# Solid State Theory

**Spring Semester 2011**

Manfred Sigrist
Institut für Theoretische Physik HIT K23.8
Tel.: 044-633-2584
Email: sigrist@itp.phys.ethz.ch
Website: http://www.itp.phys.ethz.ch/research/condmat/strong/

**Lecture Website:**
http://www.itp.phys.ethz.ch/education/lectures_fs11/solid

**Literature:**

- N.W. Ashcroft and N.D. Mermin: *Solid State Physics*, HRW International Editions, 1976.
- C. Kittel: *Einführung in die Festkörperphysik*, R. Oldenburg Verlag, 1983.
- C. Kittel: *Quantentheorie der Festkörper*, R. Oldenburg, 1970.
- O. Madelung: *Introduction to solid-state theory*, Springer 1981; auch in Deutsch in drei Bänden: *Festköperphysik I-III*, Springer.
- J.M. Ziman: *Prinzipien der Festkörpertheorie*, Verlag Harri Deutsch, 1975.
- M.P. Marder: *Condensed Matter Physics*, John Wiley & Sons, 2000.
- G. Grosso & G.P. Parravicini: *Solid State Physics*, Academic Press, 2000.
- G. Czychol: *Theoretische Festkörperphysik*, Springer 2004.
- P.L. Taylor & O. Heinonen, *A Quantum Approach to Condensed Matter Physics*, Cambridge Press 2002.
- numerous specialized books.

# Contents

# Introduction

Solid state physics (or condensed matter physics) is one of the most active and versatile branches of modern physics that have developed in the wake of the discovery of quantum mechanics. It deals with problems concerning the properties of materials and, more generally, systems with many degrees of freedom, ranging from fundamental questions to technological applications. This richness of topics has turned solid state physics into the largest subfield of physics; furthermore, it has arguably contributed most to technological development in industrialized countries.

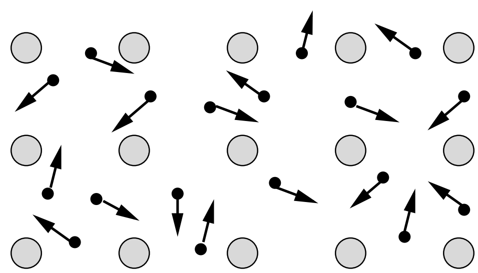


Figure 1: Atom cores and the surrounding electrons.

Condensed matter (solid bodies) consists of atomic nuclei (ions), usually arranged in a regular (elastic) lattice, and of electrons (see Figure 1). As the macroscopic behavior of a solid is determined by the dynamics of these constituents, the description of the system requires the use of quantum mechanics. Thus, we introduce the Hamiltonian describing nuclei and electrons,

$$\widehat{H} = \widehat{H}_{\mathrm{e}} + \widehat{H}_{\mathrm{n}} + \widehat{H}_{\mathrm{n-e}}, \tag{1}$$

with

$$\begin{aligned} \widehat{H}_{\mathrm{e}} &= \sum_i \frac{\widehat{\boldsymbol{p}}_i^2}{2m} + \frac{1}{2}\sum_{i\neq i'} \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|}, \\ \widehat{H}_{\mathrm{n}} &= \sum_j \frac{\widehat{\boldsymbol{P}}_j^2}{2M_j} + \frac{1}{2}\sum_{j\neq j'} \frac{Z_j Z_{j'} e^2}{|\boldsymbol{R}_j - \boldsymbol{R}_{j'}|}, \\ \widehat{H}_{\mathrm{n-e}} &= -\sum_{i,j} \frac{Z_j e^2}{|\boldsymbol{r}_i - \boldsymbol{R}_j|}, \end{aligned} \tag{2}$$

where $\widehat{H}_{\mathrm{e}}$ ($\widehat{H}_{\mathrm{n}}$) describes the dynamics of the electrons (nuclei) and their mutual interaction and $\widehat{H}_{\mathrm{n-e}}$ includes the interaction between ions and electrons. The parameters appearing are

| | | |
|---|---|---|
| $m$ | free electron mass | $9.1094 \times 10^{-31}$kg |
| $e$ | elementary charge | $1.6022 \times 10^{-19}$As |
| $M_j$ | mass of $j$-th nucleus | $\sim 10^3 - 10^4 \times$m |
| $Z_j$ | atomic (charge) number of $j$-th nucleus | |

The characteristic scales known from atomic and molecular systems are

| | |
|---|---|
| Length: Bohr radius | $a_B = \hbar^2/me^2 \approx 0.5 \times 10^{-10}$m |
| Energy: Hartree | $e^2/a_B = me^4/\hbar^2 = mc^2\alpha^2 \approx 27$eV $= 2$Ry |

with the fine structure constant $\alpha = e^2/\hbar c = 1/137$. The energy scale of one Hartree is much less than the (relativistic) rest mass of an electron ($\sim 0.5$MeV), which in turn is considered small in particle physics. In fact, in high-energy physics even physics at the Planck scale is considered, at least theoretically. The Planck scale is an energy scale so large that even gravity is thought to be affected by quantum effects, as

$$E_{\text{Planck}} = c^2\sqrt{\frac{\hbar c}{G}} \sim 10^{19}\text{GeV}, \qquad l_{\text{Planck}} = \sqrt{\frac{\hbar G}{c^3}} \sim 1.6 \times 10^{-35}\text{m}, \tag{3}$$

where $G = 6.673 \times 10^{-11}\text{m}^3\text{kg}^{-1}\text{s}^{-2}$ is the gravitational constant. This is the realm of the GUT (grand unified theory) and string theory. The goal is not to provide a better description of electrons or atomic cores, but to find the most fundamental theory of physics.

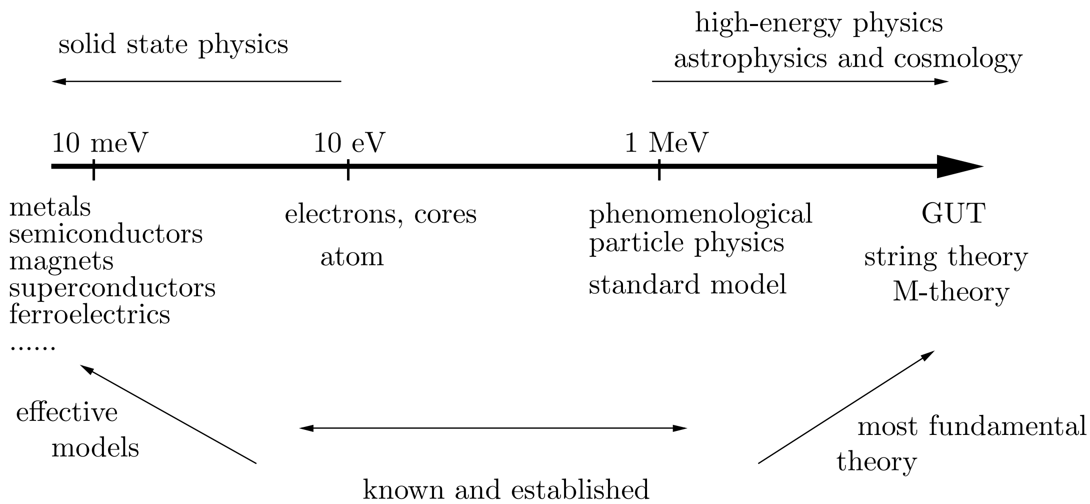



Figure 2: Energy scales in physics.

In contrast, in solid state physics we are dealing with phenomena occurring at room temperature ($T \sim 300$K) or below, i.e., at characteristic energies of about $E \sim k_BT \sim 0.03\text{eV} = 30\text{meV}$, which is even much smaller than the energy scale of one Hartree. Correspondingly, the important length scales are given by the extension of the system or of the electronic wave functions. The focus is thus quite different from the one of high-energy physics.
There, a highly successful phenomenological theory for low energies, the so-called standard model, exists, whereas the underlying theory for higher energies is unknown. In solid state physics, the situation is reversed. The Hamiltonian (1, 3) describes the known 'high-energy' physics (on the energy scale of Hartree), and one aims at describing the low-energy properties using reduced (effective, phenomenological) theories. Both tasks are far from trivial.
Among the various states of condensed matter that solid state theory seeks to describe are metals, semiconductors, and insulators. Furthermore, there are phenomena such as magnetism, superconductivity, ferroelectricity, charge ordering, and the quantum Hall effect. All of these states share a common origin: Electrons interacting among themselves and with the ions through the Coulomb interaction. More often than not, the microscopic formulation in (1) is too complicated to allow an understanding of the low-energy behavior from first principles. Consequently, the formulation of effective (reduced) theories is an important step in condensed matter theory. On the one hand, characterizing the ground state of a system is an important goal in itself. However, measurable quantities are determined by excited states, so that the concept of 'elementary excitations' takes on a central role. Some celebrated examples are Landau's quasiparticles for

Fermi liquids, the phonons connected to lattice vibrations, and magnons in ferromagnets. The idea is to treat the ground state as an effective vacuum in the sense of second quantization, with the elementary excitations as particles on that vacuum. Depending on the system, the vacuum may be the Fermi sea or some state with a broken symmetry, like a ferromagnet, a superconductor, or the crystal lattice itself.

According to P. W. Anderson,[1] the description of the properties of materials rests on two principles: The principle of *adiabatic continuity* and the principle of *spontaneously broken symmetry*. By adiabatic continuity we mean that complicated systems may be replaced by simpler systems that have the same essential properties in the sense that the two systems may be adiabatically deformed into each other without changing qualitative properties. Arguably the most impressive example is Landau's Fermi liquid theory mentioned above. The low-energy properties of strongly interacting electrons are the same as those of non-interacting fermions with renormalized parameters. On the other hand, phase transitions into states with qualitatively different properties can often be characterized by broken symmetries. In magnetically ordered states the rotational symmetry and the time-reversal invariance are broken, whereas in the superconducting state the global gauge symmetry is. In many cases the violation of a symmetry is a guiding principle which helps to simplify the theoretical description considerably. Moreover, in recent years some systems have been recognized as having topological order which may be considered as a further principle to characterize low-energy states of matter. A famous example for this is found in the context of the Quantum Hall effect.

The goal of these lectures is to introduce these basic concepts on which virtually all more elaborate methods are building up. In the course of this, we will cover a wide range of frequently encountered ground states, starting with the theory of metals and semiconductors, proceeding with magnets, Mott insulators, and finally superconductors.

[1] P.W. Anderson: *Basic Notions of Condensed Matter Physics*, Frontiers in Physics Lecture Notes Series, Addison-Wesley (1984).

# Chapter 1

# Electrons in the periodic crystal

In this chapter we discuss the properties of extended electron states in a regular lattice of ions. The presence of the lattice modifies the spectrum of the electrons as compared to that of free particles, leading to separate energy bands, which determine the qualitative properties of a solid. In particular, the structure of the electron bands allows a basic distinction between metals, insulators, and semiconductors.
In the initial considerations we neglect the interactions among the electrons as well as the dynamics of the ions. This simplification leads to a single particle description, to which Bloch's theorem can be applied.

## 1.1 Bloch states

### 1.1.1 Crystal symmetry

Many solid bodies can be described by a periodic array of positively charged ions forming a perfect crystal (kept together by electrons either forming chemical bonds or mobile conduction electrons). By definition, a crystal is characterized by the so-called *space group* $\mathcal{R}$, a group of operations (translations, rotations, the inversion or combinations) under which the crystal is left invariant. In three dimensions, there are 230 different space groups[1] (cf. Table 1.1). Translations are represented by a basic set of primitive translation vectors $\{\boldsymbol{a}_i\}$, which leave the lattice invariant. A translation by one of these vectors shifts a unit cell of the lattice to a neighboring cell. Any translation that maps the lattice onto itself is a linear combination of the $\{\boldsymbol{a}_i\}$ with integer coefficients,

$$\boldsymbol{a} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3. \tag{1.1}$$

A general symmetry transformation including the other elements of the space group may be written in the notation due to Wigner,

$$\boldsymbol{r}' = g\boldsymbol{r} + \boldsymbol{a} = \{g|\boldsymbol{a}\}\boldsymbol{r}, \tag{1.2}$$

where $g$ represents a rotation, reflection or inversion. The elements $g$ form the so-called generating *point group* $\mathcal{P}$. In three dimensions there are 32 point groups. The different types of symmetry operations involve

| | |
|---|---|
| basic translations | $\{E\|\boldsymbol{a}\}$, |
| rotations, reflections, inversions | $\{g\|\boldsymbol{0}\}$, |
| screw axes, glide planes | $\{g\|\boldsymbol{a}\}$, |

[1] A short guide to group theory is given in the textbook
- Peter Y. Yu and Manuel Cardona: *Fundamentals of Semiconductors*, Springer.

where $E$ is the unit element of $\mathcal{P}$. A screw axis is a symmetry operation of a rotation followed by a translation along the rotation axis. A glide plane is a symmetry operation with reflection at the plane followed by a translation along the plane. The symmetry operations $\{g|\boldsymbol{a}\}$, together with the associative multiplication

$$\{g|\boldsymbol{a}\}\{g'|\boldsymbol{a}'\} = \{gg'|g\boldsymbol{a}' + \boldsymbol{a}\} \tag{1.3}$$

form a group with unit element $\{E|\boldsymbol{0}\}$. In general, these groups are non-Abelian, i.e., the group elements do not commute with each other. However, there is always an Abelian subgroup of $\mathcal{R}$, the group of translations $\{E|\boldsymbol{a}\}$. The elements $g \in \mathcal{P}$ do not necessarily form a subgroup, because some of these elements (e.g., screw axes or glide planes) leave the lattice invariant only in combination with a translation. Nevertheless,

$$\{g|\boldsymbol{a}\}\{E|\boldsymbol{a}'\}\{g|\boldsymbol{a}\}^{-1} = \{E|g\boldsymbol{a}'\} \tag{1.4}$$

$$\{g|\boldsymbol{a}\}^{-1}\{E|\boldsymbol{a}'\}\{g|\boldsymbol{a}\} = \{E|g^{-1}\boldsymbol{a}'\} \tag{1.5}$$

always holds. If $\mathcal{P}$ is a subgroup of $\mathcal{R}$, then $\mathcal{R}$ is said to be *symmorphic*. In this case, the space group contains only primitive translations $\{E|\boldsymbol{a}\}$ and neither screw axes nor glide planes. The 14 Bravais lattices[2] are symmorphic. Among the 230 space groups 73 are symmorphic and 157 are non-symmorphic.

| crystal system (# point groups, # space groups) | point groups Schönflies symbols | space group numbers international tables |
|---|---|---|
| triclinic (2,2) | $C_1, C_{\bar{1}}$ | 1-2 |
| monoclinic (3,13) | $C_2, C_s, C_{2h}$ | 3-15 |
| orthorhombic (3,59) | $D_2, C_{2v}, D_{2h}$ | 16-74 |
| tetragonal (7,68) | $C_4, S_4, C_{4h}, D_4, C_{4v}, D_{2d}, D_{4h}$ | 75-142 |
| trigonal (5,25) | $C_3, S_6, D_3, C_{3v}, D_{3d}$ | 143-167 |
| hexagonal (7,27) | $C_6, C_{3h}, C_{6h}, D_6, C_{6v}, D_{3h}, D_{6h}$ | 168-194 |
| cubic (5, 36) | $T, T_h, O, T_d, O_h$ | 195-230 |

Table 1.1: List of the point and space groups for each crystal system.

### 1.1.2 Bloch's theorem

We consider a Hamiltonian $\mathcal{H}$ which is invariant under a discrete set of lattice translations $\{\{E|\boldsymbol{a}\}\}$. This discrete translational invariance can be induced by the periodic ionic potential and implies, that the corresponding translation operator $\widehat{T}_{\boldsymbol{a}}$ on the Hilbert space commutes with the Hamiltonian $\mathcal{H} = \mathcal{H}_{\mathrm{e}} + \mathcal{H}_{\mathrm{ie}}$ (purely electronic Hamiltonian $\mathcal{H}_{\mathrm{e}}$, interaction between electrons and ions $\mathcal{H}_{\mathrm{ie}}$),

$$[\widehat{T}_{\boldsymbol{a}}, \mathcal{H}] = 0. \tag{1.6}$$

[2]Crystal systems, crystals, Bravais lattices are discussed in more detail in

- Czycholl, *Theoretische Festkörperphysik*, Springer

Neglecting the interactions between the electrons, which in general is contained in $\mathcal{H}_\mathrm{e}$, we are left with a single particle problem

$$\mathcal{H} \to \mathcal{H}_0 = \frac{\widehat{\boldsymbol{p}}^2}{2m} + V(\widehat{\boldsymbol{r}}). \tag{1.7}$$

where $\widehat{\boldsymbol{r}}$ and $\widehat{\boldsymbol{p}}$ are position and momentum operators, while

$$V(\boldsymbol{r}) = \sum_j V_\mathrm{ion}(\boldsymbol{r} - \boldsymbol{R}_j), \tag{1.8}$$

describes the potential landscape of the single particle in the ionic background. With $\boldsymbol{R}_j$ the position of the $j$-th ion, the potential $V(\boldsymbol{r})$ is by construction translation invariant, or equivalently $V(\boldsymbol{r}+\boldsymbol{a}) = V(\boldsymbol{r})$ for all lattice translations $\boldsymbol{a}$. Therefore, $\mathcal{H}_0$ commutes with $\widehat{T}_{\boldsymbol{a}}$. For a Hamiltonian $\mathcal{H}_0$ commuting with the translation operator $\widehat{T}_{\boldsymbol{a}}$, the extended eigenstates of $\mathcal{H}_0$ are simultaneous eigenstates of $\widehat{T}_{\boldsymbol{a}}$. Bloch's theorem states that the eigenvalues of $\widehat{T}_{\boldsymbol{a}}$ lie on the unit circle of the complex plane. In other words, the wave function can be expressed as a product of a plane wave $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ and a periodic Bloch-factor $u_{\boldsymbol{k}}(\boldsymbol{r})$

$$\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{1.9}$$

with

$$\widehat{T}_{\boldsymbol{a}} u_{n,\boldsymbol{k}}(\boldsymbol{r}) = u_{n,\boldsymbol{k}}(\boldsymbol{r}-\boldsymbol{a}) = u_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{1.10}$$

$$\widehat{T}_{\boldsymbol{a}} \psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \psi_{n,\boldsymbol{k}}(\boldsymbol{r}-\boldsymbol{a}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{a}} \psi_{n,\boldsymbol{k}}(\boldsymbol{r}), \tag{1.11}$$

$$\mathcal{H}_0 \psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \epsilon_{n,\boldsymbol{k}} \psi_{n,\boldsymbol{k}}(\boldsymbol{r}). \tag{1.12}$$

