

$$\Omega_{\kappa\mu}(\boldsymbol{n}) \;=\; \begin{bmatrix} \mathrm{sgn}(-\kappa)\left[\dfrac{\kappa+\frac{1}{2}-\mu}{2\kappa+1}\right]^{1/2} Y_{l,\mu-\frac{1}{2}}(\boldsymbol{n}) \\[2ex] \left[\dfrac{\kappa+\frac{1}{2}+\mu}{2\kappa+1}\right]^{1/2} Y_{l,\mu+\frac{1}{2}}(\boldsymbol{n}) \end{bmatrix} \;\to\; \Omega_{\kappa\mu}(\theta=0) \;=\; \left[\frac{|\kappa|}{4\pi}\right]^{1/2} \begin{bmatrix} \mathrm{sgn}(-\kappa)\,\delta_{\mu,\frac{1}{2}} \\ \delta_{\mu,-\frac{1}{2}} \end{bmatrix}, \tag{C.18}$$

where we used $Y_{lm}(\theta=0) \;=\; \left[\dfrac{2l+1}{4\pi}\right]^{1/2}\delta_{m0}$. Thus

$$\sigma_z\,\Omega_{\kappa\mu}(\theta=0) \;=\; \left[\frac{|\kappa|}{4\pi}\right]^{1/2} \begin{bmatrix} \mathrm{sgn}(-\kappa)\,\delta_{\mu,\frac{1}{2}} \\ -\delta_{\mu,-\frac{1}{2}} \end{bmatrix} \;=\; -\left[\frac{|\kappa|}{4\pi}\right]^{1/2} \begin{bmatrix} \mathrm{sgn}(\kappa)\,\delta_{\mu,\frac{1}{2}} \\ \delta_{\mu,-\frac{1}{2}} \end{bmatrix} \;=\; -\Omega_{-\kappa\mu}(\theta=0). \tag{C.19}$$

¶ $\Omega_{\kappa\mu}(\boldsymbol{n})$ are (orthogonal) eigenvectors of $\boldsymbol{J}^2$ and J_z. Since $(\boldsymbol{\sigma}\cdot\boldsymbol{n})$ is a scalar operator, see PS2, the Wigner-Eckart theorem states that $(\Omega_{\kappa\mu},(\boldsymbol{\sigma}\cdot\boldsymbol{n})\,\Omega_{\kappa'\mu'})=c(j)\,\delta_{jj'}\,\delta_{\mu\mu'}$, where the relation between js and κs is given by (C.3) and c is a number only dependent on $j(\kappa)$ (reduced matrix element). This implies (C.16), but with the μ-dependence of c_1, c_2 ruled out.

D

Spherical harmonics

Spherical harmonics $Y_{lm} \equiv Y_L$, with $l = 0, 1, 2, \ldots, \infty$ and $m = -l, \ldots, l$, are functions defined on a unit sphere

$$Y_{lm}(\theta,\phi) \;=\; Y_L(\hat{\boldsymbol{r}}) \;=\; \left[\frac{2l+1}{4\pi}\,\frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m[\cos\theta]\,e^{im\phi} \;=\; N_{lm}P_l^m[\cos\theta]\,e^{im\phi}, \quad \text{(D.1)}$$

where $\hat{\boldsymbol{r}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, and P_l^m are the associated Legendre polynomials

$$P_l^m(x) \;=\; (-)^m(1-x^2)^{m/2}\,\frac{\mathrm{d}^m P_l(x)}{\mathrm{d}x^m}, \quad m \;=\; 0, 1, \ldots, l, \quad \text{(D.2)}$$

defined in terms of the ordinary Legendre polynomials

$$P_l(x) \;=\; \frac{1}{2^l l!}\,\frac{\mathrm{d}^l (x^2-1)^l}{\mathrm{d}x^l}; \quad \text{(D.3)}$$

insertion of (D.3) into (D.2) allows defining P_l^m also for negative ms, as needed in (D.1):

$$P_l^m(x) \;=\; \frac{(-)^m}{2^l l!}(1-x^2)^{m/2}\,\frac{\mathrm{d}^{l+m}(x^2-1)^l}{\mathrm{d}x^{l+m}}, \quad m \;=\; -l, \ldots, l. \quad \text{(D.4)}$$

[One can show $P_l^{-m}(x) = (-)^m\,\dfrac{(l-m)!}{(l+m)!}\,P_l^m(x)$; $N_{lm}P_l^m = (-)^m N_{l-m}P_l^{-m}$.]