The integer $n$ is a quantum number called band index, $\boldsymbol{k}$ is the pseudo-momentum (wave vector) and $\Omega$ represents the volume of the system. Note that the eigenvalue of $\psi_{n,\boldsymbol{k}}(\boldsymbol{r})$ with respect to $\widehat{T}_{\boldsymbol{a}}$, $e^{-i\boldsymbol{k}\cdot\boldsymbol{a}}$, implies periodicity in $\boldsymbol{k}$-space. The vectors $\boldsymbol{G}$ which satisfy $e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{a}} = e^{i\boldsymbol{k}\cdot\boldsymbol{a}}$ are termed reciprocal lattice vectors and they form a vector space. A basis of the reciprocal lattice vector space is found from the relation

$$e^{i\boldsymbol{G}_j\cdot\boldsymbol{a}_i} = 1 \tag{1.13}$$

or

$$\boldsymbol{G}_j \cdot \boldsymbol{a}_i = 2\pi\delta_{ij}. \tag{1.14}$$

The primitive basis set of the reciprocal space is given by

$$\boldsymbol{G}_i = 2\pi \frac{\boldsymbol{a}_j \times \boldsymbol{a}_k}{(\boldsymbol{a}_1 \times \boldsymbol{a}_2)\cdot\boldsymbol{a}_3} \tag{1.15}$$

where $i$, $j$, and $k$ are cyclic permutations of the indices 1, 2, and 3. These reciprocal basis vectors allow the definition of the *first Brillouin zone*: One draws lines joining $\boldsymbol{k} = 0$ and all neighboring reciprocal lattice points spanned by $\{\boldsymbol{G}_i\}$. The Brillouin zone is the smallest cell bounded by the planes that intersect these lines in their middle and which are orthogonal to them. This particular choice of a unit cell possesses all symmetry properties of the crystal. In the one dimensional simple periodic lattice of lattice spacing $a$, the interval $[-\pi/a, \pi/a]$ represents the first Brillouin zone in $\boldsymbol{k}$-space.
Bloch's theorem simplifies the initial problem to the so-called Bloch equation for the periodic function $u_{\boldsymbol{k}}$,

$$\left(\frac{(\widehat{\boldsymbol{p}} + \hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r})\right) u_{\boldsymbol{k}}(\boldsymbol{r}) = \epsilon_{\boldsymbol{k}} u_{\boldsymbol{k}}(\boldsymbol{r}), \tag{1.16}$$

where we suppressed the band index to simplify the notation. This equation follows from the relation

$$\widehat{\boldsymbol{p}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}(\widehat{\boldsymbol{p}} + \hbar\boldsymbol{k}), \tag{1.17}$$

which can be used for more complex forms of the Hamiltonian as well.[3]

## 1.2 Nearly free electron approximation

Various numerical methods allow to compute rather efficiently the band energies $\epsilon_{n,\boldsymbol{k}}$ for a given Hamiltonian $\mathcal{H}$. In order to understand some of the most essential aspects of the band structure of electrons in a crystal, we start with a simple analytical approach, the so-called nearly free electron approximation. We start by noting that the periodic potential can be expanded as

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}} e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{1.22}$$

$$V_{\boldsymbol{G}} = \frac{1}{\Omega_{\mathrm{UC}}} \int_{\mathrm{UC}} d^3r\, V(\boldsymbol{r}) e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{1.23}$$

where we sum over all reciprocal lattice vectors and the domain of integration is the unit cell (UC) with volume $\Omega_{\mathrm{UC}}$. We assume that the lattice is real and invariant under inversion ($V(\boldsymbol{r}) = V(-\boldsymbol{r})$), so that $V_{\boldsymbol{G}} = V^*_{-\boldsymbol{G}} = V_{-\boldsymbol{G}}$. Note that the uniform component $V_0$ corresponds to an irrelevant energy shift and may be set to zero. Because of its periodicity, the Bloch function $u_{\boldsymbol{k}}(\boldsymbol{r})$ can be expanded in the same way,

$$u_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} c_{\boldsymbol{G}} e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{1.24}$$

where the coefficients $c_{\boldsymbol{G}} = c_{\boldsymbol{G}}(\boldsymbol{k})$ are functions of $\boldsymbol{k}$. Inserting this Ansatz and the expansion (1.23) into the Bloch equation, (1.16), we obtain a linear eigenvalue problem for the band energies $\epsilon_{\boldsymbol{k}}$,

$$\left(\frac{\hbar^2}{2m}(\boldsymbol{k}+\boldsymbol{G})^2 - \epsilon_{\boldsymbol{k}}\right) c_{\boldsymbol{G}} + \sum_{\boldsymbol{G}'} V_{\boldsymbol{G}-\boldsymbol{G}'} c_{\boldsymbol{G}'} = 0. \tag{1.25}$$

The solution $c_{\boldsymbol{G}}(\boldsymbol{k})$ of equation (1.25) requires the determination of the eigenvalues of an infinite dimensional matrix. The resulting band energies $\epsilon_{\boldsymbol{k}}$ include corrections to the parabolic dispersion $\epsilon^{(0)}_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ due to the potential $V(\boldsymbol{r})$. It is straightforward to see that parabolic

[3] $\mathcal{H}_0$ may also contain the spin-orbit coupling, a relativistic correction which leads to an additional term

$$\mathcal{H}_0 = \frac{\widehat{\boldsymbol{p}}^2}{2m} + V(\widehat{\boldsymbol{r}}) + \frac{\hbar}{4m^2c^2}\{\boldsymbol{\sigma} \times \boldsymbol{\nabla} V(\widehat{\boldsymbol{r}})\} \cdot \widehat{\boldsymbol{p}}, \tag{1.18}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.19}$$

The Bloch equation in this case is given by

$$\left\{\frac{(\widehat{\boldsymbol{p}}+\hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r}) + \frac{\hbar}{4m^2c^2}(\boldsymbol{\sigma} \times \boldsymbol{\nabla} V(\boldsymbol{r})) \cdot (\widehat{\boldsymbol{p}} + \hbar\boldsymbol{k})\right\} u_{\boldsymbol{k}}(\boldsymbol{r}) = \epsilon_{\boldsymbol{k}} u_{\boldsymbol{k}}(\boldsymbol{r}). \tag{1.20}$$

The energy eigenstates are no longer spin eigenstates. Instead, they are of pseudo spinor form

$$u_{\boldsymbol{k},\pm}(\boldsymbol{r}) = \chi_{\boldsymbol{k},\pm\uparrow}(\boldsymbol{r})|\uparrow\rangle + \chi_{\boldsymbol{k},\pm\downarrow}(\boldsymbol{r})|\downarrow\rangle, \tag{1.21}$$

where $\sigma_z|\uparrow\rangle = +|\uparrow\rangle$ und $\sigma_z|\downarrow\rangle = -|\downarrow\rangle$. Upon adiabatically switching off spin-orbit coupling, the states with index $+/-$ turn into the usual spin eigenfunctions $|\uparrow\rangle$ and $|\downarrow\rangle$.

dispersion $\epsilon^{(0)}_{\boldsymbol{k}}$ is the correct energy spectrum in absence of the potential $V(\boldsymbol{r})$.
Under the assumption that the periodic modulation of the potential is weak, i.e., $|V_{\boldsymbol{G}}| \ll \hbar^2 \boldsymbol{G}^2/2m$, the problem (1.25) can be simplified. Here, we consider two limits for the wave vector $\boldsymbol{k}$ which are typically of interest. First, we choose $\boldsymbol{k}$ small, i.e., near the center of the Brillouin zone. The approximate solution of the equation (1.25) is then given by

$$c_{\boldsymbol{G}} \approx \begin{cases} 1 & \text{for } \boldsymbol{G} = \boldsymbol{0} \\ -\dfrac{2mV_{\boldsymbol{G}}}{\hbar^2\{(\boldsymbol{k}+\boldsymbol{G})^2 - \boldsymbol{k}^2\}} \ll 1 & \text{for } \boldsymbol{G} \neq \boldsymbol{0} \end{cases} \tag{1.26}$$

leading to the energy eigenvalue $\epsilon_{\boldsymbol{k}} \approx \hbar^2 \boldsymbol{k}^2/2m$ corresponding to the original parabolic band. Note that this form of $c_{\boldsymbol{G}\neq\boldsymbol{0}}$ resembles the lowest order correction in the Rayleigh-Schrödinger perturbation theory. This example corresponds to the lowest branch of the band structure within this approach.
Next, we consider the case when the denominator of the expression in equation (1.26) is small, i.e., $\boldsymbol{k}$ is in a range of the Brillouin zone where $\boldsymbol{k}^2 \approx (\boldsymbol{k}+\boldsymbol{G})^2$ for some reciprocal $\boldsymbol{G}$. This means that the parabolas centered around $\boldsymbol{0}$ and $-\boldsymbol{G}$ cross at $\boldsymbol{k} = -\boldsymbol{G}/2$. Choosing for $\boldsymbol{G}$ a primitive vector of the reciprocal lattice, the crossing point lies on the Brillouin zone boundary and represents a point of high symmetry within the Brillouin zone. This situation requires to take into account $c_{\boldsymbol{0}}$ and $c_{\boldsymbol{G}}$ on an equal footing, while other coefficients are still negligible. Then, the problem reduces to the coupled equations for these two coefficients,

$$\begin{pmatrix} \dfrac{\hbar^2 \boldsymbol{k}^2}{2m} - \epsilon_{\boldsymbol{k}} & V_{-\boldsymbol{G}} \\ V_{\boldsymbol{G}} & \frac{\hbar^2}{2m}(\boldsymbol{k}+\boldsymbol{G})^2 - \epsilon_{\boldsymbol{k}} \end{pmatrix} \begin{pmatrix} c_0 \\ c_{\boldsymbol{G}} \end{pmatrix} = 0 \tag{1.27}$$

Remember that $V_{\boldsymbol{G}} = V^*_{\boldsymbol{G}} = V_{-\boldsymbol{G}}$. From equation (1.27), the secular equation

$$\det \begin{bmatrix} \dfrac{\hbar^2 \boldsymbol{k}^2}{2m} - \epsilon_{\boldsymbol{k}} & V^*_{\boldsymbol{G}} \\ V_{\boldsymbol{G}} & \dfrac{\hbar^2(\boldsymbol{k}+\boldsymbol{G})^2}{2m} - \epsilon_{\boldsymbol{k}} \end{bmatrix} = 0 \tag{1.28}$$

follows, which allows us to determine

$$\epsilon_{\boldsymbol{k}} = \frac{1}{2}\left(\frac{\hbar^2}{2m}(\boldsymbol{k}^2 + (\boldsymbol{k}+\boldsymbol{G})^2) \pm \sqrt{[\frac{\hbar^2}{2m}(\boldsymbol{k}^2 - (\boldsymbol{k}+\boldsymbol{G})^2)]^2 + 4|V_{\boldsymbol{G}}|^2}\right). \tag{1.29}$$

For the symmetry point $\boldsymbol{k} = -\boldsymbol{G}/2$ and for $V_{\boldsymbol{G}} < 0$ we obtain

$$\epsilon_{-\boldsymbol{G}/2,\pm} = \frac{\hbar^2}{2m}\left(\frac{\boldsymbol{G}}{2}\right)^2 \pm |V_{\boldsymbol{G}}| \tag{1.30}$$

and

$$u_{\boldsymbol{k}}(\boldsymbol{r}) = e^{i\frac{\boldsymbol{G}\cdot\boldsymbol{r}}{2}} \begin{cases} \sin\frac{\boldsymbol{G}\cdot\boldsymbol{r}}{2}, & + \text{ anti-bonding}, \\ \cos\frac{\boldsymbol{G}\cdot\boldsymbol{r}}{2}, & - \text{ bonding}. \end{cases} \tag{1.31}$$

This result is equivalent to the splitting of a degenerate level through a symmetry breaking interaction (hybridization). Note that the scheme applied here is quite analogous to Rayleigh-Schrödinger perturbation theory for (nearly) degenerate energy levels.
The band structure can thus be constructed by the superposition of parabolic energy spectra centered around all reciprocal lattice points. At the crossing points of the parabolas we find a "band splitting" due to the periodically modulated potential. This leads to energy ranges where no Bloch states exist, so-called band gaps. An illustrative and simple band structure of this kind can straightforwardly be calculated in a one-dimensional regular lattice with a weak potential $V(x)$ as shown in Fig. 1.1.

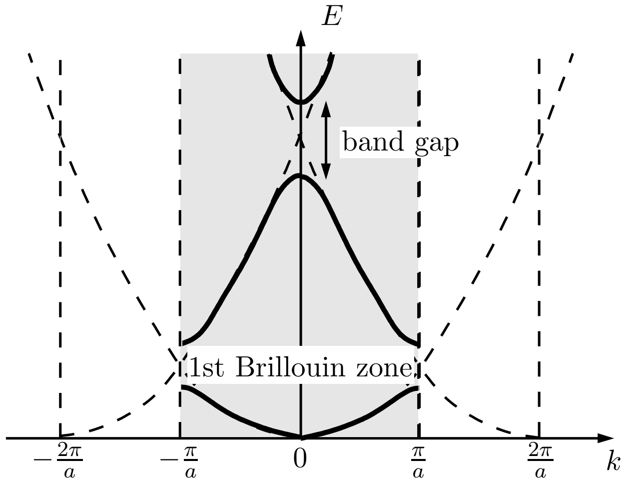



Figure 1.1: Band structure obtained by the nearly free electron approximation for a regular one-dimensional lattice.

## 1.3 Symmetry properties of the band structure

The symmetry properties of crystals are a helpful tool for the analysis of their band structure. They emerge from the symmetry group (space and point group) of the crystal lattice. Consider the action $\widehat{S}_{\{g|\boldsymbol{a}\}}$ of an element $\{g|\boldsymbol{a}\}$ of the space group on a Bloch wave function $\psi_{\boldsymbol{k}}(\boldsymbol{r})$

$$\widehat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}}(\{g|\boldsymbol{a}\}^{-1}\boldsymbol{r}) = \psi_{\boldsymbol{k}}(g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}). \tag{1.32}$$

Because $\{g|\boldsymbol{a}\}$ belongs to the space group of the crystal, we have $[\widehat{S}_{\{g|\boldsymbol{a}\}}, \mathcal{H}_0] = 0$. Applying a pure translation $\widehat{T}_{\boldsymbol{a}'} = \widehat{S}_{\{E|\boldsymbol{a}'\}}$ to this new wave function

$$\begin{aligned}\widehat{T}_{\boldsymbol{a}'}\widehat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) &= \widehat{S}_{\{g|\boldsymbol{a}\}}\widehat{T}_{g^{-1}\boldsymbol{a}'}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \widehat{S}_{\{g|\boldsymbol{a}\}}e^{-i\boldsymbol{k}\cdot(g^{-1}\boldsymbol{a}')}\psi_{\boldsymbol{k}}(\boldsymbol{r}) \\ &= \widehat{S}_{\{g|\boldsymbol{a}\}}e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}\psi_{\boldsymbol{k}}(\boldsymbol{r}) \\ &= e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}\widehat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}),\end{aligned} \tag{1.33}$$

latter is found to be an eigenfunction of $\widehat{T}_{\boldsymbol{a}'}$ with eigenvalue $e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}$. Remember, that, according to the Bloch theorem, we chose a basis $\{\psi_{\boldsymbol{k}}\}$ diagonalizing both $\widehat{T}_{\boldsymbol{a}}$ and $\mathcal{H}_0$. Thus, apart from a phase factor the action of a symmetry transformation $\{g|\boldsymbol{a}\}$ on the wave function[4] corresponds to a rotation from $\boldsymbol{k}$ to $g^{-1}\boldsymbol{k}$.

$$\widehat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \lambda_{\{g|\boldsymbol{a}\}}\psi_{g\boldsymbol{k}}(\boldsymbol{r}), \tag{1.35}$$

[4]Symmetry behavior of the wave function.

$$\begin{aligned}\widehat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) &= \frac{1}{\sqrt{\Omega}}\widehat{S}_{\{g|\boldsymbol{a}\}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}}c_{\boldsymbol{G}}(\boldsymbol{k})\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}} = \frac{1}{\sqrt{\Omega}}e^{i\boldsymbol{k}\cdot\left(g^{-1}\boldsymbol{r}-g^{-1}\boldsymbol{a}\right)}\sum_{\boldsymbol{G}}c_{\boldsymbol{G}}(\boldsymbol{k})\,e^{i\boldsymbol{G}\cdot\left(g^{-1}\boldsymbol{r}-g^{-1}\boldsymbol{a}\right)} \\ &= \frac{1}{\sqrt{\Omega}}e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}}e^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}}c_{\boldsymbol{G}}(\boldsymbol{k})\,e^{i(g\boldsymbol{G})\cdot\boldsymbol{r}} = e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}}\frac{1}{\sqrt{\Omega}}e^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}}c_{g^{-1}\boldsymbol{G}}(\boldsymbol{k})\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \\ &= e^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}}\frac{1}{\sqrt{\Omega}}e^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}}c_{\boldsymbol{G}}(g\boldsymbol{k})\,e^{i\boldsymbol{G}\cdot\boldsymbol{r}} = \lambda_{g|\boldsymbol{a}}\psi_{g\boldsymbol{k}}(\boldsymbol{r}),\end{aligned} \tag{1.34}$$

where we use the fact that $c_{\boldsymbol{G}} = c_{\boldsymbol{G}}(\boldsymbol{k})$ is a function of $\boldsymbol{k}$ with the property $c_{g^{-1}\boldsymbol{G}}(\boldsymbol{k}) = c_{\boldsymbol{G}}(g\boldsymbol{k})$ i.e. $\widehat{S}_{\{g|\boldsymbol{a}\}}u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{g\boldsymbol{k}}(\boldsymbol{r})$.

with $|\lambda_{\{g|\boldsymbol{a}\}}|^2 = 1$, or

$$\widehat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{k}\rangle = \lambda_{\{g|\boldsymbol{a}\}}|g\boldsymbol{k}\rangle. \tag{1.36}$$

in Dirac notation.[5] Then it is easy to see that

$$\epsilon_{g\boldsymbol{k}} = \langle g\boldsymbol{k}|\mathcal{H}_0|g\boldsymbol{k}\rangle = \langle \boldsymbol{k}|\widehat{S}^{-1}_{\{g|\boldsymbol{a}\}}\mathcal{H}_0\widehat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{k}\rangle = \langle \boldsymbol{k}|\mathcal{H}_0|\boldsymbol{k}\rangle = \epsilon_{\boldsymbol{k}}. \tag{1.40}$$

Consequently, there is a starlike structure of equivalent points $g\boldsymbol{k}$ with the same band energy (→ degeneracy) for each $\boldsymbol{k}$ in the Brillouin zone (cf. Fig. 1.2).

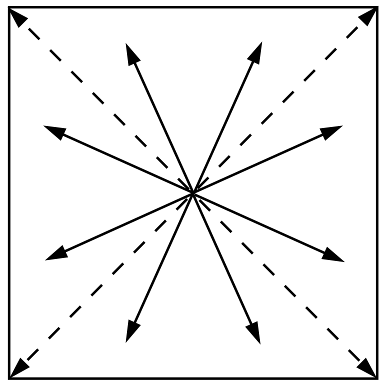


Figure 1.2: Star of $\boldsymbol{k}$-points in the Brillouin zone with degenerate band energy.