The spherical harmonics are orthonormal on a unit sphere, S_1,

$$\int \mathrm{d}\hat{\boldsymbol{r}}\; Y_L^*(\hat{\boldsymbol{r}})\,Y_{L'}(\hat{\boldsymbol{r}}) \;=\; \delta_{LL'}, \quad \text{(D.5)}$$

and form a (countable) basis of $L^2(S_1)$ functions.

The spherical harmonics (D.1) are complex-valued. Often it is more practical to use their real-valued linear combinations, here denoted S_L, normalized in the same way as Ys are. Each such combination necessarily involves Y_{lm} and $Y_{l-m} = (-)^m Y_{lm}^*$ (except for the $m = 0$ case, which is already real). One may, e.g., have $S_{m>0} \propto \cos(|m|\phi)$ and $S_{m<0} \propto \sin(|m|\phi)$:

$$S_{lm} = \begin{cases} \dfrac{1}{\sqrt{2}}\left(Y_{l|m|} + Y_{l|m|}^*\right) \;=\; \sqrt{2}\,N_{l|m|}P_l^{|m|}\cos(|m|\phi) & \text{if } m > 0, \\ Y_{l0} \;=\; N_{l0}P_{l0} & \text{if } m = 0, \\ \dfrac{1}{i\sqrt{2}}\left(Y_{l|m|} - Y_{l|m|}^*\right) \;=\; \sqrt{2}\,N_{l|m|}P_l^{|m|}\sin(|m|\phi) & \text{if } m < 0. \end{cases} \quad \text{(D.6)}$$

These functions are orthonormal, $\int \mathrm{d}\hat{\boldsymbol{r}}\; S_L S_{L'} = \delta_{LL'}$, and form a basis of $L^2(S_1)$, too.

E

Note on Matrix Signatures

Let us define a (generally complex) matrix M to be *non-negative*, $M \geq 0$, if for all *complex* vectors x $(x, Mx) \in \mathrm{R}$ and ≥ 0. [Similarly define a *non-positive* matrix, $M \leq 0$.] Such a matrix is then necessarily hermitian: The matrix can always be decomposed to its hermitian and antihermitian part, $M = \mathrm{Re}M + i\mathrm{Im}M$, $\mathrm{Re}M = \frac{1}{2}(M + M^+)$, $\mathrm{Im}M = \frac{1}{2i}(M - M^+)$. Both $\mathrm{Re}M$ and $\mathrm{Im}M$ are now hermitian. In order that (x, Mx) be real $(x, \mathrm{Im}Mx)$ must be zero. But for any matrix N $(x, Nx) = 0$ for all x implies $N = 0$.* Thus $\mathrm{Im}M = 0$ and M is hermitian.

If M satisfies $M \geq 0$ or $M \leq 0$, we call M *definite*. If for M neither $M \geq 0$ nor $M \leq 0$ holds, we call M *indefinite*.

1 $\boxed{\mathrm{Im}G^R(E) \leq 0}$

Follows from

$$G^R = \frac{1}{E - H + i0} = \mathrm{P}\frac{1}{E - H} - i\pi\delta(E - H),$$

since

$$-\frac{1}{\pi}\mathrm{Im}G^R = \delta(E - H) = A(E) = A^+(E) \geq 0.$$

One should think of such operators as of operator densities, so that the very operators (well behaved) are then e.g. like $O(E) = \int_{\Delta E} A(E)\, dE$.

2 $\boxed{M \geq 0 \rightarrow AMA^+ \geq 0 \qquad M \text{ hermitian, } A \text{ arbitrary}}$ i.e. property of non-negativity is basis-independent

Follows from

$$(x, AMA^+ x) = (y, My) \geq 0, \text{ where } y = A^+ x.$$

---

* Here we need less since $\mathrm{Im}M$ is hermitian and as for such this property can be seen from the fact that any hermitian matrix can be diagonalized and must then have zero eigenvalues). Anyway, for a general matrix:

Either by a brute force: $\sum_{ij} x_i^* N_{ij} x_j = 0$. First, choose $x_i = \delta_{ik}$ giving $N_{kk} = 0$. Second, choose $x_i = \delta_{ik} + \delta_{il}$, giving $N_{kl} + N_{lk} = 0$. Finally, choose $x_i = \delta_{ik} + i\delta_{il}$, giving $N_{kl} - N_{lk} = 0$. Altogether, $N_{kl} = 0$ for any kl.

Or: from the decomposition: $(x, Nx) = (x, \mathrm{Re}Nx) + i(x, \mathrm{Im}Nx)$. Since both $\mathrm{Re}N$ and $\mathrm{Im}N$ are hermitian, both scalar products are real, and must be thus both be zero. But then also the matrices $\mathrm{Re}N$ and $\mathrm{Im}N$ themselves must be zero, thanks to the reasoning from the beginning of this footnote.

3 $\boxed{\text{Im } AMA^+ = A\ (\text{Im } M)\ A^+ \qquad M, A \text{ arbitrary}}$

Follows from

$$\text{Im } AMA^+ = \frac{1}{2i}(AMA^+ - (AMA^+)^+) = \frac{1}{2i}(AMA^+ - AM^+A^+) = A\frac{1}{2i}(M - M^+)A^+ =$$
$$= A\ (\text{Im } M)\ A^+ .$$

4 $\boxed{\text{Im}A \geq 0 \rightarrow \text{Im}A^{-1} \leq 0 \qquad \text{arbitrary A having an inverse}}$

Follows from, $B = A^{-1}$, $B^+ = (A^{-1})^+ = (A^+)^{-1}$,

$$B - B^+ = A^{-1} - (A^+)^{-1} = A^{-1}A^+(A^+)^{-1} - A^{-1}A(A^+)^{-1} = A^{-1}(A^+ - A)(A^+)^{-1} = B(A^+ - A)B^+$$

and thus

$$\text{Im}A^{-1} = \frac{1}{2i}(B - B^+) = \frac{1}{2i}B(A^+ - A)B^+ = B(-\text{Im}A)B^+ , \qquad (*)$$

meaning that ImA and ImA^{-1} have opposite (semi-)definiteness.

5 $\boxed{\text{Re}A \geq 0 \rightarrow \text{Re}A^{-1} \geq 0 \qquad \text{arbitrary A having an inverse}}$

... along the same lines as in the previous case, only with a + sign.

6 $\boxed{\text{Im}A \text{ is indefinite} \rightarrow \text{Im}A^{-1} \text{ is indefinite} \qquad \text{arbitrary A having an inverse}}$

Proof by contradiction: Suppose ImA^{-1} is definite, i.e Im$A^{-1} \geq (\leq) 0$. But then, according to the point 4, Im$(A^{-1})^{-1} \leq (\geq) 0$, i.e. Im$A \leq (\geq) 0$. This contradicts the assumption of ImA's indefiniteness.

7 $\boxed{\text{Re}A \text{ is indefinite} \rightarrow \text{Re}A^{-1} \text{ is indefinite} \qquad \text{arbitrary A having an inverse}}$

... along the same lines as in the previous case.

8 [In the following, by a vector we mean a set of numbers, not an abstract object.]
Let us have two *hermitely congruent* matrices A and B, i.e. there exists a non-singular X such that $A = XBX^+$. Suppose further that A or B is in addition hermitian (then both are necessarily hermitian). It can be shown then that the number of positive, negative and zero eigenvalues, the so-called *index of inertia, is the same for the two matrices*:

The matrices are diagonalizable, say $A = U^+MU$, $B = VNV^+$, where U and V are unitary and M and N are diagonal with (real) by-magnitude-ordered eigenvalues of A and B, respectively, in descending order (this can always be achieved with a permutation matrix, which is itself unitary). Thus we have $M = RNR^+$ with an invertible $R = UXV$. Suppose there is a different number of positive values in M and N, i and j, respectively, and without loss of generality let us suppose that $i > j$.

The i vectors $\begin{bmatrix}1\\0\\0\\0\\0\\\vdots\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\\0\\\vdots\end{bmatrix}, \ldots , \begin{bmatrix}0\\0\\\vdots\\1\\0\\\vdots\end{bmatrix}$ form a basis of an i-dimensional positive subspace Ω of M. On the other hand, the $(n-j)$ vectors $\begin{bmatrix}0\\\vdots\\1\\0\\0\\\vdots\end{bmatrix}, \begin{bmatrix}0\\\vdots\\0\\1\\0\\\vdots\end{bmatrix}, \ldots , \begin{bmatrix}0\\\vdots\\0\\0\\\vdots\\1\end{bmatrix}$ form a basis of an $(n-j)$-dimensional non-positive subspace Υ of N, where n is the number of matrix rows. Since R has an inverse, the space $(R^+)^{-1}\Upsilon$ is $(n-j)$-dimensional. Since $i + (n-j) > n$, there must be a non-trivial vector in $\Omega \cap (R^+)^{-1}\Upsilon$. Call it w. But then $(w, Mw) > 0$ and $(w, RNR^+w) \leq 0$, which contradicts $M = RNR^+$. Thus i cannot differ from j. For the number of negative values the argument would be similar, equality of the

number of zero values then follows immediately.