For a general point $\boldsymbol{k}$ the number of equivalent points in the star equals the number of point group elements for this $\boldsymbol{k}$ (without inversion). If $\boldsymbol{k}$ lies on points or lines of higher symmetry, it is left invariant under a subgroup of the point group. Consequently, the number of beams of the star is smaller. The subgroup of the point group leaving $\boldsymbol{k}$ unchanged is called *little group* of $\boldsymbol{k}$. If inversion is part of the point group, $-\boldsymbol{k}$ is always contained in the star of $\boldsymbol{k}$. In summary, we have the simple relations

$$\epsilon_{n\boldsymbol{k}} = \epsilon_{n,g\boldsymbol{k}}, \quad \epsilon_{n\boldsymbol{k}} = \epsilon_{n,-\boldsymbol{k}}, \quad \epsilon_{n\boldsymbol{k}} = \epsilon_{n,\boldsymbol{k}+\boldsymbol{G}}. \tag{1.41}$$

### Simple cubic lattice

Next, we will consider the energy bands $\epsilon_{n\boldsymbol{k}}$ on points and along lines of high symmetry in a simple cubic (sc) lattice (point group $O_h$) with lattice constant $a$, using the nearly free electron method. The reciprocal lattice is again sc with lattice constant $G = 2\pi/a$. Without periodic potential, the energy spectrum $\epsilon^{(0)}_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ is parabolic in $\boldsymbol{k}$.

[5] In Dirac notation we write for the Bloch state with pseudo-momentum $\boldsymbol{k}$

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \langle \boldsymbol{r}|\psi_{\boldsymbol{k}}\rangle. \tag{1.37}$$

The action of the operator $\widehat{S}_{\{g|\boldsymbol{a}\}}$ on the state $|\boldsymbol{r}\rangle$ is given by

$$\widehat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{r}\rangle = |g\boldsymbol{r}+\boldsymbol{a}\rangle \quad \text{and} \quad \langle \boldsymbol{r}|\widehat{S}_{\{g|\boldsymbol{a}\}} = \langle g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}|, \tag{1.38}$$

such that

$$\langle \boldsymbol{r}|\widehat{S}_{\{g|\boldsymbol{a}\}}|\psi_{\boldsymbol{k}}\rangle = \psi_{\boldsymbol{k}}(g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}). \tag{1.39}$$

The same holds for pure translations.

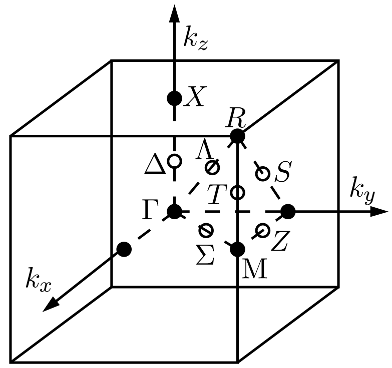



Figure 1.3: Points and lines of high symmetry within the first Brillouin zone (cube) of a simple cubic lattice.

### Γ-point

First, we consider the center of the Brillouin zone, usually called the Γ-point (cf. Figure 1.3). The lowest band at the Γ-point with energy $E_0 = \epsilon_{0\boldsymbol{k}=0} = 0$ belongs to the parabola around the center of the first Brillouin zone ($\epsilon_{0\boldsymbol{k}} \approx \hbar^2\boldsymbol{k}^2/2m$) and is non-degenerate. The next higher energy level for free electrons in a cubic crystal of lattice constant $a$ is

$$E_1 = \frac{\hbar^2 G^2}{2m} \tag{1.42}$$

where $G = 2\pi/a$ is the reciprocal unit vector and originates from the crossing of the parabolas centered around the six nearest neighbor points of the reciprocal lattice

$$\begin{aligned} \boldsymbol{G}_1 &= G(1,0,0), & \boldsymbol{G}_2 &= G(-1,0,0), \\ \boldsymbol{G}_3 &= G(0,1,0), & \boldsymbol{G}_4 &= G(0,-1,0), \\ \boldsymbol{G}_5 &= G(0,0,1), & \boldsymbol{G}_6 &= G(0,0,-1). \end{aligned} \tag{1.43}$$

The relevant basis functions for the expansion of the Bloch function are given by $e^{i\boldsymbol{r}\cdot\boldsymbol{G}_n}$ with $n = 1, \ldots, 6$ and we find

$$u_{\boldsymbol{k}=0}(\boldsymbol{r}) = \sum_{n=1}^{6} c_n e^{i\boldsymbol{r}\cdot\boldsymbol{G}_n}. \tag{1.44}$$

where $c_n = c_{\boldsymbol{G}_n}$. Assuming that all other contributions ($c_{\boldsymbol{G}\neq\boldsymbol{G}_n} \approx 0$) to be irrelevant, the secular equation from (1.25) reads

$$\det \begin{bmatrix} E_1-E & v & u & u & u & u \\ v & E_1-E & u & u & u & u \\ u & u & E_1-E & v & u & u \\ u & u & v & E_1-E & u & u \\ u & u & u & u & E_1-E & v \\ u & u & u & u & v & E_1-E \end{bmatrix} = 0 \tag{1.45}$$

with $v = V_{2\boldsymbol{G}_n}$ and $u = V_{\boldsymbol{G}_n+\boldsymbol{G}_{n'}}$ ($n \neq n'$). Solving this secular equation, we find, that there are three eigenvalues with corresponding eigenvectors (listed in Table 1.2). The six-fold degeneracy of the energy at the Γ-point of a free electron gas is partially lifted due to the potential components $v$ and $u$. Every energy eigenspace corresponds to an irreducible representations (Γ) with dimension $d_\Gamma$ of the point group at the Γ-point (→ degeneracy). The Γ-point shares the symmetry of the point group[6] of the crystal, which in this case is the cubic group $O_h$. A set of even

[6]Literature on point groups:

| $E=\epsilon_{1\boldsymbol{k}=0}$ | $(c_1,c_2,c_3,c_4,c_5,c_6)$ | $u_{\boldsymbol{k}=0}(\boldsymbol{r})$ | $\Gamma$ | $d_\Gamma$ |
|---|---|---|---|---|
| $E_1+v+4u$ | $(1,1,1,1,1,1)/\sqrt(6)$ | $\phi_0=\cos Gx+\cos Gy+\cos Gz$ | $\Gamma_1^+$ | 1 |
| $E_1+v-2u$ | $(-1,-1,-1,-1,2,2)/2\sqrt{3}$<br>$(1,1,-1,-1,0,0)/2$ | $\phi_{3z^2-r^2}=2\cos Gz-\cos Gx-\cos Gy$ ,<br>$\phi_{\sqrt{3}(x^2-y^2)}=\sqrt{3}(\cos Gx-\cos Gy)$ | $\Gamma_3^+$ | 2 |
| $E_1-v$ | $(1,-1,0,0,0,0)/\sqrt{2}$<br>$(0,0,1,-1,0,0)/\sqrt{2}$<br>$(0,0,0,0,1,-1)/\sqrt{2}$ | $\phi_x=\sin Gx$<br>$\phi_y=\sin Gy$<br>$\phi_z=\sin Gz$ | $\Gamma_4^-$ | 3 |

Table 1.2: Table of the eigenenergies, eigenvectors and corresponding Bloch functions at the Γ-point.

and odd irreducible representations belongs to this group. An irreducible representation can be specified by a vector space of functions of the vector $(x, y, z)$ or the pseudo-vector $(s_x, s_y, s_z)$ that are left invariant by symmetry operations of the group (see Table 1.3).

Note that each eigenvalue of the above secular equation can be attributed to one of the irre-

| even | basis function | odd | basis function |
|---|---|---|---|
| $\Gamma_1^+$ | $1, x^2+y^2+z^2$ | $\Gamma_1^-$ | $xyz(x^2-y^2)(y^2-z^2)(z^2-x^2)$ |
| $\Gamma_2^+$ | $(x^2-y^2)(y^2-z^2)(z^2-x^2)$ | $\Gamma_2^-$ | $xyz$ |
| $\Gamma_3^+$ | $\{2z^2-x^2-y^2, \sqrt{3}(x^2-y^2)\}$ | $\Gamma_3^-$ | $xyz\{2z^2-x^2-y^2, \sqrt{3}(x^2-y^2)\}$ |
| $\Gamma_4^+$ | $\{s_x, s_y, s_x\}$ | $\Gamma_4^-$ | $\{x, y, z\}$ |
| $\Gamma_5^+$ | $\{yz, zx, xy\}$ | $\Gamma_5^-$ | $xyz(x^2-y^2)(y^2-z^2)(z^2-x^2)\{yz, zx, xy\}$ |

Table 1.3: Irreducible representations and representative basis functions of the corresponding vector spaces for the point group $O_h$.

ducible representations. The corresponding wave functions of the eigenstates form a vector space and transform according to the properties of the representation under symmetry operations.

### Δ-line

Now we investigate the evolution of the band energies when $\boldsymbol{k}$ moves away from the Γ-point along the $(0, 0, 1)$ direction in $\boldsymbol{k}$-space. This line is called the Δ-line. There, some of the degeneracies present at the Γ-point will be lifted (cf. Table 1.5) because the symmetry operations leaving $\boldsymbol{k}$ unchanged are now restricted to a subgroup of $O_h$. In the case at hand, this so-called little group of $\boldsymbol{k}$ is isomorphic to $C_{4v}$, which is part of the tetragonal crystal system. Note that the inversion, acting as $\boldsymbol{k} \to -\boldsymbol{k}$, is not an element of the little group. The group $C_{4v}$ has four one-dimensional and one two-dimensional representations. We denote these representations by $\Delta_1, \ldots, \Delta_5$ (cf. Table 1.4). It follows that five bands emanate from the three energy levels at the Γ-point, one of which is two-fold degenerate (Figure 1.4 of).

### X-point

Once we reach the Brillouin zone boundary at the X-point, the symmetry is again larger than on the Δ-line, namely $D_{4h}$, the full tetragonal point group which for each parity has five irreducible representations, four of them are one-dimensional while the remaining one is two-dimensional (cf.

- Landau & Lifschitz: Vol. III Chapt. XII;
- Dresselhaus & Jorio, *Group Theory - Applications to the Physics of Condensed Matter*, Springer;
- Koster et al., *Properties of the thirty-two point groups*, MIT Press (Table book)

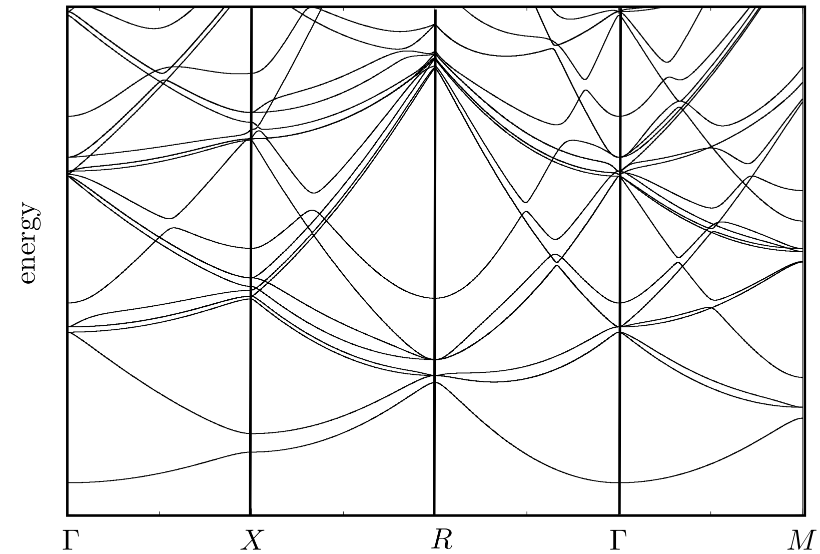



Figure 1.4: The band structure of the simple cubic lattice.

| $\Delta$ | basis function | $d_\Delta$ |
|---|---|---|
| $\Delta_1$ | $1, z$ | 1 |
| $\Delta_2$ | $xy(x^2-y^2)$ | 1 |
| $\Delta_3$ | $x^2-y^2$ | 1 |
| $\Delta_4$ | $xy$ | 1 |
| $\Delta_5$ | $\{x,y\}$ | 2 |

Table 1.4: Irreducible representations of $C_{4v}$ and their basis functions.

| $O_h$ | $C_{4v}$ |
|---|---|
| $\Gamma_1^+$ | $\Delta_1$ |
| $\Gamma_3^+$ | $\Delta_1 \oplus \Delta_3$ |
| $\Gamma_4^-$ | $\Delta_1 \oplus \Delta_5$ |

Table 1.5: Partial lifting of degeneracy leaving the Γ-point along the Δ-line.

Table 1.6). Note that $C_{4v}$ is a subgroup of $D_{4h}$ as well as $D_{4h}$ is a subgroup of $O_h$. Furthermore, the inversion is an element of $D_{4h}$, as for the X-point $\boldsymbol{k}$ is equivalent to $-\boldsymbol{k}$ because $\boldsymbol{k}-(-\boldsymbol{k}) = 2\boldsymbol{k}$ is a reciprocal lattice vector.

The set of states with the lowest energy is equivalent to the problem discussed above in

| even | basis function | $d_X$ | odd | basis function | $d_X$ |
|---|---|---|---|---|---|
| $X_1^+$ | $1$ | 1 | $X_1^-$ | $xyz(x^2-y^2)$ | 1 |
| $X_2^+$ | $xy(x^2-y^2)$ | 1 | $X_2^-$ | $z$ | 1 |
| $X_3^+$ | $x^2-y^2$ | 1 | $X_3^-$ | $xyz$ | 1 |
| $X_4^+$ | $xy$ | 1 | $X_4^-$ | $z(x^2-y^2)$ | 1 |
| $X_5^+$ | $\{zx, zy\}$ | 2 | $X_5^-$ | $\{x,y\}$ | 2 |

Table 1.6: Irreducible representations of $D_{4h}$ and their basis functions.

equations (1.27), (1.28) and (1.31). We consider $\boldsymbol{G}_1 = 0$ and $\boldsymbol{G}_2 = G(0,0,1)$ (remember that $G = 2\pi/a$) with energy $(\hbar^2/2m(G/2)^2)$ at the X-point. The levels are split into an (even) bonding state and an (odd) anti-bonding state

$$\mathrm{X}_1^+ : \; E = \frac{\hbar^2}{2m}\left(\frac{G}{2}\right)^2 - |V_{\boldsymbol{G}_2}|, \qquad\qquad e^{iGz/2}\cos\left(\frac{Gz}{2}\right), \tag{1.46}$$

$$\mathrm{X}_2^- : \; E = \frac{\hbar^2}{2m}\left(\frac{G}{2}\right)^2 + |V_{\boldsymbol{G}_2}|, \qquad\qquad e^{iGz/2}\sin\left(\frac{Gz}{2}\right). \tag{1.47}$$

The next higher energy states are centered around $E = (\hbar^2/2m)(\sqrt{5}G/2)^2$ and belong to the next-to-nearest neighbors of the X-point in the reciprocal lattice. There are eight such points, namely

$$\begin{aligned} &\boldsymbol{G}_1 = G(1,0,0), \quad &&\boldsymbol{G}_2 = G(1,0,1), \quad &&\boldsymbol{G}_3 = G(-1,0,0), \quad &&\boldsymbol{G}_4 = G(-1,0,1), \\ &\boldsymbol{G}_5 = G(0,1,0), \quad &&\boldsymbol{G}_6 = G(0,1,1), \quad &&\boldsymbol{G}_7 = G(0,-1,0), \quad &&\boldsymbol{G}_8 = G(0,-1,1). \end{aligned} \tag{1.48}$$

To find the splitting of the energy levels, we project the base functions in (1.48) onto those of the irreducible representations and find the results displayed in Table 1.7.

This analysis shows that there are six energy levels where two of them are two-fold degenerate.

| X | $u_{\boldsymbol{k}=(G/2)(0,0,1)}(\boldsymbol{r})$ | $d_{\mathrm{X}}$ |
|---|---|---|
| $X_1^+$ | $(\cos(Gx)+\cos(Gy))e^{iGz/2}\cos(Gz/2)$ | 1 |
| $X_3^+$ | $(\cos(Gx)-\cos(Gy))e^{iGz/2}\cos(Gz/2)$ | 1 |
| $X_5^+$ | $\{\sin(Gx)e^{-iGz/2}\sin(Gz/2), \sin(Gy)e^{iGz/2}\sin(Gz/2)\}$ | 2 |
| $X_2^-$ | $(\cos(Gx)+\cos(Gy))e^{iGz/2}\sin(Gz/2)$ | 1 |
| $X_4^-$ | $(\cos(Gx)-\cos(Gy))e^{iGz/2}\sin(Gz/2)$ | 1 |
| $X_5^-$ | $\{\sin(Gx)e^{iGz/2}\cos(Gz/2), \sin(Gy)e^{iGz/2}\cos(Gz/2)\}$ | 2 |

Table 1.7: Projections of the base functions 1.48 onto the ones of the irreducible representations at the X-point ($G = 2\pi/a$).

This kind of analysis can be applied to all symmetry lines, so that a good qualitative picture of the symmetries of the bands can be obtained. For more quantitative information, knowledge of the specific form of the periodic potential is necessary and also more advanced techniques beyond the nearly free electron approach are required. Nevertheless, the nearly free electron method can give important qualitative insights into the symmetry related properties of the band structure (see Figure 1.4 for a full band structure).