9 Using the knowledge of point 8 and the equation (*) of point 4 (and its analogy for the Re part as needed in the point 5), one can generalize the statements of points 4, 5, 6 and 7. It follows that always

$\xi(\mathrm{Im}A^{-1}) = -\xi(\mathrm{Im}A)$ and $\xi(\mathrm{Re}A^{-1}) = \xi(\mathrm{Re}A)$, for any invertible A, where ξ denotes the index of inertia.

List of Figures

List of Abbreviations

AMR	anisotropic magnetoresistance
AP	anti-parallel (configuration)
ASA	atomic-spheres approximation
BSF	Bloch spectral function
C	conductance
CDFT	current-density functional theory
CPA	coherent-potential approximation
DOS	density of states
DFT	density functional theory
FM	ferromagnetic (metallic material)
GMR	giant magneto-resistance (ratio)
I	insulator (or semiconductor)
KSD	Kohn−Sham−Dirac
LDA	local-density approximation
LMTO	linear muffin-tin orbital
LSDA	local-spin-density approximation
MBE	molecular beam epitaxy
MR	magnetoresistance
MTO	muffin-tin orbital
NM	nonmagnetic (metallic material)
P	parallel (configuration)
R	resistance
QED	quantum electrodynamics
RAM	random-access memory
RDFT	relativistic density functional theory
SDFT	spin-density functional theory
SO	spin-orbit
SR	scalar-relativistic
SRA	scalar-relativistic approximation
TAMR	tunnelling anisotropic magnetoresistance
TB	tight-binding
TMR	tunnelling magnetoresistance
VWN	Vosko−Wilk−Nusair

Bibliography

Ande71. O. K. Andersen and R. V. Kasowski, *Phys. Rev. B* **4**, p. 1064 (1971).

Ande73. O. K. Andersen and R. G. Woolley, *Molecular Physics* **26**, p. 905 (1973).

Ande75. O. K. Andersen, *Phys. Rev. B* **12**, p. 3060 (1975).

Ande84. O. K. Andersen, *Phys. Rev. Lett.* **53**, p. 2571 (1984).

Ande85. O. K. Andersen, O. Jepsen, and D. Glötzel, in *Highlights of Condensed-Matter Theory*, edited by F. Bassani, F. Fumi and M. P. Tosi, North-Holland, New York (1985).

Ande92. O. K. Andersen, A. V. Postnikov, and S. Yu. Savrasov, *Mat. Res. Soc. Symp. Proc.* **253**, p. 37 (1992).

Ande99. O. K. Andersen, T. Saha-Dasgupta, R. W. Tank, C. Arcangeli, O. Jepsen, and G. Krier, *Developing the MTO Formalism*, Lecture Notes in Physics 535, Springer-Verlag (1999). H. Dreyseé (Ed.)

Ande65. D. G. Anderson, *J. Assoc. Comput. Mach.* **12**(4), p. 547 (1965).

Ande03. P. W. Anderson, *Concepts in Solids*, World Scientific (2003).

Ashc76. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College Publishing (1976).

Baib88. M. N. Baibich and *et al.*, *Phys. Rev. Lett.* **61**, p. 2472 (1988).

Banh96. J. Banhart, A. Vernes, and H. Ebert, *Solid State Commun.* **98**, p. 129 (1996).

Banh97. J. Banhart, H. Ebert, and A. Vernes, *Phys. Rev. B* **56**, p. 10165 (1997).

Bara89. H. U. Baranger and A. D. Stone, *Phys. Rev. B* **40**, p. 8169 (1989).

Bart72. U. von Barth and L. Hedin, *J. Phys. C* **5**, p. 1629 (1972).

Bart94. U. von Barth, *Lectures on Methods of Electronic Structure Calculations*, World Scientific (1994). edited V. Kumar O. K. Andersen A. Mookerjee

Bass99. J. Bass and W. P. Pratt, Jr., *J. Magn. Magn. Mater.* **200**, p. 274 (1999).

Bela05. K. D. Belashchenko, J. Velev, and E. Y. Tsymbal, *Phys. Rev. B* **72**, p. 140404(R) (2005).

Bina89. G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, *Phys. Rev. B* **39**, p. 4828 (1989).

Blüg06. S. Blügel and others (Eds.), *Computational Condensed Matter Physics, 37th IFF Spring School of the Institude of Solid State Research*, Forschungszentrum Jülich (2006).

Bogo70. N. N. Bogoliubov, *Lectures on Quantum Statistics*, Vol. 2, Gordon and Breach, New York (1970).

Broy65. C. G. Broyden, *Math. Comp.* **19**, p. 577 (1965).