## 1.4 $\boldsymbol{k}\cdot\boldsymbol{p}$ - expansion

Near points of high symmetry in the Brillouin zone (such as the Γ-point), energy bands can be approximated by considering a perturbative formulation. For illustration, we expand the Hamiltonian (1.20) around the Γ-point $\boldsymbol{k} = 0$ up to second order in $\boldsymbol{k}$ and split it into the three parts

$$\mathcal{H}_0 = \frac{\widehat{\boldsymbol{p}}^2}{2m} + \widehat{V}, \tag{1.49}$$

$$\mathcal{H}_1 = \frac{\hbar}{m}\widehat{\boldsymbol{\pi}}\cdot\boldsymbol{k}, \tag{1.50}$$

$$\mathcal{H}_2 = \frac{\hbar^2\boldsymbol{k}^2}{2m}, \tag{1.51}$$

where $\widehat{\boldsymbol{\pi}} = \widehat{\boldsymbol{p}}$ in this example. In a more general context (for example considering spin-orbit coupling (1.18)) the operator $\widehat{\boldsymbol{\pi}}$ in equation (1.50) may differ from $\widehat{\boldsymbol{p}}$. We assume that the Hamiltonian $\mathcal{H}_0$ can be solved exactly such that

$$\mathcal{H}_0|n0\rangle = \epsilon_n|n0\rangle, \tag{1.52}$$

where $|n0\rangle$ are states at $\boldsymbol{k} = 0$ with the band index (quantum number) $n$. Compared to $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are small perturbations (small $\boldsymbol{k}$). Note that $\mathcal{H}_2$ is not an operator, but simply a $\boldsymbol{k}$-dependent contribution to the energy. For simplicity, we assume the states $|n0\rangle$ to be non-degenerate, so that Rayleigh-Schrödinger perturbation theory yields the perturbed energy

$$\epsilon_{\boldsymbol{k}} \approx E_{\boldsymbol{k}} = \epsilon_n + \frac{\hbar^2\boldsymbol{k}^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\sum_{\mu,\nu}\frac{\langle n0,|\widehat{\pi}_\mu|n'0\rangle\langle n'0|\widehat{\pi}_\nu|n0\rangle}{\epsilon_n - \epsilon_{n'}}k_\mu k_\nu, \tag{1.53}$$

$$= \epsilon_n + \frac{\hbar^2}{2m}\sum_{\mu,\nu}\left(\frac{m}{m^*}\right)_{\mu\nu}k_\mu k_\nu \tag{1.54}$$

where we introduced a mass-tensor of the form

$$\left(\frac{m}{m^*}\right)_{\mu\nu} = \delta_{\mu\nu} + \frac{2}{m}\sum_{n'\neq n}\frac{\langle n0|\widehat{\pi}_\mu|n'0\rangle\langle n'0|\widehat{\pi}_\nu|n0\rangle}{\epsilon_n - \epsilon_{n'}}. \tag{1.55}$$

Similarly, the eigenstates $|n\boldsymbol{k}\rangle$ can be approximated using the Rayleigh-Schrödinger method, resulting in

$$|n\boldsymbol{k}\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\left\{|n0\rangle + \frac{\hbar}{m}\sum_{n'\neq n}|n'0\rangle\frac{\langle n'0|\widehat{\boldsymbol{\pi}}\cdot\boldsymbol{k}|n0\rangle}{\epsilon_n - \epsilon_{n'}}\right\}. \tag{1.56}$$

Thus, the electronic band structure in the vicinity of the $\Gamma$-point can be expressed by a mass-tensor. If the latter is diagonal and isotropic, we recover a free dispersion relation but with an modified mass $m^*$ instead of $m$. Expressing the mass tensor as

$$\left(\frac{m}{m^*}\right)_{\mu\nu} = \frac{2m}{\hbar^2}\frac{\partial^2\epsilon_{\boldsymbol{k}}}{\partial\boldsymbol{k}_\mu\partial\boldsymbol{k}_\nu}, \tag{1.57}$$

it gives a measure of the curvature of the energy band around a symmetry point. The resulting effective mass $m^*$ can strongly deviate from the bare electron mass $m$ and may even become negative. The $\boldsymbol{k}\cdot\boldsymbol{p}$-approximation is valid for other symmetry points, too. Later, we will find this approximation very convenient when dealing with problems for which states around the upper or lower band edges are important which are often, but not always, high-symmetry points.

Note that, at the band edges, all eigenvalues of the mass tensor have the same sign. There are other symmetry points (usually located at the boundary of the Brillouin zone) where the mass tensor has both positive and negative eigenvalues. These are called saddle points, which play an important role in connection with van Hove singularities in the density of states.

In the derivation we used the property that at symmetry points, the energy shift is only quadratic in $\mathcal{H}_1$, as

$$\langle n0|\widehat{\boldsymbol{\pi}}|n0\rangle = 0. \tag{1.58}$$

This is because of parity and $\widehat{\boldsymbol{\pi}}$ being a rank one tensor operator, as can be easily verified by noting that $\widehat{\boldsymbol{\pi}}\cdot\boldsymbol{k}$ should be a scalar in equation (1.51).[7]

[7] In this case, $\hat{P}\widehat{\boldsymbol{\pi}}\hat{P} = -\widehat{\boldsymbol{\pi}}$ holds for the parity operator $\hat{P}$. But $\hat{P}|n0\rangle = \pm|n0\rangle$ ($|n0\rangle$ is a parity eigenstate whenever the system has inversion symmetry, which carries over to the little group of $\boldsymbol{k} = 0$). Then,

$$\langle n0|\widehat{\boldsymbol{\pi}}|n0\rangle = -\langle n0|\hat{P}\widehat{\boldsymbol{\pi}}\hat{P}|n0\rangle = -\langle n0|\widehat{\boldsymbol{\pi}}|n0\rangle, \tag{1.59}$$

so that the matrix element vanishes.

**Degenerate bands**

If an energy level (say at the Γ-point) is degenerate a special treatment is needed because the previous Rayleigh-Schrödinger approximation assumed non-degenerate levels. As an example, we consider a three-fold degenerate level corresponding to the irreducible representation $\Gamma_4^-$ with $|n_\mu 0\rangle$ ($\mu = x, y, z$) in a cubic crystal with the generating point group $O_h$.
Then, the Rayleigh-Schrödinger perturbation theory leads to the problem of diagonalizing the $3 \times 3$-matrix

$$H_{\mu\nu} = \frac{\hbar^2}{m^2} \sum_{n' \neq n} \frac{\langle n_\mu 0|\widehat{\boldsymbol{\pi}} \cdot \boldsymbol{k}|n'0\rangle\langle n'0|\widehat{\boldsymbol{\pi}} \cdot \boldsymbol{k}|n_\nu 0\rangle}{\epsilon_n - \epsilon_{n'}}, \tag{1.60}$$

where the sum runs over all intermediate states $|n'0\rangle$ not degenerate with $|n_\mu 0\rangle$. The cubic symmetry of the Γ-point leads to the following structure of the matrix:

$$H_{\mu\nu} = (A\boldsymbol{k}^2 + Bk_\mu^2)\delta_{\mu\nu} + Ck_\mu k_\nu \tag{1.61}$$

and the correction to the electronic spectrum can be obtained from the secular equation

$$\det(H_{\mu\nu} - E\delta_{\mu\nu}) = 0 . \tag{1.62}$$

This equation is difficult to solve in general. Nevertheless, for special directions in the Brillouin zone, we obtain simple spectra. First we consider the Δ-line, $\boldsymbol{k} = (0, 0, k_z)$. We find two eigenvalues, leading to

$$\epsilon_{\boldsymbol{k}} = \epsilon_0 + \left(\frac{\hbar^2}{2m} + A\right) k_z^2 + \begin{cases} 0 & \text{two-fold degenerate} \\ \\ (B + C)\, k_z^2 & \text{non-degenerate} \end{cases} \tag{1.63}$$

where the non-degenerate band belong to $\Delta_1$ and the two-fold degenerate branch to $\Delta_5$ of $C_{4v}$. Small deviations from the Δ-line can be treated under certain conditions exactly. For example, for $\boldsymbol{k} = (k_x, 0, k_z)$ we find,

$$\epsilon_{\boldsymbol{k}} = \epsilon_0 + \left(\frac{\hbar^2}{2m} + A\right) \boldsymbol{k}^2 + \begin{cases} (B + C)\boldsymbol{k}^2 - \sqrt{(B + C)^2\boldsymbol{k}^4 - 4B(B + 2C)k_x^2 k_z^2} \\ \\ 0 \\ \\ (B + C)\boldsymbol{k}^2 + \sqrt{(B + C)^2\boldsymbol{k}^4 - 4B(B + 2C)k_x^2 k_z^2} \end{cases} \tag{1.64}$$

which correspond to three non-degenerate branches of the electron bands.

## 1.5 Approximate band calculations

While the approximation of nearly free electrons gives a qualitatively reasonable picture of the band structure, it rests on the assumption that the periodic potential is weak, and thus may be treated as a small perturbation. However, in reality the ionic potential is strong compared to the electrons' kinetic energy. This leads to strong modulations of the wave function around the ions, which is not well described by slightly perturbed plane waves.

### 1.5.1 Pseudo-potential

In order to overcome this weakness of the plane wave solution, we would have to superpose a very large number of plane waves, which is not an easy task to put into practice. Alternatively, we can divide the electronic states into the ones corresponding to filled low-lying energy states,

which are concentrated around the ionic core (core states), and into extended (and more weakly modulated) states, which form the valence and conduction bands. The core electron states may be approximated by atomic orbitals of isolated atoms. For a metal such as aluminum (Al: $1s^2 2s^2 2p^6 3s^2 3p$) the core electrons correspond to the $1s$-, $2s$-, and $2p$-orbitals, whereas the $3s$- and $3p$-orbitals contribute dominantly to the extended states of the valence- and conduction bands. We will focus on the latter, as they determine the low-energy physics of the electrons. The core electrons are deeply bound and can be considered inert.
We introduce the core electron states as $|\phi_j\rangle$, with $\mathcal{H}|\phi_j\rangle = E_j|\phi_j\rangle$ where $\mathcal{H}$ is the Hamiltonian of the single atom. The remaining states have to be orthogonal to these core states, so that we make the Ansatz

$$|\phi_{n,\boldsymbol{k}}\rangle = |\chi_{n\boldsymbol{k}}\rangle - \sum_j |\phi_j\rangle\langle\phi_j|\chi_{n,\boldsymbol{k}}\rangle, \tag{1.65}$$

with $|\chi_{n,\boldsymbol{k}}\rangle$ an orthonormal set of states. Then, $\langle\phi_{n,\boldsymbol{k}}|\phi_j\rangle = 0$ holds for all $j$. If we choose plane waves for the $|\chi_{n\boldsymbol{k}}\rangle$, the resulting $|\phi_{n,\boldsymbol{k}}\rangle$ are so-called orthogonalized plane waves (OPW). The Bloch functions are superpositions of these OPW,

$$|\psi_{n,\boldsymbol{k}}\rangle = \sum_{\boldsymbol{G}} b_{\boldsymbol{k}+\boldsymbol{G}}|\phi_{n,\boldsymbol{k}+\boldsymbol{G}}\rangle, \tag{1.66}$$

where the coefficients $b_{\boldsymbol{k}+\boldsymbol{G}}$ converge rapidly, such that, hopefully, only a small number of OPWs is needed for a good description.
First, we again consider an arbitrary $|\chi_{n\boldsymbol{k}}\rangle$ and insert it into the eigenvalue equation $\mathcal{H}|\phi_{n\boldsymbol{k}}\rangle = E_{n\boldsymbol{k}}|\phi_{n\boldsymbol{k}}\rangle$,

$$\mathcal{H}|\chi_{n\boldsymbol{k}}\rangle - \sum_j \mathcal{H}|\phi_j\rangle\langle\phi_j|\chi_{n,\boldsymbol{k}}\rangle = E_{n\boldsymbol{k}}\Big(|\chi_{n\boldsymbol{k}}\rangle - \sum_j |\phi_j\rangle\langle\phi_j|\chi_{n,\boldsymbol{k}}\rangle\Big) \tag{1.67}$$

or

$$\mathcal{H}|\chi_{n\boldsymbol{k}}\rangle + \sum_j (E_{n\boldsymbol{k}} - E_j)\,|\phi_j\rangle\langle\phi_j|\chi_{n,\boldsymbol{k}}\rangle = E_{n\boldsymbol{k}}|\chi_{n\boldsymbol{k}}\rangle. \tag{1.68}$$

We introduce the integral operator in real space $\widehat{V}' = \sum_j (E_{n\boldsymbol{k}} - E_j)|\phi_j\rangle\langle\phi_j|$, describing a non-local and energy-dependent potential. With this operator we can rewrite the eigenvalue equation in the form

$$(\mathcal{H} + \widehat{V}')|\chi_{n,\boldsymbol{k}}\rangle = (\mathcal{H}_0 + \widehat{V} + \widehat{V}')|\chi_{n,\boldsymbol{k}}\rangle = (\mathcal{H} + \widehat{V}_{\text{ps}})|\chi_{n,\boldsymbol{k}}\rangle = E_{n\boldsymbol{k}}|\chi_{n\boldsymbol{k}}\rangle. \tag{1.69}$$

This is an eigenvalue equation for the so-called pseudo-wave function (or pseudo-state) $|\chi_{n\boldsymbol{k}}\rangle$, instead of the Bloch state $|\psi_{n\boldsymbol{k}}\rangle$, where the modified potential $\widehat{V}_{\text{ps}} = \widehat{V} + \widehat{V}'$ is called *pseudo-potential*. The attractive core potential $\widehat{V} = V(\widehat{\boldsymbol{r}})$ is always negative. On the other hand, $E_{n\boldsymbol{k}} > E_j$, such that $\widehat{V}'$ is positive. It follows that $\widehat{V}_{\text{ps}}$ is weaker than both $\widehat{V}$ and $\widehat{V}'$.
An arbitrary number of core states $\sum_j a_j|\psi_j\rangle$ may be added to $|\chi_{n\boldsymbol{k}}\rangle$ without violating the orthogonality condition (1.65). Consequently, neither the pseudo-potential nor the pseudo-states are uniquely determined and may be optimized variationally with respect to the set $\{a_j\}$ in order to optimally reduce the spatial modulation of either the pseudo-potential or the wave-function.

If we are only interested in states inside a small energy window, the energy dependence of the pseudo-potential can be neglected, and $V_{\text{ps}}$ may be approximated by a standard potential (see Figure 1.5). Such a simple Ansatz is exemplified by the atomic pseudo-potential, proposed by Ashcroft, Heine and Abarenkov (AHA). The potential of a single ion is assumed to be of the form

$$v_{\text{ps}}(r) = \begin{cases} V_0 & r < R_c, \\ -\frac{Z_{\text{ion}}e^2}{r} & r > R_c, \end{cases} \tag{1.70}$$

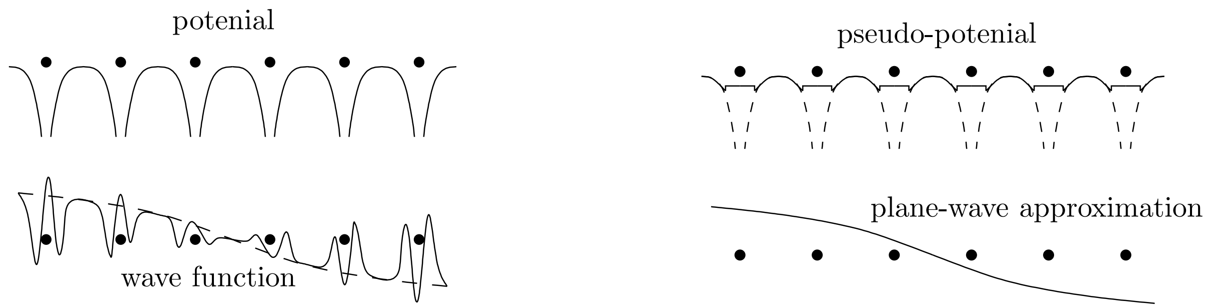



Figure 1.5: Illustration of the pseudo-potential.

where $Z_{\text{ion}}$ is the charge of the ionic core and $R_c$ its effective radius (determined by the core electrons). The constants $R_c$ and $V_0$ are chosen such that the energy levels of the outermost electrons are reproduced correctly for the single-atom calculations. For example, the $1s$-, $2s$-, and $2p$-electrons of Na form the ionic core. $R_c$ and $V_0$ are adjusted such that the one-particle problem $\boldsymbol{p}^2/2m + v_{\text{ps}}(r)$ leads to the correct ionization energy of the $3s$-electron. More flexible approaches allow for the incorporation of more experimental input into the pseudo-potential. The full pseudo-potential of the lattice can be constructed from the contribution of the individual atoms,

$$V_{\text{ps}}(\boldsymbol{r}) = \sum_n v_{\text{ps}}(|\boldsymbol{r} - \boldsymbol{R}_n|), \tag{1.71}$$

where $\boldsymbol{R}_n$ is the lattice vector. For the method of nearly free electrons we need the Fourier transform of the potential evaluated at the reciprocal lattice vectors,

$$V_{\text{ps},\boldsymbol{G}} = \frac{1}{\Omega}\int d^3r\; V_{\text{ps}}(\boldsymbol{r})e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} = \frac{N}{\Omega}\int d^3r\; v_{\text{ps}}(\boldsymbol{r})e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}. \tag{1.72}$$

For the AHA form (1.70), this is given by

$$V_{\text{ps},\boldsymbol{G}} = -\frac{4\pi Z_{\text{ion}}e^2}{G^2}\Bigg(\cos(GR_c) + \frac{V_0}{Z_{\text{ion}}e^2 G}\left((R_c^2G^2-2)\cos(GR_c) + 2 - 2R_cG\sin(GR_c)\right)\Bigg). \tag{1.73}$$

For small reciprocal lattice vectors, the zeroes of the trigonometric functions on the right-hand side of (1.73) reduce the strength of the potential. For large $G$, the pseudo-potential decays $\propto 1/G^2$. It is thus clear that the pseudo-potential is always weaker than the original potential. Extending this theory for complex unit cells containing more than one atom, the pseudo-potential may be written as

$$V_{\text{ps}}(\boldsymbol{r}) = \sum_{n,\alpha} v_{\text{ps}}^{\alpha}(|\boldsymbol{r} - (\boldsymbol{R}_n + \boldsymbol{R}_\alpha)|), \tag{1.74}$$

where $\boldsymbol{R}_\alpha$ denotes the position of the $\alpha$-th base atom in the unit cell. Here, $v_{\text{ps}}^{\alpha}$ is the pseudo-potential of the $\alpha$-th ion. In reciprocal space,

$$V_{\text{ps},\boldsymbol{G}} = \frac{N}{\Omega}\sum_\alpha e^{-i\boldsymbol{G}\cdot\boldsymbol{R}_\alpha}\int d^3r\; v_{\text{ps}}^{\alpha}(|\boldsymbol{r}|)e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \tag{1.75}$$

$$= \sum_\alpha e^{-i\boldsymbol{G}\cdot\boldsymbol{R}_\alpha}F_{\alpha,\boldsymbol{G}}. \tag{1.76}$$

The form factor $F_{\alpha,\boldsymbol{G}}$ contains the information of the base atoms and may be calculated or obtained by fitting experimental data.

### 1.5.2 Augmented plane wave

We now consider a method introduced by Slater in 1937. It is an extension of the so-called Wigner-Seitz cell method (1933) and consists of approximating the crystal potential by a so-called muffin-tin potential. Latter is a periodic potential, which is taken to be spherically symmetric and position dependent around each atom up to a distance $r_s$, and constant for larger distances. The spheres of radius $r_s$ are taken to be non-overlapping and are contained completely in the Wigner-Seitz cell[8] (Figure 1.6).

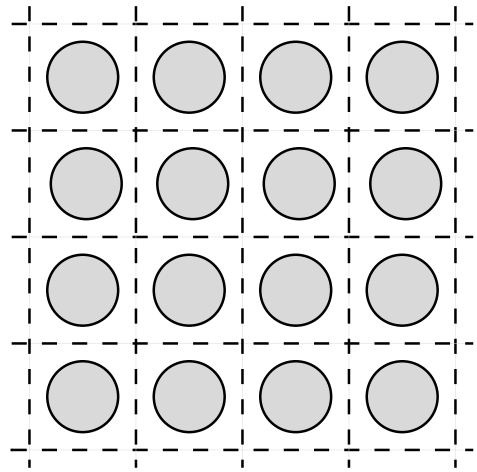


Figure 1.6: Muffin-tin potential.