Butl01. W. H. Butler, X. G. Zhang, T. C. Schulthess, and J. M. MacLaren, *Phys. Rev. B* **63**, p. 054416 (2001).

Büt86. M. Büttiker, *Phys. Rev. Lett.* **57**, p. 1761 (1986).

Büt88. M. Büttiker, *IBM J. Res. Dev.* **32**, p. 317 (1988).

Camp70. I. A. Campbell, A. Fert, and O. Jaoul, *J. Phys. C: Solid State Phys.* **3**, p. S95 (1970).

Cape01. K. Capelle and G. Vignale, *Phys. Rev. Lett.* **86**, p. 5546 (2001).

Caro71. C. Caroli, R. Combescot, P.Nozieres, and D. Saint-James, *J. Phys. C: Solid St. Phys.* **4**, p. 916 (1971).

Carv06. K. Carva, I. Turek, J. Kudrnovský, and O. Bengone, *Phys. Rev. B* **73**, p. 144421 (2006).

Chad08. S. Chadov, J. Minár, M.I. Katsnelson, H. Ebert, D. Ködderitzsch, and A. I. Lichtenstein, *EPL* **82**, p. 37001 (2008).

Chan07a. A. N. Chantis, K. D. Belashchenko, E. Y. Tsymbal, and M. van Schilfgaarde, *Phys. Rev. Lett.* **98**, p. 046601 (2007).

Chan07b. A. N. Chantis, K. D. Belashchenko, D. L. Smith, E. Y. Tsymbal, M. van Schilfgaarde, and R. C. Albers, *Phys. Rev. Lett.* **99**, p. 196603 (2007).

Chay85. J. T. Chayes, L. Chayes, and M. B. Ruskai, *J. Stat. Phys.* **38**, p. 497 (1985).

Chia80. T. C. Chiang, J. A. Knapp, M. Aono, and D. E. Eastman, *Phys. Rev. B* **21**, p. 3513 (1980).

Chri84. N. E. Christensen, *Int. J. of Quantum Chemistry* **25**, p. 233 (1984).

Cohe77. C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics*, Hermann and John Wiley & Sons (1977).

Cohe97. C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Phonon and Atoms, Introduction to Quantum Electrodynamics*, John Wiley & Sons (1997).

Cohe08. A. J. Cohen and *et al.*, *Science* **321**, p. 792 (2008).

Cré01. A. Crépieux and P. Bruno, *Phys. Rev. B* **64**, p. 014416 (2001).

Datt97. S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press (1997).

Datt00. S. Datta, *Superlattices and Microstructures* **28**, p. 253 (2000).

Datt07. S. Datta, *Quantum Transport, Atom to Transistor*, Cambridge University Press (2007).

Dede02. P. H. Dederichs, P. Mavropoulos, O. Wunnicke, N. Papanikolaou, V. Bellini, R. Zeller, V. Drchal, and J. Kudrnovský, *J. Magn. Mater.* **240**, p. 108 (2002).

Dira30. P. A.M. Dirac, *Proc. Cambridge Phil. Soc.* **26**, p. 376 (1930).

Drch02. V. Drchal, J. Kudrnovský, P. Bruno, P. H. Dederichs, I. Turek, and P. Weinberger, *Phys. Rev. B* **65**, p. 214414 (2002).

Drei90. R. M. Dreizler and E. K. U. Gross, *Density Functional Theory, An Approach to the Quantum Many-Body Problem*, Springer−Verlag (1990).

Eber99. H. Ebert, *Fully Relativistic Band Structure Calculations for Magnetic Solids − Formalism and Application*, Lecture Notes in Physics 535, Springer-Verlag (1999). H. Dreyseé (Ed.)

Eber10. H. Ebert, S. Bornemann, J. Braun, D. Ködderitzsch, S. Lowitzer, S. Mankovskyy, J. Minár, M. Offenberger, S. Polesya, and V. Popescu, *Psi-k Newsletter* **97**, p. 79, http://www.psi-k.org/newsletters.shtml (2010).

Enge11. E. Engel and R. M. Dreizler, *Density Functional Theory: An Advanced Course*, Springer, Berlin (2011).

Engl83. H. Englisch and R. Englisch, *Physica* **121A**, p. 253 (1983).

Engl84. H. Englisch and R. Englisch, *Phys. Stat. Sol. (b)* **123**, p. 711, continued in *Phys. Stat. Sol. (b)* **124**, p. 373 (1984).

Erik99. O. Eriksson and J. Wills, *First Principles Theory of Magneto-Crystalline Anisotropy*, Lecture Notes in Physics 535, Springer-Verlag (1999). H. Dreyseé (Ed.)

Esch85. H. Eschrig, *Solid State Commun.* **56**, p. 777 (1985).

Esch01. H. Eschrig and W. E. Pickett, *Solid State Commun.* **118**, p. 123 (2001).

Esch03. H. Eschrig, *The Fundamentals of Density Functional Theory*, Edition am Gutenbergplatz (2003).

Esch05. H. Eschrig, M. Sargolzaei, K. Koepernik, and M. Richter, *Europhys. Lett.* **72**, p. 611 (2005).

Ferm27. E. Fermi, *Rend. Accad. Naz. Linzei* **6**, p. 602 (1927).

Ferm28. E. Fermi, *Z. Phys.* **48**, p. 73 (1928).

Fiol03. C. Fiolhais, F. Nogueira, and M. Marques (eds.), *A Primer in Density Functional Theory*, Lecture Notes in Physics 620, Springer-Verlag (2003).

Fish81. D. S. Fisher and P. A. Lee, *Phys. Rev. B* **23**, p. 6851 (1981).

Form00. J. Formánek, *Úvod do relativistické kvantové mechaniky a kvantové teorie pole 1*, (in Czech), Karolinum, Praha (2000).

Gilb75. T. L. Gilbert, *Phys. Rev. B* **12**, p. 2111 (1975).

Giov02. C. Giovanardi, A. di Bona, T. S. Moia, S. Valeri, C. Pisani, M. Sgroi, and M. Busso, *Surf. Science* **505**, p. L209 (2002).

Gord28. W. Gordon, *Z. Physik* **50**, p. 630 (1928).

Goul04. C. Gould, C. Rüster, T. Jungwirth, E. Girgis, G. M. Schott, R. Giraud, K. Brunner, G. Schmidt, and L. W. MolenKamp, *Phys. Rev. Lett.* **93**, p. 117203 (2004).

Gree58. D. A. Greenwood, *Proc. Phys. Soc.* **71**, p. 585 (1958).

Grü86. P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sowers, *Phys. Rev. Lett.* **57**, p. 2442 (1986).

Gunn74. O. Gunnarsson, B. I. Lundquist, and J. W. Wilkins, *Phys. Rev. B* **10**, p. 1319 (1974).

Gunn76. O. Gunnarsson and B. I. Lundquist, *Phys. Rev. B* **13**, p. 4274 (1976).

Gunn79. O. Gunnarsson, M. Jonson, and B. I. Lundquist, *Phys. Rev. B* **20**, p. 3136 (1979).

Haya06. K. Hayashi and S. Sasa, *Physica A* **370**, p. 407 (2006).

Hedi69. L. Hedin and S. Lundquist, *Solid State Physics* **23**, p. 1, edited by H. Ehenreich, F. Seitz and D. Turnbull (1969).

Hedi71. L. Hedin and B. I. Lundquist, *J. Phys. C* **4**, p. 2064 (1971).

Heil06. Ch. Heiliger, P. Zahn, and I. Mertig, *Materials Today* **9**, p. 46 (2006).

Heil08. Ch. Heiliger, P. Zahn, B. Y. Yavorsky, and I. Mertig, *Phys. Rev. B* **77**, p. 224407 (2008).

Hein60. V. Heine, *Group Theory in Quantum Mechanics*, Pergamon Press, New York (1960).

Hohe64. P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, p. B864 (1964).

Iked08. S. Ikeda and *et al.*, *Appl. Phys. Lett.* **93**, p. 082508 (2008).

Imry99. Y. Imry and R. Landauer, *Rev. Mod. Phys.* **71**, p. S306 (1999).

Jone89. R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, p. 689 (1989).

Jone06. R. O. Jones, ‘‘Introduction to Density Functional Theory and Exchange-Correlation Energy Functionals,’’ *Computational Nanoscience: Do It Yourself!, NIC Series* **31**, p. 45, Lecture notes, ed. by J. Grotendorst, S. Blügel and D. Marx, John von Neumann Institude for Computing, Jülich, available online: http://www2.fz-juelich.de/nic-series/volume31/volume31.html (2006).

Jull75. M. Jullière, *Phys. Lett. A* **54**, p. 225 (1975).

Kada62. L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics*, W. A. Benjamin, Inc. (1962).

Kaso72. R. V. Kasowski and O. K. Andersen, *Solid State Commun.* **11**, p. 799 (1972).

Khan08. M. N. Khan, J. Henk, and P. Bruno, *J. Phys.: Condens. Matter* **20**, p. 155208 (2008).

Koel77. D. D. Koelling and B. N. Harmon, *J. Phys. C: Solid State Phys.* **10**, p. 3107 (1977).