The advantage of this decomposition is that the problem can be solved using a divide-and-conquer strategy. Inside the muffin-tin radius we solve the spherically symmetric problem, while the solutions on the outside are given by plane waves; the remaining task is to match the solutions at the boundaries.
The spherically symmetric problem for $|\boldsymbol{r}| < r_s$ is solved with the standard Ansatz

$$\varphi(\boldsymbol{r}) = \frac{u_l(r)}{r} Y_{lm}(\theta, \phi), \tag{1.77}$$

where $(r, \theta, \phi)$ are the spherical coordinates or $\boldsymbol{r}$ and the radial part $u_l(r)$ of the wave function obeys the differential equation

$$\left[ -\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) - E \right] u_l(r, E) = 0. \tag{1.78}$$

We define an augmented plane wave (APW) $A(\boldsymbol{k}, \boldsymbol{r}, E)$, which is a pure plane wave with wave vector $\boldsymbol{k}$ outside the Muffin-tin sphere. For this, we employ the representation of plane waves by spherical harmonics,

$$e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(kr) Y^*_{lm}(\hat{\boldsymbol{k}}) Y_{lm}(\hat{\boldsymbol{r}}), \tag{1.79}$$

[8]The Wigner-Seitz cell is the analogue of the Brillouin zone in real space. One draws planes cutting each line joining two atoms in the middle, and orthogonal to them. The smallest cell bounded by these planes is the Wigner-Seitz cell.

where $j_l(x)$ is the $l$-th spherical Bessel function. We parametrize

$$A(\boldsymbol{k},\boldsymbol{r},E) = \begin{cases} \dfrac{4\pi}{\sqrt{\Omega_{\mathrm{UC}}}} \displaystyle\sum_{l,m} i^l j_l(kr_s) \frac{r_s u_l(r,E)}{r u_l(r_s,E)} Y^*_{lm}(\hat{\boldsymbol{k}}) Y_{lm}(\hat{\boldsymbol{r}}), & r < r_s, \\ \\ \dfrac{4\pi}{\sqrt{\Omega_{\mathrm{UC}}}} \displaystyle\sum_{l,m} i^l j_l(kr) Y^*_{lm}(\hat{\boldsymbol{k}}) Y_{lm}(\hat{\boldsymbol{r}}), & r > r_s, \end{cases} \tag{1.80}$$

where $\Omega_{\mathrm{UC}}$ is the volume of the unit cell. Note that the wave function is always continuous at $r = r_s$, but that its derivatives are in general not continuous. We can use an expansion of the wave function $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ similar to the one in the nearly free electron approximation (see equations (1.9) and (1.24)),

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(\boldsymbol{k}) A(\boldsymbol{k}+\boldsymbol{G},\boldsymbol{r},E), \tag{1.81}$$

where the $\boldsymbol{G}$ are reciprocal lattice vectors. The unknown coefficients can be determined variationally by solving the system of equations

$$\sum_{\boldsymbol{G}} \langle A_{\boldsymbol{k}}(E)|\mathcal{H}-E|A_{\boldsymbol{k}+\boldsymbol{G}}(E)\rangle c_{\boldsymbol{G}}(\boldsymbol{k}) = 0, \tag{1.82}$$

where

$$\langle A_{\boldsymbol{k}}(E)|\mathcal{H}-E|A_{\boldsymbol{k}'}(E)\rangle = \left(\frac{\hbar^2 \boldsymbol{k}\cdot\boldsymbol{k}'}{2m} - E\right) \Omega_{\mathrm{UC}} \delta_{\boldsymbol{k},\boldsymbol{k}'} + V_{\boldsymbol{k},\boldsymbol{k}'} \tag{1.83}$$

with

$$\begin{aligned} V_{\boldsymbol{k},\boldsymbol{k}'} = 4\pi r_s^2 \Bigg[ & -\left(\frac{\hbar^2 \boldsymbol{k}\cdot\boldsymbol{k}'}{2m} - E\right) \frac{j_1(|\boldsymbol{k}-\boldsymbol{k}'|r_s)}{|\boldsymbol{k}-\boldsymbol{k}'|} \\ & + \sum_{l=0}^{\infty} \frac{\hbar^2}{2m} (2l+1) P_l(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}') j_l(kr_s) j_l(k'r_s) \frac{u_l'(r_s,E)}{u_l(r_s,E)} \Bigg]. \end{aligned} \tag{1.84}$$

Here, $P_l(z)$ is the $l$-th Legendre polynomial and $u' = du/dr$. The solution of (1.82) yields the energy bands. The most difficult parts are the approximation of the crystal potential by the muffin-tin potential and the computation of the matrix elements in (1.82). The rapid convergence of the method is its big advantage: just a few dozens of $\boldsymbol{G}$-vectors are needed and the largest angular momentum needed is about $l = 5$. Another positive aspect is the fact that the APW-method allows the interpolation between the two extremes of extended, weakly bound electronic states and tightly bound states.

## 1.6 Tightly bound electrons

If the electrons in the valence and conduction bands are strongly bound to the ions, another very efficient approximation to the band structure exists. In this case, it is easier to approach the problem in real space instead of reciprocal space. This leads to the so-called tight-binding model. We introduce the Wannier functions as 'Fourier transforms' of the Bloch functions,

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_j e^{i\boldsymbol{k}\cdot\boldsymbol{R}_j} w(\boldsymbol{r}-\boldsymbol{R}_j), \tag{1.85}$$

where $w(\boldsymbol{r}-\boldsymbol{R}_j)$ is the Wannier function centered around the $j$-th atom. There is a Wannier function for each atomic orbital. For the sake of simplicity, we restrict ourselves to the case of one orbital per atom. The Wannier function obeys the orthogonality relation

$$\int d^3r\ w^*(\boldsymbol{r}-\boldsymbol{R}_j)w(\boldsymbol{r}-\boldsymbol{R}_l)=\delta_{jl}. \tag{1.86}$$

We assume the one-particle Hamiltonian to be of the form $\mathcal{H}=-\hbar^2\boldsymbol{\nabla}^2/2m+V(\boldsymbol{r})$, with a periodic potential $V(\boldsymbol{r})$. Then, $\epsilon_{\boldsymbol{k}}$ can be expressed through

$$\epsilon_{\boldsymbol{k}}=\int d^3r\ \psi^*_{\boldsymbol{k}}(\boldsymbol{r})\mathcal{H}\psi_{\boldsymbol{k}}(\boldsymbol{r})=\frac{1}{N}\sum_{j,l}e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}_j-\boldsymbol{R}_l)}\int d^3r\ w^*(\boldsymbol{r}-\boldsymbol{R}_j)\mathcal{H}w(\boldsymbol{r}-\boldsymbol{R}_l). \tag{1.87}$$

With the definitions

$$\epsilon_0=\int d^3r\ w^*(\boldsymbol{r}-\boldsymbol{R}_j)\mathcal{H}w(\boldsymbol{r}-\boldsymbol{R}_j), \tag{1.88}$$

$$t_{jl}=\int d^3r\ w^*(\boldsymbol{r}-\boldsymbol{R}_j)\mathcal{H}w(\boldsymbol{r}-\boldsymbol{R}_l)\quad \text{for}\quad j\neq l \tag{1.89}$$

we immediately find, that the band energy can be written as a discrete sum,

$$\epsilon_{\boldsymbol{k}}=\epsilon_0+\frac{1}{N}\sum_{j,l}t_{jl}e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}_j-\boldsymbol{R}_l)}=\epsilon_0+\sum_{l}t_{0l}e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l}, \tag{1.90}$$

where $\boldsymbol{R}_0=0$ is assumed without loss of generality. It is straightforward to show that $\epsilon_{\boldsymbol{k}+\boldsymbol{G}}=\epsilon_{\boldsymbol{k}}$. The quantities $t_{jl}$ are called *hopping matrix elements*. It is possible to construct an effective Hamiltonian based on the above findings, which describes the band structure of independent electrons, as

$$\mathcal{H}=\sum_{i,j}\sum_{s}t_{ij}c^\dagger_{is}c_{js}, \tag{1.91}$$

where $c_{is}$ ($c^\dagger_{is}$) annihilates (creates) an electron with spin $s$ at the lattice point $i$. The Hamiltonian describes the hopping of electrons from site $j$ to site $i$. This formulation is advantageous, if the hopping matrix elements fall off rapidly with the distance between the lattice points. This should be the case for electronic states which are tightly bound to the ions.
Consider a simple cubic lattice, assuming that $t_{jl}=-t$ for nearest neighbors and zero otherwise. The band energy follows from a Fourier transform and is given by

$$\epsilon_{\boldsymbol{k}}=\epsilon_0-2t\big(\cos(k_xa)+\cos(k_ya)+\cos(k_za)\big), \tag{1.92}$$

where $a$ is the lattice constant. The same can be applied to more complicated lattices and systems with several relevant orbitals per atom.

## 1.7 Semi-classical description of band electrons

In quantum mechanics, the Ehrenfest theorem shows that the expectation values of the position and momentum operators obey equations similar to the equation of motion in Newtonian mechanics. An analogous formulation holds for electrons in a periodic potential, where we assume that the electron may be described as a wave packet of the form

$$\psi_{\boldsymbol{k}}(\boldsymbol{r},t)=\sum_{\boldsymbol{k}'}g_{\boldsymbol{k}}(\boldsymbol{k}')e^{i\boldsymbol{k}'\cdot\boldsymbol{r}-i\epsilon_{\boldsymbol{k}'}t}, \tag{1.93}$$

where $g_{\boldsymbol{k}}(\boldsymbol{k}')$ is centered around $\boldsymbol{k}$ in reciprocal space and has a width of $\Delta k$. $\Delta k$ should be much smaller than the size of the Brillouin zone for this Ansatz to make sense, i.e., $\Delta k \ll 2\pi/a$. Therefore, the wave packet is spread over many unit cells of the lattice since Heisenberg's uncertainty principle $(\Delta k)(\Delta x) > 1$ implies $\Delta x \gg a/2\pi$. In this way, the pseudo-momentum $\boldsymbol{k}$ of the wave packet remains well defined. Furthermore, the applied electric and magnetic fields have to be small enough not to induce transitions between different bands. The latter condition is not very restrictive in practice.

### 1.7.1 Semi-classical equations of motion

We introduce the rules of the semi-classical motion of electrons with applied electric and magnetic fields without proof:

- The band index of an electron is conserved, i.e., there are no transitions between the bands.
- The equations of motion read[9]

$$\dot{\boldsymbol{r}} = \boldsymbol{v}_n(\boldsymbol{k}) = \frac{\partial \epsilon_{n\boldsymbol{k}}}{\partial \hbar \boldsymbol{k}}, \tag{1.96}$$

$$\hbar \dot{\boldsymbol{k}} = -e\boldsymbol{E}(\boldsymbol{r},t) - \frac{e}{c}\boldsymbol{v}_n(\boldsymbol{k}) \times \boldsymbol{H}(\boldsymbol{r},t). \tag{1.97}$$

- All electronic states have a wave vector that lies in the first Brillouin zone, as $\boldsymbol{k}$ and $\boldsymbol{k}+\boldsymbol{G}$ label the same state for all reciprocal lattice vectors $\boldsymbol{G}$.
- In thermal equilibrium, the electron density per spin in the $n$-th band in the volume element $d^3k/(2\pi)^3$ around $\boldsymbol{k}$ is given by

$$n_F[\epsilon_n(\boldsymbol{k})] = \frac{1}{e^{[\epsilon_n(\boldsymbol{k})-\mu]/k_BT} + 1}, \tag{1.98}$$

  Each state of given $\boldsymbol{k}$ and spin can be occupied only once (Pauli principle).

Note that $\hbar\boldsymbol{k}$ is not the momentum of the electron, but the so-called lattice momentum or pseudo momentum in the Bloch theory of bands. It is connected with the eigenvalue of the translation operator on the state. Consequently, the right-hand side of the equation (1.97) is not the force that acts on the electron, as the forces exerted by the periodic lattice potential is not included. The latter effect is contained implicitly through the form of the band energy $\epsilon(\boldsymbol{k})$, which governs the first equation.

#### Bloch oscillations

The fact that the band energy is a periodic function of $\boldsymbol{k}$ leads to a strange oscillatory behavior of the electron motion in a static electric field. For illustration, consider a one-dimensional system

[9] A plausibility argument concerning the conservation of energy leading to the equation (1.97) is given here. The time derivative of the energy (kinetic and potential)

$$E = \epsilon_n(\boldsymbol{k}(t)) - e\phi(\boldsymbol{r}(t)) \tag{1.94}$$

has to vanish, i.e.,

$$0 = \frac{dE}{dt} = \frac{\partial \epsilon_n(\boldsymbol{k})}{\partial \boldsymbol{k}} \cdot \dot{\boldsymbol{k}} - e\boldsymbol{\nabla}\phi \cdot \dot{\boldsymbol{r}} = \boldsymbol{v}_n(\boldsymbol{k}) \cdot \left(\hbar\dot{\boldsymbol{k}} - e\boldsymbol{\nabla}\phi\right). \tag{1.95}$$

From this, equation (1.97) follows directly for the electric field $\boldsymbol{E} = \boldsymbol{\nabla}\phi$ and the Lorentz force is allowed because the force is always perpendicular to the velocity $\boldsymbol{v}_n$.

where the band energy $\epsilon_k = -2\cos ka$ leads to the solution of the semi-classical equations (1.96,1.97)

$$\hbar\dot{k} = -eE \tag{1.99}$$

$$\Rightarrow \qquad k = -\frac{eEt}{\hbar} \tag{1.100}$$

$$\Rightarrow \qquad \dot{x} = -\frac{2a}{\hbar}\sin\left(\frac{eEat}{\hbar}\right), \tag{1.101}$$

in the presence of a homogenous electric field $E$. It follows immediately, that the position $x$ of the electron oscillates like

$$x(t) = \frac{1}{eE}\cos\left(\frac{eEat}{\hbar}\right). \tag{1.102}$$

This behavior is called *Bloch oscillation* and means that the electron oscillates around its initial position rather than moving in one direction when subjected to a static electric field. This effect can only be observed under very special conditions where the probe is absolutely clean. The effect is easily destroyed by damping or scattering.

### 1.7.2 Current densities

We will see in chapter 6 that homogenous steady current carrying states of electron systems can be described by the momentum distribution $n(\boldsymbol{k})$. Assuming this property, the current density follows from

$$\boldsymbol{j} = -2e\int_{\text{BZ}}\frac{d^3k}{(2\pi)^3}\boldsymbol{v}(\boldsymbol{k})n(\boldsymbol{k}) = -2e\int_{\text{BZ}}\frac{d^3k}{(2\pi)^3}n(\boldsymbol{k})\frac{\partial\epsilon(\boldsymbol{k})}{\partial\hbar\boldsymbol{k}}, \tag{1.103}$$

where the integral extends over all $\boldsymbol{k}$ in the Brillouin zone (BZ) and the factor 2 originates from the two possible spin states of the electrons. Note that for a finite current density $\boldsymbol{j}$, the momentum distribution $n(\boldsymbol{k})$ has to deviate from the equilibrium Fermi-Dirac distribution in equation (1.98). It is straight forward to show that the current density vanishes for an empty band. The same holds true for a completely filled band ($n(\boldsymbol{k}) = 1$) where equation (1.96) implies

$$\boldsymbol{j} = -2e\int_{\text{BZ}}\frac{d^3k}{(2\pi)^3}\frac{1}{\hbar}\frac{\partial\epsilon(\boldsymbol{k})}{\partial\boldsymbol{k}} = 0 \tag{1.104}$$

because $\epsilon(\boldsymbol{k})$ is periodic in the Brillouin zone, i.e., $\epsilon(\boldsymbol{k}+\boldsymbol{G}) = \epsilon(\boldsymbol{k})$ when $\boldsymbol{G}$ is a reciprocal lattice vector. Thus, neither empty nor completely filled bands can carry currents.
An interesting aspect of band theory is the picture of holes. We compute the current density for a partially filled band in the framework of the semi-classical approximation,

$$\boldsymbol{j} = -e\int_{\text{BZ}}\frac{d^3k}{4\pi^3}n(\boldsymbol{k})\boldsymbol{v}_n(\boldsymbol{k}) \tag{1.105}$$

$$= -e\Big[\int_{\text{BZ}}\frac{d^3k}{4\pi^3}\boldsymbol{v}(\boldsymbol{k}) - \int_{\text{BZ}}\frac{d^3k}{4\pi^3}[1-n(\boldsymbol{k})]\boldsymbol{v}(\boldsymbol{k})\Big] \tag{1.106}$$

$$= +e\int_{\text{BZ}}\frac{d^3k}{4\pi^3}[1-n(\boldsymbol{k})]\boldsymbol{v}(\boldsymbol{k}). \tag{1.107}$$

This suggests that the current density comes either from electrons in filled states with charge $-e$ or from 'holes', missing electrons carrying positive charge and sitting in the unoccupied electronic states. In band theory, both descriptions are equivalent. However, it is usually easier to work with holes if a band is almost filled, and with electrons if the filling of an energy band is small.

## 1.8 Metals and semiconductors

Each state $|\psi_{n,\boldsymbol{k}}\rangle$ can be occupied by two electrons, one with spin state $|\uparrow\rangle$ and $|\downarrow\rangle$. In the ground state, all lowest states up to a Fermi energy $E_F$ are filled. The nature of the ground state of electrons in a solid depends on the number of electrons per atom. Usually, this number is an integer, so that in the simplest cases we distinguish only two different situations: First, the bands can be either completely filled or empty when the number of electrons per atom is even. In this case, the Fermi energy lies in a band gap (cf. Figure 1.7), and a finite energy is needed to add, to remove or to excite electrons. If the band gap $E_g$ is much smaller than the bandwidth, we call the material a *semiconductor*. for $E_g$ of the order of the bandwidth, it is an *(band) insulator*. In both cases, for temperatures $T \ll E_g/k_B$ the application of a small electric voltage will not produce an electric transport. The highest filled band is called *valence* band, whereas the lowest empty band is termed *conduction* band. Note that we will encounter another form of an insulator, the Mott insulator, whose insulating behavior is not governed by a band structure effect, but by a correlation effect through strong Coulomb interaction.
Secondly, if the number of electrons per atom is odd, the uppermost non-empty band is half filled (see Figure 1.7). Then the system is a *metal*, as charges can move without overcoming a band gap and electrons can be excited with arbitrarily small energy. The electrons remain mobile down to arbitrarily low temperatures. The standard example of a metal are the Alkali metals in the first column of the periodic table (Li, Na, K, Rb, Cs), as all of them have the configuration [noble gas] $(\mathrm{ns})^1$, i.e., one mobile electron per ion.