Kohn65. W. Kohn and L. Sham, *Phys. Rev.* **140**, p. B1133 (1965).

Kohn85. W. Kohn, p. 1 in *Highlights of Condensed Matter Theory*, ed. by F. Bassani, F. Fumi and M. P. Tosi, North-Holland, Amsterdam (1985).

Kohn99. W. Kohn, *Rev. Mod. Phys.* **71**, p. 1253 (1999).

Kohn04. W. Kohn, A. Savin, and C. A. Ullrich, *Int. J. Quant. Chem.* **100**, p. 20 (2004).

Kost57. G. F. Koster, *Solid State Physics* **5**, p. 173, Academic Press, New York, edited by F. Seitz and D. Turnbull (1957).

Kubo57. R. Kubo, *J. Phys. Soc. Jpn.* **12**, p. 570 (1957).

Kudr00. J. Kudrnovský, V. Drchal, C. Blass, P. Weinberger, I. Turek, and P. Bruno, *Phys. Rev. B* **62**, p. 15084 (2000).

Kuma94. V. Kumar, O. K. Andersen, and A. Mookerjee , *Lectures on Methods of Electronic Structure Calculations*, World Scientific (1994).

Land57. R. Landauer, *IBM J. Res. Dev.* **1**, p. 223 (1957).

Land87. R. Landauer, *Z. Phys. B − Condensed Matter* **68**, p. 217 (1987).

Land88. R. Landauer, *IBM J. Res. Dev.* **32**, p. 306 (1988).

Lax58. M. Lax, *Phys. Rev.* **109**, p. 1921 (1958).

Levy79. M. Levy, *Proc. Natl. Acad. Sci. USA* **76**, p. 6062 (1979).

Levy82. M. Levy, *Phys. Rev. A* **26**, p. 1200 (1982).

Lieb82. E. H. Lieb, p. 111 in *Physics as Natural Philosophy*, ed. by A. Shimoni and H. Feshbach, MIT Press, Cambridge (1982). A revised version appeared in *Int. J. Quant. Chem.* **24**, p. 243 (1983) and an extended version in *Density Functional Methods in Physics*, NATO ASI Series B123, ed. by R. M. Dreizler and J. da Providencia, p. 31, Plenum, New York (1985).

MacD79. A. H. MacDonald and S. H. Vosko, *J. Phys. C: Solid State Phys.* **12**, p. 2977 (1979).

Manc11. A. Manchon, *Phys. Rev. B* **83**, p. 172403 (2011).

Mart04. R. M. Martin, *Electronic Structure, Basic Theory and Practical Methods*, Cambridge University Press (2004).

Math96. J. Mathon, *Phys. Rev. B* **54**, p. 55 (1996).

Math97. J. Mathon, *Phys. Rev. B* **55**, p. 960 (1997).

Math01. J. Mathon and A. Umerski, *Phys. Rev. B* **63**, p. 220403 (2001).

Mato09. A. Matos-Abiague and J. Fabian, *Phys. Rev. B* **79**, p. 155303 (2009).

Mavr04. P. Mavropoulos, N. Papanikolaou, and P. H. Dederichs, *Phys. Rev. B* **69**, p. 125104 (2004).

McGu75. T. R. McGuire and R. I. Potter, *IEEE Transactions on Magnetism* **MAG-11**, p. 1018 (1975).

Meye02. H. L. Meyerheim, R. Popescu, N. Jedrecy, M. Vedpathak, M. Sauvage-Simkin, R. Pinchaux, B. Heinrich, and J. Kirschner, *Phys. Rev. B* **65**, p. 144433 (2002).

Miro10. I. M. Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, *Nat. Mater.* **9**, p. 230 (2010).

Mose07. J. Moser, A. Matos-Abiague, D. Schuh, W. Wegschreider, J. Fabian, and D. Weiss, *Phys. Rev. Lett.* **99**, p. 056601 (2007).

Orti94. G. Ortiz and P. Ballone, *Phys. Rev. B* **50**, p. 1391 (1994).

Park11. B. G. Park, J. Wunderlich, X. Martí, V. Holý, Y. Kurosaki, M. Yamada, H. Yamamoto, A. Nishide, J. Haykawa, H. Takahashi, A. B. Shick, and T. Jungwirth, *Nat. Mater.* **10**, p. 347 (2011).

Perd81. J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, p. 5048 (1981).

Pyyk88. P. Pyykkö, *Chem. Rev.* **88**, p. 563 (1988).

Raja73. A. K. Rajagopal and J. Callaway, *Phys. Rev. B* **7**, p. 1912 (1973).

Saku67. J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley Publishing (1967).