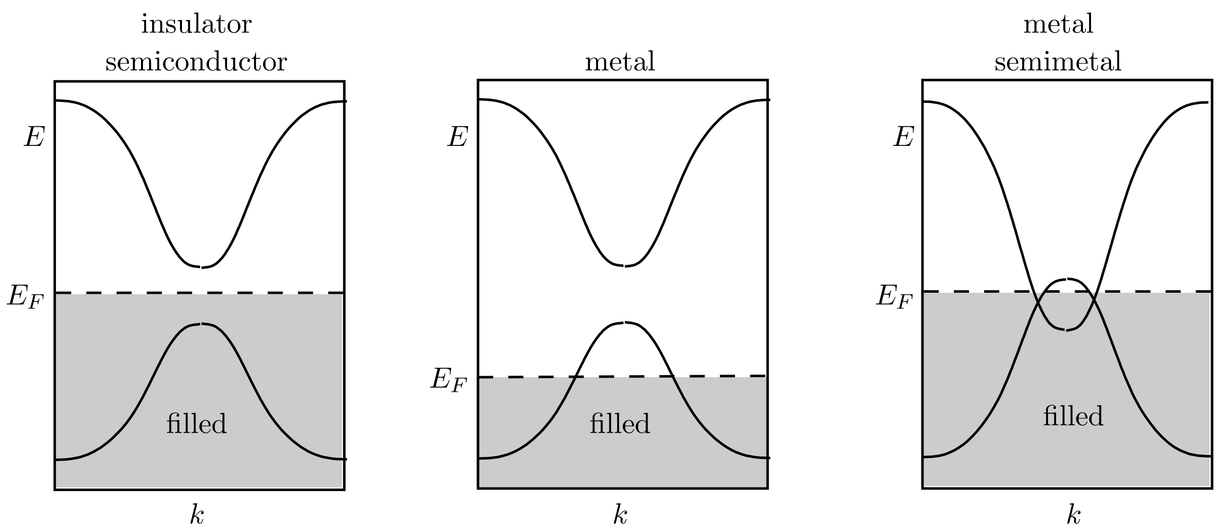



Figure 1.7: Material classes according to band filling: left panel: insulator or semiconductor (Fermi level in band gap); center panel: metal (Fermi level inside band); right panel: metal or semimetal (Fermi level inside two overlapping bands).

In general, band structures are more complex. Different bands need not to be separated by energy gaps, but can overlap instead. In particular, this happens if different orbitals are involved in the structure of the bands. In these systems, bands can have any fractional filling (not just filled or half-filled). The earth alkaline metals are an example for this (second column of the periodic table, Be, Mg, Ca, Sr, Ba), which are metallic despite having two $(n, s)$-electrons per unit cell. Systems, where two bands overlap at the Fermi energy but the overlap is small, are termed *semi-metals*. The extreme case, where valence and conduction band touch in isolated points so that there are no electrons at the Fermi energy and still the band gap is zero, is realized in graphite.

# Chapter 2

# Semiconductors

The technological relevance of semiconductors can hardly be overstated. In this chapter, we review some of their basic properties. Regarding the electric conductivity, semiconductors are placed in between metals and insulators. Normal metals are good conductors at all temperatures, and the conductivity usually increases with decreasing temperature. On the other hand, for semiconductors and insulators the conductivity decreases upon cooling (see Figure 2.1).

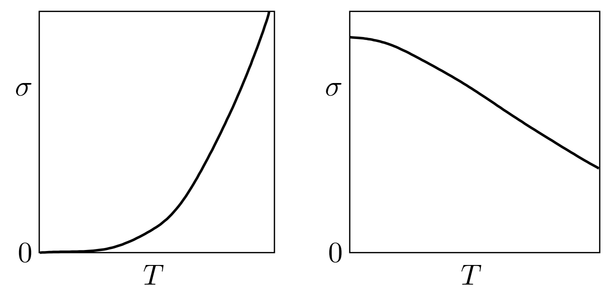



Figure 2.1: Schematic temperature dependence of the electric conductivity for semiconductors and metals.

We will see that the conductivity may be written as

$$\sigma = \frac{ne^2\tau}{m}, \tag{2.1}$$

where $n$ is the density of mobile electrons, $\tau$ is the average time between two scattering events of the electrons, and $m$ and $e$ are the electronic mass and charge, respectively. In metals, $n$ is independent of temperature, whereas the scattering time $\tau$ decreases with increasing temperature. Thus, $\tau$ determines majorly the temperature dependence of the conductivity in metals. On the other hand, insulators and semiconductors have no mobile charges at $T = 0$. At finite temperature, charges are induced by thermal excitations which have to overcome the band gap[1] $E_g$ between the valence and the conduction band, yielding

$$n \sim n_0 \left(\frac{T}{T_0}\right)^{3/2} e^{-E_g/2k_BT}, \tag{2.3}$$

[1] Actually, one has to count both the excited electrons in the conduction band and the resulting holes in the valence band, as both contribute to the current,

$$\boldsymbol{j} = (\sigma_+ + \sigma_-)\boldsymbol{E}, \quad \text{with} \quad \sigma_\pm = \frac{n_\pm e^2\tau_\pm}{m_\pm}, \tag{2.2}$$

where $+$ and $-$ stand for holes and electrons, respectively, and $n_+ = n_-$ holds for thermal excitation. Note that, in general, the effective masses and scattering times are not the same for the valence and conduction bands.

where $T_0 = 300\text{K}$ and the electron density in the material $n_0$ is typically $10^{20}$ cm$^{-3}$. For insulators, the energy gap is huge, e.g., 5.5 eV for diamond. Consequently, the charge carrier density at room temperature $T = 300\text{K}$ is around $n \sim 10^{-27}\text{cm}^{-3}$. For a higher charge carrier density $n \sim 10^3 - 10^{11}\text{cm}^{-3}$, smaller gaps $E_g \sim 0.5 - 1\text{eV}$ are necessary. Materials with a band gap in this regime are not fully isolating and therefore are termed semiconductors. However, the carrier densities of both insulators and semiconductors are dwarfed by the electron density in metals contributing to current transport ($n_{\text{metal}} \sim 10^{23} - 10^{24}\text{cm}^{-3}$). Adding a small amount of impurities in semiconductors, a process called doping with acceptors or donators, their conductivity can be engineered in various ways, rendering them useful as components in innumerable applications.

## 2.1 The band structure of the elements in group IV

### 2.1.1 Crystal and band structure

The most important semiconductor for technological applications is silicon (Si) that – like carbon (C), germanium (Ge) and tin (Sn) – belongs to the group IV of the periodic table. These elements have four electrons in their outermost shell in the configuration $(\text{ns})^2(\text{np})^2$ (n=2 for C, n=3 for Si, n=4 for Ge, and n=5 for Sn). All four elements form crystals with a diamond structure (cf. Figure 2.2), i.e., a face-centered cubic lattice with a unit cell containing two atoms located at $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ (for Sn this is called $\alpha$-Sn). The crystal structure is stabilized by hybridization of the four valence electrons, leading to covalent bonding of oriented orbitals,

$$\begin{aligned} |\psi_1\rangle &= |\text{ns}\rangle + |\text{np}_x\rangle + |\text{np}_y\rangle + |\text{np}_z\rangle, & |\psi_2\rangle &= |\text{ns}\rangle + |\text{np}_x\rangle - |\text{np}_y\rangle - |\text{np}_z\rangle, \\ |\psi_3\rangle &= |\text{ns}\rangle - |\text{np}_x\rangle + |\text{np}_y\rangle - |\text{np}_z\rangle, & |\psi_4\rangle &= |\text{ns}\rangle - |\text{np}_x\rangle - |\text{np}_y\rangle + |\text{np}_z\rangle. \end{aligned} \tag{2.4}$$

Locally, the nearest neighbors of each atom form a tetrahedron around it, which leads to the diamond structure of the lattice.

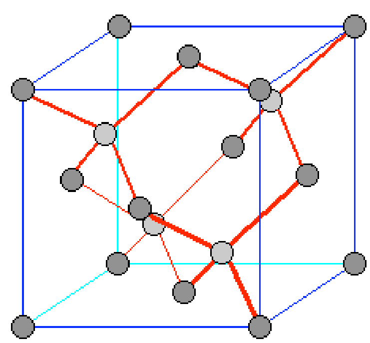


Figure 2.2: The crystal structure of diamond corresponds to two face-centered cubic latices shifted by a quarter of lattice spacing along the (1,1,1) direction.

The band structure of these materials can be found around the $\Gamma$-point by applying the free-electron approximation discussed in the previous chapter. $\Gamma_1$ is the corresponding representation to the parabolic band centered around the center of the Brillouin zone $(0, 0, 0)$. Note that the reciprocal lattice of a face-centered cubic (fcc) lattice is body-centered cubic (bcc). The Brillouin zone of a fcc crystal is embedded in a bcc lattice as illustrated in Figure 2.3).
The next multiplet with an energy of $\epsilon = 6\pi^2\hbar^2/ma^2$ derives from the parabolic bands emanating from the neighboring Brillouin zones, with $\boldsymbol{G} = G(\pm1, \pm1, \pm1)$ and $G = 2\pi/a$. The eight states

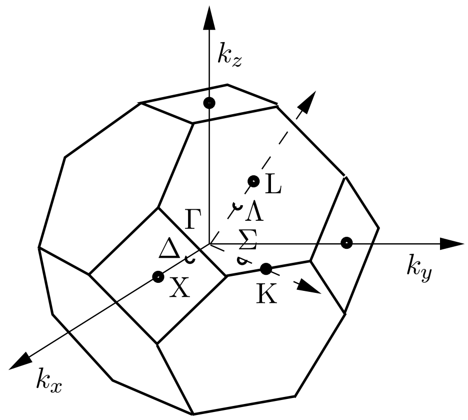



Figure 2.3: The Brillouin zone of a face-centered cubic crystal (in real space) is embedded in a bcc lattice.

are split into $\Gamma_1 \oplus \Gamma_2 \oplus \Gamma_4 \oplus \Gamma_5$. The magnitude of the resulting energies are obtained from band structure calculations, lifting the energy degeneracy $\epsilon_{\Gamma_5} < \epsilon_{\Gamma_4} < \epsilon_{\Gamma_2} < \epsilon_{\Gamma_1}$ in the presence of a periodic potential $V(\boldsymbol{r})$. The essential behavior of the low-energy band structure of C and Si is shown in Figure 2.4.

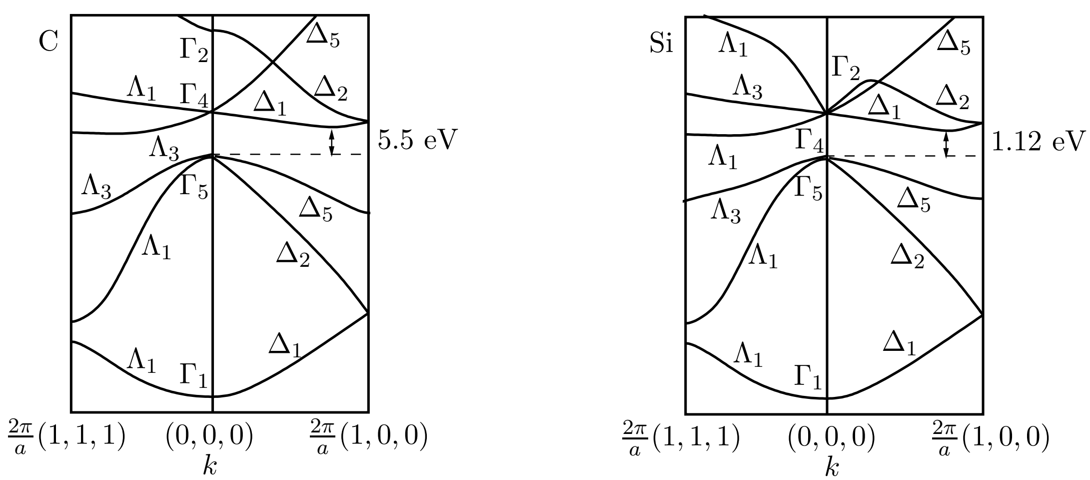



Figure 2.4: Band structure of C and Si.

With two atoms per unit cell and four valence electrons, there are totally eight electrons per unit cell. It follows that the bands belonging to $\Gamma_1$ (non-degenerate) and $\Gamma_5$ (threefold degenerate) are completely filled. The maximum of the valence band is located at the Γ-point and belongs to the representation $\Gamma_5$. Because of the existence of an energy gap between valence and conduction band, the system is not conducting at $T = 0$. The gap is *indirect*, meaning that the minimum of the conduction band and the maximum of the valence band lie at different points in the Brillouin zone, i.e., the gap is minimal between the Γ-point of the valence band and some finite momentum $\hbar\boldsymbol{k}_0$ along the [100]-direction of the conduction band.[2] A typical example of a semiconductor with a direct gap is GaAs, composed of one element of the third and fifth group, respectively. Next we list some facts about these materials of the group IV:

- Carbon has an energy gap of around 5.5eV in the diamond structure. Thus, in this

[2]Energy gaps in semiconductors and insulators are said to be *direct* if the wave-vector connecting the maximum of the valence band and the minimum of the conduction band vanishes. Otherwise a gap is called *indirect* (see Figure 2.5).

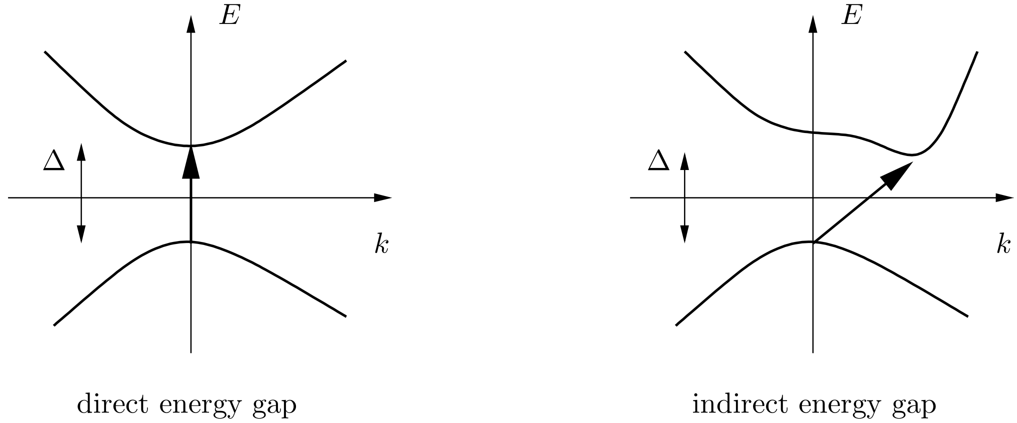



Figure 2.5: Illustration of direct and indirect band gaps.

configuration it is not a semiconductor but an insulator. The large energy gap causes the transparency of diamond in the visible range (1.5-3.5eV), as the electromagnetic energy in this range cannot be absorbed by the electrons.

- The energy gap of silicon is 1.12eV and thus much smaller; furthermore, it is indirect.
- Germanium has an indirect gap of 0.67eV.

### 2.1.2 $\boldsymbol{k}\cdot\boldsymbol{p}$ - expansion

The band structure in the vicinity of the band edges can be very well described using the $\boldsymbol{k}\cdot\boldsymbol{p}$ method, as we show now for silicon. First, we consider the maximum of the valence band at $\boldsymbol{k}=0$ ($\Gamma$-point), with electronic states

$$\{\, |yz\rangle, |zx\rangle, |xy\rangle \,\} \tag{2.5}$$

belonging to the representation $\Gamma_5^+$ (see Table 1.3). By symmetry considerations we obtain, in second order perturbation theory, the secular equation for the degenerate subspace,

$$\det\begin{bmatrix} ak_x^2+b(k_y^2+k_z^2)-E & ck_xk_y & ck_xk_z \\ ck_xk_y & ak_y^2+b(k_x^2+k_z^2)-E & ck_yk_z \\ ck_xk_z & ck_yk_z & ak_z^2+b(k_x^2+k_y^2)-E \end{bmatrix}=0. \tag{2.6}$$

The general form of the eigenvalues is complicated, but it can be shown that the threefold degeneracy of the energies is lifted when moving away from the $\Gamma$-point along the high-symmetry lines $\Delta$ and $\Lambda$. On the $\Delta$- ($C_{4v}$) and $\Lambda$-lines ($C_{3v}$) one twofold degenerate and one non-degenerate band emerge (cf. Figure 2.4):

$$\Delta\text{-line } (\boldsymbol{k}=k(1,0,0)) \qquad E_1(\Delta_2)=ak^2, \qquad E_{2,3}(\Delta_5)=bk^2, \tag{2.7}$$

$$\Lambda\text{-line } \left(\boldsymbol{k}=\frac{k}{\sqrt{3}}(1,1,1)\right) \qquad E_1(\Lambda_1)=(a+2b+2c)k^2, \qquad E_{2,3}(\Lambda_3)=(a+2b-c)k^2, \tag{2.8}$$

where $\Delta_i$ and $\Lambda_i$ are irreducible representations of the point group $C_{4v}$ and $C_{3v}$, respectively.[3]

On the other hand, the minimum of the conduction band is located on the $\Delta$-line at $\boldsymbol{k}_0 =$

[3] Spin-orbit coupling has been neglected so far. Including the spin degrees of freedom would lead to a splitting of the energies at $k=0$ into a two-fold degenerate level ($\Gamma_6^+$) and a four-fold degenerate one ($\Gamma_8^+$).

$k_0(1,0,0)$ with $k_0 \approx 0.8\overline{\Gamma X}$. Apart from spin, the corresponding band is non-degenerate. It follows that the $\boldsymbol{k}\cdot\boldsymbol{p}$-approximation is given by

$$E_{\boldsymbol{k}} = a'(k_x - k_0)^2 + b'(k_y^2 + k_z^2), \tag{2.9}$$

due to the symmetry around $\boldsymbol{k}_0 = k_0(1,0,0)$. The electronic properties of these semiconductors are determined by the states close to the band edges, so that these $\boldsymbol{k}\cdot\boldsymbol{p}$ - approximations – even if they only describe the energy bands very locally – play an important role in the physics of semiconductors.

## 2.2 Elementary excitations in semiconductors

We consider a simple two-band model to illustrate the most basic properties of the excitation spectrum of a semiconductor. The Hamiltonian is given by

$$\mathcal{H} = \sum_{\boldsymbol{k},s} \epsilon_{V,\boldsymbol{k}} \widehat{c}^{\dagger}_{V,\boldsymbol{k},s} \widehat{c}_{V,\boldsymbol{k},s} + \sum_{\boldsymbol{k},s} \epsilon_{C,\boldsymbol{k}} \widehat{c}^{\dagger}_{C,\boldsymbol{k},s} \widehat{c}_{C,\boldsymbol{k},s}, \tag{2.10}$$

where $\epsilon_{V,\boldsymbol{k}}$ and $\epsilon_{C,\boldsymbol{k}}$ are the band energies of the valence band and conduction band, respectively. The operator $c^{\dagger}_{n\boldsymbol{k}s}$ ($c_{n\boldsymbol{k}s}$) creates (annihilates) an electron with (pseudo-)momentum $\boldsymbol{k}$ and spin $s$ in the band $n$, $n \in \{V, C\}$. In the ground state $|\Phi_0\rangle$,

$$|\Phi_0\rangle = \prod_{\boldsymbol{k},s} \widehat{c}^{\dagger}_{V,\boldsymbol{k},s} |0\rangle, \tag{2.11}$$

the valence band is completely filled, whereas the conduction band is empty. The product on the right-hand side runs over all wave vectors in the first Brillouin zone. The ground state energy is given by

$$E_0 = 2 \sum_{\boldsymbol{k}} \epsilon_{V,\boldsymbol{k}}. \tag{2.12}$$

The total momentum and spin of the ground state vanish. We next consider single electron excitations from that ground state.

### 2.2.1 Electron-hole excitations

A simple excitation of the system consists of removing an electron (i.e., creating a hole) from the valence band and inserting it into the conduction band. We write such an excitation as

$$|\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle = c^{\dagger}_{C,\boldsymbol{k}+\boldsymbol{q},s} c_{V,\boldsymbol{k},s'} |\Phi_0\rangle, \tag{2.13}$$

where we remove an electron with pseudo-momentum $\boldsymbol{k}$ and spin $s'$ in the valence band and replace it by an electron with $\boldsymbol{k}+\boldsymbol{q}$ and $s$ in the conduction band. The possibility of changing the spin from $s'$ to $s$ and of shifting the wave vector of conduction electrons by $\boldsymbol{q}$ is included. Furthermore $|\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle$ is assumed to be normalized. The electron-hole pair may either be in a spin-singlet (pure charge excitation $s = s'$) or a spin-triplet state (spin excitation $s \neq s'$). Apart from spin, the state is characterized by the wave vectors $\boldsymbol{k}$ and $\boldsymbol{q}$. The excitation energy is given by

$$E = \epsilon_{C,\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{V,\boldsymbol{k}}. \tag{2.14}$$

The spectrum of the electron-hole excitations with fixed $\boldsymbol{q}$ is determined by the spectral function

$$I(\boldsymbol{q}, E) = \sum_{\boldsymbol{k},s,s'} |\langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'| c^{\dagger}_{C,\boldsymbol{k}+\boldsymbol{q},s} c_{V,\boldsymbol{k}s'} |\Phi_0\rangle|^2 \delta(E - (\epsilon_C(\boldsymbol{k}+\boldsymbol{q}) - \epsilon_V(\boldsymbol{k})). \tag{2.15}$$

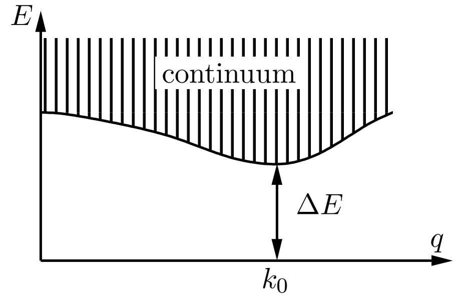



Figure 2.6: Electron-hole excitation spectrum for $\boldsymbol{k}_0 \neq 0$. Excitations exist in the shaded region, where $I(q, E) \neq 0$.

Excitations exist for all pairs $E$ and $\boldsymbol{q}$ for which $I(\boldsymbol{q}, E)$ does not vanish, thus, only above a $\boldsymbol{q}$-dependent threshold, which is minimal for $\boldsymbol{q} = \boldsymbol{k}_0$, where $\boldsymbol{k}_0 = 0$ ($\boldsymbol{k}_0 \neq 0$) for a direct (indirect) energy gap. As $\boldsymbol{k}$ is arbitrary, there is a continuum of excited states above the threshold for each $\boldsymbol{q}$ (see Figure 2.6).
For the electron-hole excitations considered here, interactions among them was assumed to be irrelevant, and the electrons involved are treated as non-interacting particles. Note the analogy with the Dirac-sea in relativistic quantum mechanics: The electron-hole excitations of a semiconductor correspond to electron-positron pair creation in the Dirac theory.

### 2.2.2 Excitons

Taking into account the Coulomb interaction between the electrons, there is another class of excitations called excitons. In order to discuss them, we extend the Hamiltonian (2.10) by the Coulomb interaction,

$$\widehat{V} = \sum_{s,s'} \int d^3r\, d^3r' \widehat{\Psi}_s^\dagger(\boldsymbol{r}) \widehat{\Psi}_{s'}^\dagger(\boldsymbol{r}') \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \widehat{\Psi}_{s'}(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}), \tag{2.16}$$

where the field operators are defined by

$$\widehat{\Psi}_s(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{n=V,C} \sum_{\boldsymbol{k}} u_{n,\boldsymbol{k}}(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \widehat{c}_{n,\boldsymbol{k}s}, \tag{2.17}$$

where $u_{n,\boldsymbol{k}}(\boldsymbol{r})$ are the Bloch functions of the band $n = C, V$. Now, we consider a general particle-hole state,

$$|\Phi_{\boldsymbol{q}}\rangle = \sum_{\boldsymbol{k}} A(\boldsymbol{k}) \widehat{c}^\dagger_{C,\boldsymbol{k}+\boldsymbol{q},s} \widehat{c}_{V,\boldsymbol{k},s'} |\Phi_0\rangle = \sum_{\boldsymbol{k}} A(\boldsymbol{k}) |\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle, \tag{2.18}$$

and demand that it satisfies the stationary Schrödinger equation $(\mathcal{H} + \widehat{V})|\Phi_{\boldsymbol{q}}\rangle = E|\Phi_{\boldsymbol{q}}\rangle$. This two-body problem can be expressed as

$$\sum_{\boldsymbol{k}'} \langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s' | \mathcal{H} + \widehat{V} | \boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s' \rangle A(\boldsymbol{k}') = E A(\boldsymbol{k}). \tag{2.19}$$

The matrix elements are given by

$$\langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s' | \mathcal{H} | \boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s' \rangle = \delta_{\boldsymbol{k},\boldsymbol{k}'} \{\epsilon_{C,\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{V,\boldsymbol{k}}\} \tag{2.20}$$

and

$$\langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'|\widehat{V}|\boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s'\rangle =$$
$$\frac{2\delta_{ss'}}{\Omega^2}\int d^3r\, d^3r'\, u^*_{C,\boldsymbol{k}+\boldsymbol{q}}(\boldsymbol{r})u_{V,\boldsymbol{k}}(\boldsymbol{r})u_{C,\boldsymbol{k}'+\boldsymbol{q}}(\boldsymbol{r}')u^*_{V,\boldsymbol{k}'}(\boldsymbol{r}')e^{-i\boldsymbol{q}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}$$
$$-\frac{1}{\Omega^2}\int d^3r\, d^3r'\, u^*_{C,\boldsymbol{k}+\boldsymbol{q}}(\boldsymbol{r})u_{V,\boldsymbol{k}}(\boldsymbol{r}')u_{C,\boldsymbol{k}'+\boldsymbol{q}}(\boldsymbol{r})u^*_{V,\boldsymbol{k}'}(\boldsymbol{r}')e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot(\boldsymbol{r}-\boldsymbol{r}')}\frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}, \quad (2.21)$$

The first term is the exchange term, and the second term the direct term of the Coulomb interaction. Now we consider a semiconductor with a direct energy gap at the Γ-point. Thus, the most important wave vectors are those around $\boldsymbol{k}=0$. We approximate

$$u^*_{n,\boldsymbol{k}'}(\boldsymbol{r})u_{n,\boldsymbol{k}}(\boldsymbol{r}) \approx \frac{1}{\Omega}\int d^3r u^*_{n,\boldsymbol{k}'}(\boldsymbol{r})u_{n,\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\Omega}\langle u_{n,\boldsymbol{k}'}|u_{n,\boldsymbol{k}}\rangle \approx 1, \quad (2.22)$$

which is reasonable for $\boldsymbol{k} \approx \boldsymbol{k}'(=\boldsymbol{k}+\boldsymbol{q})$. In the same manner, we see that

$$u^*_{C\boldsymbol{k}+\boldsymbol{q}}(\boldsymbol{r})u_{V,\boldsymbol{k}}(\boldsymbol{r}) \approx \frac{1}{\Omega}\langle u_{C,\boldsymbol{k}+\boldsymbol{q}}|u_{V,\boldsymbol{k}}\rangle \approx \frac{1}{\Omega}\langle u_{C,\boldsymbol{k}}|u_{V,\boldsymbol{k}}\rangle = 0. \quad (2.23)$$

Note that the semiconductor is a dielectric medium with a dielectric constant $\varepsilon$ ($\boldsymbol{D} = \varepsilon\boldsymbol{E}$). Classical electrodynamics states that

$$\boldsymbol{\nabla}\cdot\boldsymbol{E} = \frac{4\pi\rho}{\varepsilon}, \quad (2.24)$$

i.e., the Coulomb potential is partially screened due to dielectric polarization. Including this effect in the Schrödinger equation phenomenologically, the matrix element (2.21) takes on the form

$$-\frac{4\pi e^2}{\Omega\varepsilon|\boldsymbol{k}-\boldsymbol{k}'|^2}. \quad (2.25)$$

Thus, we can write the stationary equation (2.19) as

$$\left(\epsilon_{C,\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{V,\boldsymbol{k}} - E\right)A(\boldsymbol{k}) - \frac{1}{\Omega}\sum_{\boldsymbol{k}'}\frac{4\pi e^2}{\varepsilon|\boldsymbol{k}-\boldsymbol{k}'|^2}A(\boldsymbol{k}') = 0. \quad (2.26)$$

We include the band structure using the $\boldsymbol{k}\cdot\boldsymbol{p}$ - approximation which, for a direct energy gap, leads to

$$\epsilon_{C,\boldsymbol{k}} = \frac{\hbar^2\boldsymbol{k}^2}{2m_C} + E_0 + E_g \quad \text{and} \quad \epsilon_{V,\boldsymbol{k}} = E_0 - \frac{\hbar^2\boldsymbol{k}^2}{2m_V}, \quad (2.27)$$

where $E_0$ denotes the energy of the valence band top. We define a so-called envelope function $F(\boldsymbol{r})$ by

$$F(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}}\sum_{\boldsymbol{k}} A(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \quad (2.28)$$

This function satisfies the differential equation

$$\left[-\frac{\hbar^2\boldsymbol{\nabla}^2}{2\mu_{\text{ex}}} + \frac{\hbar^2}{2i}\left(\frac{1}{m_V} - \frac{1}{m_C}\right)\boldsymbol{q}\cdot\boldsymbol{\nabla} - \frac{e^2}{\varepsilon|\boldsymbol{r}|}\right]F(\boldsymbol{r}) = \left\{E - E_g - \frac{\hbar^2\boldsymbol{q}^2}{2\mu_{\text{ex}}}\right\}F(\boldsymbol{r}), \quad (2.29)$$

where $\mu_{\text{ex}}$ is the reduced mass, i.e., $\mu_{\text{ex}}^{-1} = m_V^{-1} + m_C^{-1}$. The term linear in $\boldsymbol{\nabla}$ can be eliminated by the transformation

$$F(\boldsymbol{r}) = F'(\boldsymbol{r})\exp\left(\frac{i}{2}\frac{m_V - m_C}{m_V + m_C}\boldsymbol{q}\cdot\boldsymbol{r}\right), \quad (2.30)$$

and after some algebraic manipulations we obtain

$$\left[-\frac{\hbar^2\boldsymbol{\nabla}^2}{2\mu_{\text{ex}}}-\frac{e^2}{\varepsilon|\boldsymbol{r}|}\right]F'(\boldsymbol{r})=\left\{E-E_g-\frac{\hbar^2\boldsymbol{q}^2}{2M_{\text{ex}}}\right\}F'(\boldsymbol{r}), \tag{2.31}$$

where $M_{\text{ex}}=m_V+m_C$.
The stationary equation (2.31) is equivalent to the Schrödinger equation of a hydrogen atom. The energy levels then are given by

$$E_{\boldsymbol{q}}=E_g-\frac{\mu_{\text{ex}}e^4}{2\varepsilon^2\hbar^2n^2}+\frac{\hbar^2\boldsymbol{q}^2}{2M_{\text{ex}}}, \tag{2.32}$$

which implies that there are excitations below the particle-hole continuum, corresponding to particle-hole bound states. This excitation spectrum is discrete and there is a well-defined relation between energy and momentum ($\boldsymbol{q}$), which is the wave vector corresponding to the center of mass of the particle-hole pair. This non-trivial quasiparticle is called *exciton*. In the present approximation it takes on the form of a simple two-particle state. In fact, however, it may be viewed as a collective excitation, as the dielectric constant includes the polarization by all electrons. When the screening is neglected, the excitonic states would not make sense as their energies would not be within the band gap but much below. For the case of weak binding considered above, the excitation is called a *Wannier exciton*. The typical binding energy is

$$E_b\sim\frac{\mu_{\text{ex}}}{m\varepsilon^2}Ry. \tag{2.33}$$

Typical values of the constants on the right-hand side are $\varepsilon\sim 10$ and $\mu_{\text{ex}}\sim m/10$, so that the binding energy is in the meV range. This energy is much smaller than the energy gap, such that the excitons are inside the gap, as shown schematically in Figure 2.7.

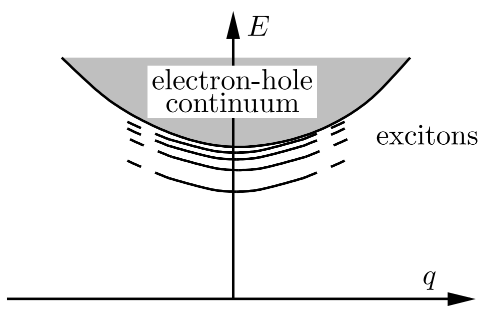



Figure 2.7: Qualitative form of the exciton spectrum below the electron-hole continuum.

The exciton levels are dispersive and their spectrum becomes increasingly dense with increasing energy, similar to the hydrogen atom. When they merge with the particle-hole continuum the bound state is 'ionized', i.e., the electron and the hole dissociate and behave like independent particles.
Strongly bound excitons are called *Frenkel excitons*. In the limit of strong binding, the pair is almost local, so that the excitation is restricted to a single atom rather than involving the whole semiconductor band structure.
Excitons are mobile, but they carry no charge, as they consist of an electron and a hole with opposite charges. Their spin quantum number depends on $s$ and $s'$. If $s=s'$ the exciton is a spin singlet, while for $s\neq'$ it has spin triplet character, both corresponding to integer spin quasiparticles. For small densities they approximately obey Bose-Einstein statistics, as they are made from two fermions. In special cases, Bose-Einstein condensation of excitons can be observed experimentally.

### 2.2.3 Optical properties

Excitation in semiconductors can occur via the absorption of electromagnetic radiation. The energy and momentum transferred by a photon is $\hbar\omega$ and $\hbar\boldsymbol{q}$, respectively. With the linear light dispersion relation $\omega = c|\boldsymbol{q}|$ and the approximation $E_g \sim 1eV \sim e^2/a$, we can estimate this momentum transfer in a semiconductor

$$q = \frac{\omega}{c} = \frac{\hbar\omega}{hc}2\pi \sim \frac{e^2}{hc}\frac{2\pi}{a} = \alpha\frac{2\pi}{a} \ll \frac{2\pi}{a}, \tag{2.34}$$

where $c$ denotes the speed of light, $a$ the lattice constant, and $\alpha \approx 1/137$ the fine structure constant. With this, the momentum transfer from a photon to the excited electron can be ignored. In other words, pure electromagnetic excitations lead only to 'direct' excitations. For semiconductors with a direct energy gap (e.g., GaAs) the photo-induced electron-hole excitation is most easy and yields absorption rates with the characteristics

$$\Gamma_{\text{abs}}(\omega) \propto \begin{cases} (\hbar\omega - E_g)^{1/2}, & \text{dipole-allowed,} \\ (\hbar\omega - E_g)^{3/2}, & \text{dipole-forbidden.} \end{cases} \tag{2.35}$$

Here, the terms "dipole-allowed" and "dipole-forbidden" have a similar meaning as in the excitation of atoms regarding whether matrix elements of the type $\langle u_{V,\boldsymbol{k}}|\boldsymbol{r}|u_{C,\boldsymbol{k}}\rangle$ are finite or vanish, respectively. Obviously, dipole-allowed transitions occur at a higher rate for photon energies immediately above the energy gap $E_g$, than for dipole-forbidden transitions.
For semiconductors with indirect energy gap (e.g., Si and Ge), the lowest energy transition connecting the top of the valence band to the bottom of the conduction band is not allowed without the help of phonons (lattice vibrations), which contribute little energy but much momentum transfer, as $\hbar\omega_{\boldsymbol{Q}} \ll \hbar\omega$ with $\omega_{\boldsymbol{Q}} = c_s|\boldsymbol{Q}|$ and the sound velocity $c_s \ll c$. The requirement of a phonon assisting in the transition reduces the transition rate to

$$\Gamma_{\text{abs}}(\omega) \propto c_+(\hbar\omega + \hbar\omega_{\boldsymbol{Q}} - E_g)^2 + c_-(\hbar\omega - \hbar\omega_{\boldsymbol{Q}} - E_g)^2, \tag{2.36}$$

where $c_\pm$ are constants and $\boldsymbol{Q}$ corresponds to the wave vector of the phonon connecting the top of the valence band and the bottom of the conduction band. There are two relevant processes: either the phonon is absorbed ($c_+$-process) or it is emitted ($c_-$-process) (see Figure 2.8).

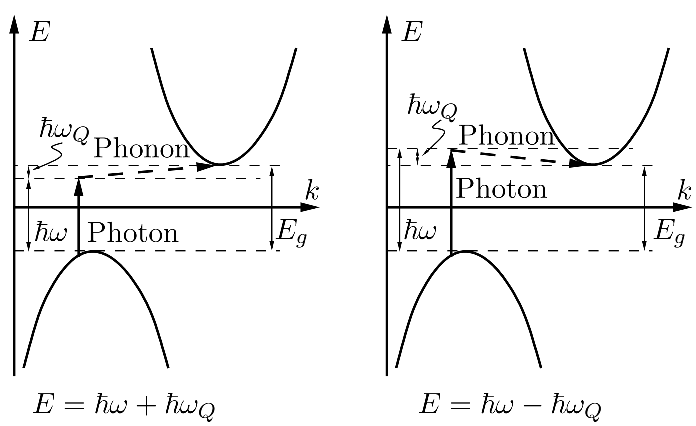



Figure 2.8: Phonon-assisted photon absorption in a semiconductor with indirect gap: phonon absorption (left panel) and phonon emission (right panel).

In addition to the absorption into the particle-hole spectrum, absorption processes inducing exciton states exist. They lead to discrete absorption peaks below the absorption continuum. In Figure 2.9, we show the situation for a direct-gap semiconductor.

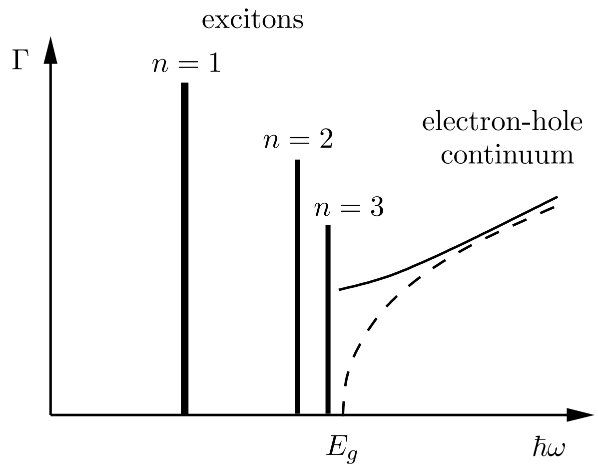



Figure 2.9: Absorption spectrum including the exciton states for a direct-gap semiconductor with dipole-allowed transitions. The exciton states appear as sharp lines below the electron-hole continuum starting at $\hbar\omega = E_g$.

Naturally, the recombination of electrons and holes is important as well; in particular, if it is a radiative recombination, i.e., leads to the emission of a photon. Additionally, other recombination channels such as recombination at impurities, interfaces and through Auger processes are possible. The radiative recombination for the direct-gap semiconductors is most relevant for applications. The photon emission rate follows the approximate law

$$\Gamma_{\text{em}}(\omega) \propto [N_\gamma(\omega) + 1](\hbar\omega - E_g)^{1/2} e^{-\hbar\omega/k_B T}, \tag{2.37}$$

with the photon density $N_\gamma(\omega)$. This yields the dominant rate for $\hbar\omega$ very close to $E_g$.