Skri84. H. L. Skriver, *The LMTO Method, Muffin-Tin Orbitals and Electronic Structure*, Springer-Verlag (1984).

Slat37. J. C. Slater, *Phys. Rev.* **51**, p. 846 (1937).

Smit51. J. Smit, *Physica* **17**, p. 612 (1951).

Ston88. A. D. Stone and A. Szafer, *IBM J. Res. Dev.* **32**, p. 384 (1988).

Stra98. P. Strange, *Relativistic Quantum Mechanics,* Cambridge University Press (1998).

Szmy07. R. Szmytkowski, *J. Math. Chem.* **42**, p. 397 (2007).

Söd92. P. Söderlind, B. Johansson, and O. Eriksson, *J. Magn. Magn. Mater.* **104-107**, p. 2037 (1992).

Sýk12a. R. Sýkora and I. Turek, *J. Nanosci. Nanotechnol.* **12**, p. ― (2012). in press,

Sýk12b. R. Sýkora and I. Turek, submitted to *New J. Phys.*, available online at http://arxiv.org/abs/1206.0615, 2012.

Thom27. L. H. Thomas, *Proc. Cambridge Phil. Soc.* **23**, p. 542 (1927).

Tius04. C. Tiusan, J. Faure-Vincent, C. Bellouard, M. Hehn, E. Jouguelet, and A. Schuhl, *Phys. Rev. Lett.* **93**, p. 106602 (2004).

Todo02. T. N. Todorov, *J. Phys.: Condens. Matter* **14**, p. 3049 (2002).

Tsym07. E. Y. Tsymbal, K. D. Belashchenko, J. P. Velev, S. S. Jaswal, M. van Schilfgaarde, I. I. Oleynik, and D. A. Stewart, *Prog. Mater. Sci.* **52**, p. 401 (2007).

Ture97. I. Turek, V. Drchal, J. Kudrnovský, M. Šob, and P. Weinberger, *Electronic Structure of Disordered Alloys, Surfaces and Interfaces,* Kluwer Academic Publishers (1997).

Ture06. I. Turek, K. Carva, and J. Kudrnovský, *Advances in Science and Technology* **52**, p. 1 (2006).

Ture08. I. Turek, V. Drchal, and J. Kudrnovský, *Philos. Magazine* **88**, p. 2787 (2008).

Ture12. I. Turek, J. Kudrnovský, and K. Carva, manuscript in preparation, 2012.

Uemu11. T. Uemura, M. Harada, T. Akiho, K. Matsuda, and M. Yamamoto, *Appl. Phys. Lett.* **98**, p. 102503 (2011).

Vent08. M. Di Ventra, *Electrical Transport in Nanoscale Systems,* Cambridge University Press (2008).

Vign88. G. Vignale and M. Rasolt, *Phys. Rev. B* **37**, p. 10685 (1988).

Vosk80. S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, p. 1200 (1980).

Wash92. S. Washburn and R. A. Webb, *Rep. Prog. Phys.* **55**, p. 1311 (1992).

Wein95. S. Weinberg, *The Quantum Theory of Fields,* Volume I, Foundations, Cambridge University Press (1995).

Weiz35. C. F. von Weizsäcker, *Z. Phys.* **96**, p. 431 (1935).

Wign34. E.P. Wigner, *Phys. Rev.* **46**, p. 1002 (1934).

Wign38. E.P. Wigner, *Trans. Faraday Soc.* **34**, p. 678 (1938).

Wort06. D. Wortmann, ‘‘*Ab Initio* Description of Electronic Transport,’’ *Computational Nanoscience: Do It Yourself!, NIC Series* **31**, p. 469, Lecture notes, ed. by J. Grotendorst, S. Blügel and D. Marx, John von Neumann Institude for Computing, Jülich, available online: http://www2.fz-juelich.de/nic-series/volume31/ volume31.html (2006).

Wunn02. O. Wunnicke, N. Papanikolaou, R. Zeller, P. H. Dederichs, V. Drchal, and J. Kudrnovský, *Phys. Rev. B* **65**, p. 064425 (2002).

Xia01. K. Xia, P. J. Kelly, G. E. W. Bauer, I. Turek, J. Kudrnovský, and V. Drchal, *Phys. Rev. B* **63**, p. 064407 (2001).

Yuss87. F. Yussouff (ed.), *Electronic Band Structure and Its Applications,* Lecture Notes in Physics 283, Springer-Verlag (1987).

Zahn98. P. Zahn, J. Binder, I. Mertig, R. Zeller, and P. H. Dederichs, *Phys. Rev. Lett.* **80**, p. 4309 (1998).