## 2.3 Doping semiconductors

Let us replace a Si atom in a Si semiconductor by aluminum Al (group III) or phosphorus P (group V), which then act as impurities in the crystal lattice. Both Al and P are in the same row of the periodic table, and their electron configurations are given by

$$\begin{aligned} \text{Al}: &\quad [(1\text{s})^2(2\text{s})^2(2\text{p})^6]\ \underline{(3\text{s})^2(3\text{p})}, \\ \text{P}: &\quad [(1\text{s})^2(2\text{s})^2(2\text{p})^6]\ \underline{(3\text{s})^2(3\text{p})^3}. \end{aligned}$$

The compound Al (P) has one electron less (more) than Si.

### 2.3.1 Impurity state

We consider the case of a P-impurity contributing an additional electron whose dynamics is governed by the conduction band of the semiconductor. For the sake of simplicity, we describe the conduction band by a single isotropic band with effective mass $m^*$,

$$\epsilon_{\boldsymbol{k}} = \frac{\hbar^2 \boldsymbol{k}^2}{2m^*} + E_g. \tag{2.38}$$

In the neutral Si background, the phosphorus (P) ion represents a positively charged center, which attracts its additional electron. In the simplest model, this situation is described by the so-called Wannier equation

$$\left\{ -\frac{\hbar^2 \boldsymbol{\nabla}^2}{2m^*} - \frac{e^2}{\varepsilon|\vec{r}|} \right\} F(\boldsymbol{r}) = EF(\boldsymbol{r}), \tag{2.39}$$

which is nothing else than the static Schrödinger equation for the hydrogen atom, where the dielectric constant $\varepsilon$ measures the screening of the ionic potential by the surrounding electrons. Analogous to the discussion of the exciton states, $F(\boldsymbol{r})$ is an envelope wave function of the electron. Therefore, the low energy states of the additional electron are bound states around the P ion. The electron may become mobile when this "reduced hydrogen atom" is ionized. The binding energy relative to the minimum of the conduction band given by

$$E_n - E_g = -\frac{m^* e^4}{2\hbar^2 \varepsilon^2 n^2} = -\frac{m^*}{m\varepsilon^2 n^2}\mathrm{Ry}, \tag{2.40}$$

for $n \in \mathbb{N}$ and the effective radius (corresponding to the renormalized Bohr radius in the material) of the lowest bound state reads

$$r_1 = \frac{\hbar^2 \varepsilon}{m^* e^2} = \frac{\varepsilon m}{m^*} a_B, \tag{2.41}$$

where $a_B = 0.53\text{Å}$ is the Bohr radius for the hydrogen atom. For Si we find $m^* \approx 0.2m$ and $\varepsilon \approx 12$, such that

$$E_1 \approx -20\mathrm{meV} \tag{2.42}$$

and

$$r_1 \approx 30\text{Å}. \tag{2.43}$$

Thus, the resulting states are weakly bound, with energies inside the band gap. We conclude that the net effect of the P-impurities is to introduce additional electrons into the crystal, whose energies lie just below the conduction band ($E_g \sim 1\mathrm{eV}$ while $E_g - E_1 \sim 10\mathrm{meV}$). Therefore, they can easily be transferred to the conduction band by thermal excitation (ionization). One speaks of an *n-doped* semiconductor (n: negative charge). In full analogy one can consider Al-impurities, thereby replacing electrons with holes: An Al-atom introduces an additional hole into the lattice which is weakly bound to the Al-ion (its energy is slightly above the band edge of the valence band) and may dissociate from the impurity by thermal excitation. This case is called *p-doping* (p: positive charge). In both cases, the chemical potential is tied to the dopand levels, i.e., it lies between the dopand level and the valence band for p-doping and between the dopand level and the conduction band in case of n-doping (Figure 2.10).

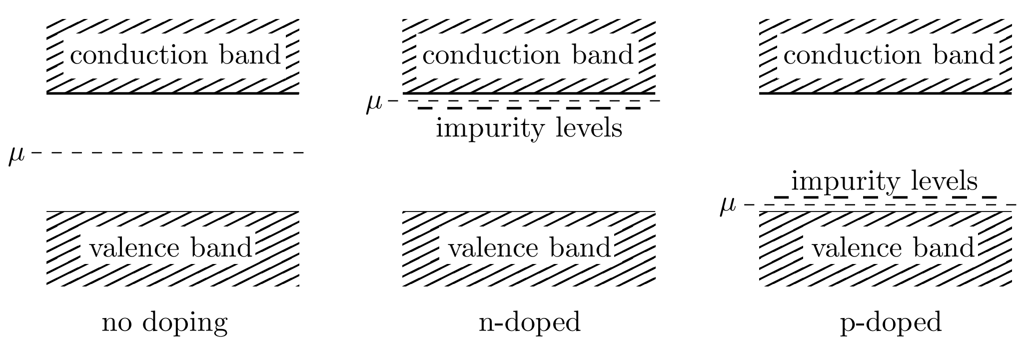



Figure 2.10: Position of the chemical potential in semiconductors.

The electric conductivity of semiconductors (in particular at room temperature) can be tuned strongly by doping with so-called 'donators' (n-doping) and 'acceptors' (p-doping). Practically all dopand atoms are ionized, with the electrons/holes becoming mobile. Combining differently doped semiconductors, the possibility to engineer electronic properties is enhanced even more. This is the basic reason for the semiconductors being ubiquitous in modern electronics.

### 2.3.2 Carrier concentration

Let us briefly compare the carrier concentration in doped and undoped semiconductors at room temperature. Carriers are always created in form of electron hole pairs, following the "reaction formula"

$$e + h \leftrightarrow \gamma, \tag{2.44}$$

where $\gamma$ denotes a photon which is absorbed ($e$-$h$-creation) or emitted ($e$-$h$-recombination) and accounts for the energy balance. The carrier concentration is described by a mass action law of the form,

$$n_e n_h = n_0^2 \left(\frac{T}{T_0}\right)^3 e^{-E_g/k_B T} = n^2(T), \tag{2.45}$$

where $T_0$, $n_0$ and $E_g$ are parameters specific to the semiconductor. In the case of undoped silicon at $T = 300\text{K}$, $n_e n_h \approx 10^{20} \left(\text{cm}^{-3}\right)^2$. Thus, for the undoped semiconductor we find $n_e = n_h \approx 10^{10}\text{cm}^{-3}$. On the other hand, for n-doped Si with a typical donor concentration of $n_D \approx 10^{17}\text{cm}^{-3}$ we can assume that most of the donors are ionized at room temperature such that

$$n_e \approx \quad n_D \quad \approx 10^{17}\text{cm}^{-3} \tag{2.46}$$

and

$$n_h = \frac{n^2(T)}{n_e} \approx 10^3\text{cm}^{-3}. \tag{2.47}$$

We conclude, that in n-doped superconductors the vast majority of mobile carriers are electrons, while the hole carriers are negligible. The opposite is true for p-doped Si.

## 2.4 Semiconductor devices

Semiconductors are among the most important components of current high-technology. In this section, we consider a few basic examples of semiconductor devices.

### 2.4.1 pn-contacts

The so-called pn-junctions, made by bringing in contact a p-doped and an n-doped version of the same semiconductor, are used as rectifiers.[4] When contacting the two types of doped semiconductors the chemical potential, which is pinned by the dopand (impurity) levels, determines the behavior of the electrons at the interface. In electrostatic equilibrium, the chemical potential is constant across the interface. This is accompanied by a "band bending" leading to the ionization of the impurity levels in the interface region (see Figure 2.11). Consequently, these ions produce an electric dipole layer which induces an electrostatic potential shift across the interface. Additionally, the carrier concentration is strongly reduced in the interface region (depletion layer).
In the absence of a voltage $U$ over the junction, the net current flow vanishes because the dipole is in electrostatic equilibrium. This can also be interpreted as the equilibrium of two oppositely directed currents, called the drift current $J_{\text{drift}}$ and the diffusion current $J_{\text{diff}}$. From the point of view of the electrons, the dipole field exerts a force pulling the electrons from the p-side to the n-side. This leads to the drift current $J_{\text{drift}}$. On the other hand, the electron concentration gradient leads to the diffusion current $J_{\text{diff}}$ from the n-side to the p-side. The diffusion current

[4] dt. Gleichrichter

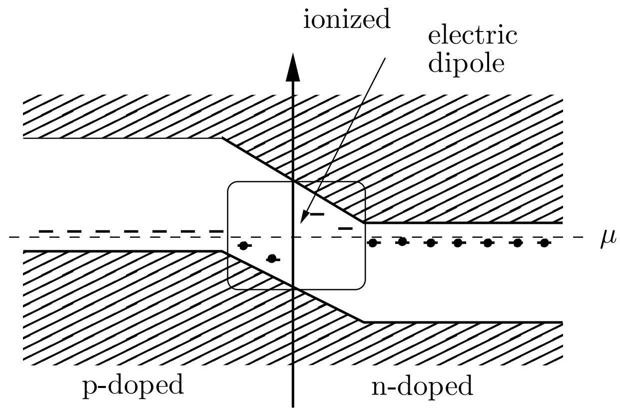



Figure 2.11: Occupation of the impurity levels of a pn-junction.

is directed against the potential gradient, so that the diffusing electrons have to overcome a potential step. The equilibrium condition for $U = 0$ is given by

$$0 = J_{\text{tot}}(U = 0) = J_{\text{diff}} + J_{\text{drift}} \propto C_1(T)e^{-E_g/k_BT} - C_2(T)e^{-E_g/k_BT} = 0, \tag{2.48}$$

where $C_1 = C_2 = C$. Both currents are essentially determined by the factor $C(T)e^{-E_g/k_BT}$. For the drift current, the exponential behavior $e^{-E_g/k_BT}$ stems from the dependence of the current on the concentration of mobile charge carriers (electrons and holes on the p-side and n-side, respectively), which are created by thermal excitation (Boltzmann factor). Applying a voltage does not change this contribution significantly. For the diffusion current however, the factor $C(T)e^{-E_g/k_BT}$ describes the thermal activation over the dipole barrier, which in turn strongly depends on the applied voltage $U$. For zero voltage, the height of the barrier $E_b$ is essentially given by the energy gap $E_b \approx E_g$. With an applied voltage, this is modified according to $E_b \approx E_g + eU$, where $eU = \mu_n - \mu_p$ is the difference of the chemical potentials between the n-side and the p-side. From these considerations, the well-known current-voltage characteristic of the pn-junctions follows directly as

$$J_{\text{tot}}(U) = C(T)e^{-E_g/k_BT}\left(e^{eU/k_BT} - 1\right). \tag{2.49}$$

For $U > 0$, the current is rapidly enhanced with increasing voltage. This is called forward bias. By contrast, charge transport is suppressed for $U < 0$ (reverse bias), leading to small currents only. The current-voltage characteristics $J(U)$ (see Figure 2.12) shows a clearly asymmetric behavior, which can be used to rectify ac-currents. Rectifiers (or diodes) are an important component of many integrated circuits.

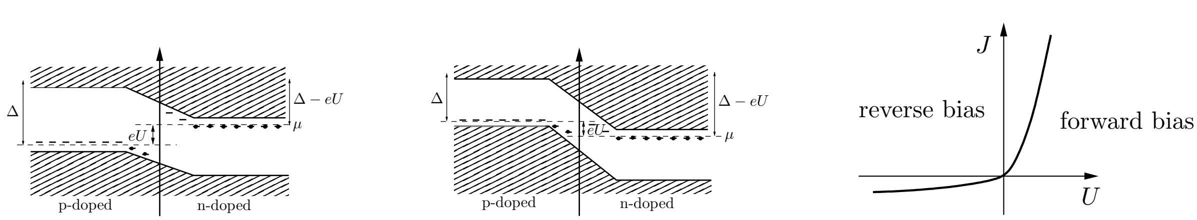



Figure 2.12: The pn-junction with an applied voltage and the resulting $J$-$U$ characteristics.

### 2.4.2 Semiconductor diodes

#### Light emitting diode

As mentioned above, the recombination of electrons and holes can lead to the emission of photons (radiative recombination) with a rather well-defined frequency essentially corresponding to the energy gap $E_g$. An excess of electron-hole pairs can be produced in pn-diodes by running a current in forward direction. Using different semiconductors with different energy gaps allows to tune the color of the emitted light. Direct-gap semiconductors are most suitable for this kind of devices. Well-know are the semiconductors of the GaAs-GaN series (see table 2.1). These techniques are commonly used in LED (light emitting diode) lamps.

There appear efficiency problems concerning the emission of light by semiconductors. In

| semiconductor | GaAs | $GaAs_{0.6}P_{0.4}$ | $GaAs_{0.4}P_{0.6}$ | GaP | GaN |
|---|---|---|---|---|---|
| wave length (nm) | 940 | 660 | 620 | 550 | 340 |
| color | infrared | red | yellow | green | ultraviolet |

Table 2.1: Materials commonly used for LEDs and their light emitting properties.

particular, the difference in refractive indices inside $n_{\mathrm{SC}} \approx 3$ and outside $n_{\mathrm{air}} \approx 1$ the device leads to large reflective losses. Thus, the efficacy of diode light sources, defined as the number of photons emitted per created particle-hole pair, is small, but still larger than the efficiency of conventional dissipative light bulbs.

#### Solar cell

Inversely to the previous consideration, the population of charge carriers can be changed by the absorption of light. Suppose that the n-side of a diode is exposed to irradiation by light, which leads to an excess of hole carriers (minority charge carriers). Some of these holes will diffuse towards the pn-interface and will be drawn to the p-side by the dipole field. In this way, they induce an additional current $J_L$ modifying the current-voltage characteristics to

$$J_{\mathrm{tot}} = J_{pn} - J_L = J_s(e^{eU/k_BT} - 1) - J_L. \tag{2.50}$$

It is important for the successful migration of the holes to the interface dipole that they do not recombine too quickly. When $J_{\mathrm{tot}} = 0$, the voltage drop across the diode is

$$U_L = \frac{k_B T}{e} \ln\left(\frac{J_L}{J_s} + 1\right). \tag{2.51}$$

The maximum efficiency is reached by applying an external voltage $U_c < U_L$ such that the product $J_c U_c$ is maximized, where $J_c = J_{\mathrm{tot}}(U = U_c)$ (cf. Figure 2.13).

### 2.4.3 MOSFET

The arguably most important application of semiconductors is the transistor, an element existing with different architectures. Here we shortly introduce the MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor). A transistor is a switch allowing to control the current through the device by switching a small control voltage. In the MOSFET, this is achieved by changing the charge carrier concentration in a p-doped semiconductor using a metallic gate. The basic design of a MOSFET is as follows (see Figure 2.14): A thin layer of $SiO_2$ is deposited on the surface of a p-type semiconductor. $SiO_2$ is a good insulator that is compatible with the lattice structure of Si. Next, a metallic layer, used as a gate electrode, is deposited on top of the insulating layer.

The voltage between the Si semiconductor and the metal electrode is called gate voltage $U_G$.

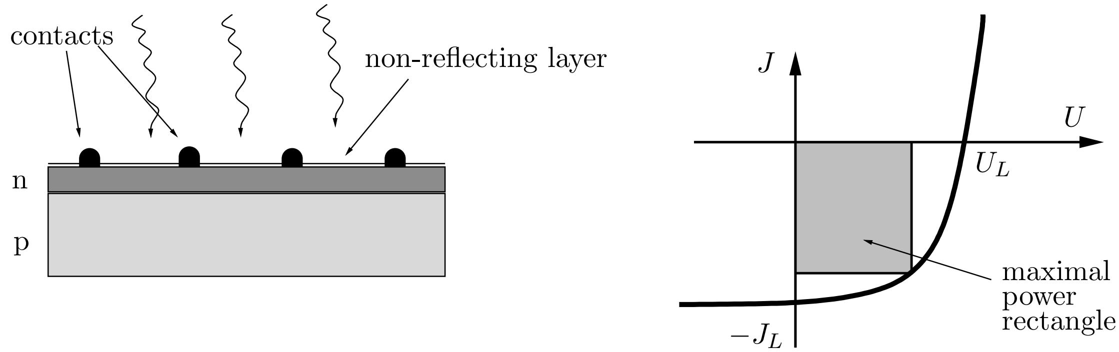



Figure 2.13: Solar cell design and shifted current-voltage characteristics. The efficiency is maximal for a maximal area of the power rectangle.

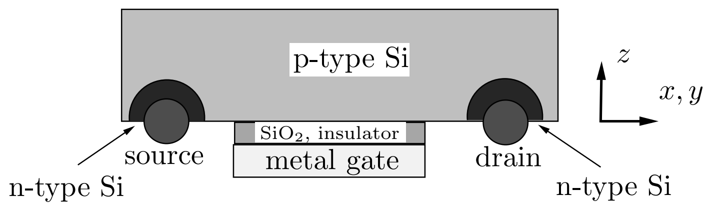



Figure 2.14: Schematic design of a MOSFET device.

The insulating $SiO_2$ layer ensures that no currents flow between the electrode and the semiconductor when a gate voltage is applied. The switchable currents in the MOSFET flow between the source and the drain which are heavily n-doped semiconductor regions.

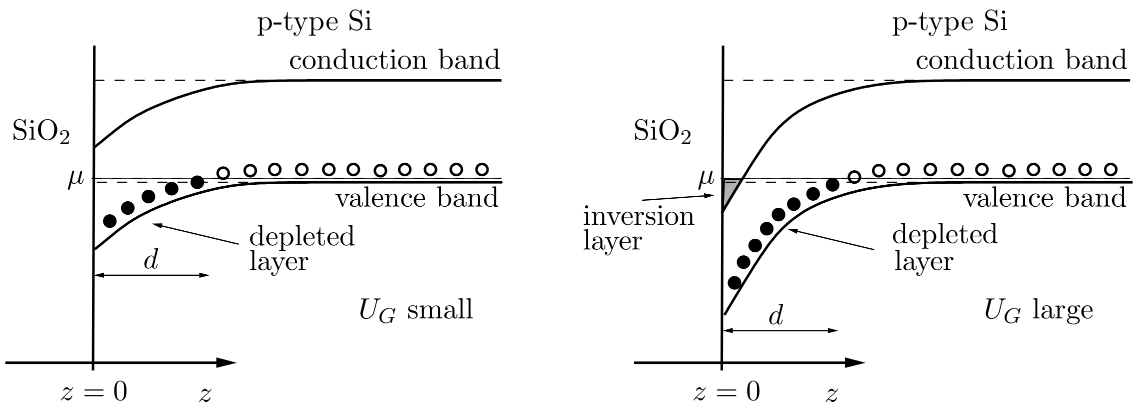



Figure 2.15: Depletion layer at the $SiO_2$-Si interface for $0 < e\,U_G < E_g$ (left panel) and the inversion layer $E_g < e\,U_G$ (right panel).

Depending on the applied gate voltage $U_G$ three different regimes can be realized:

**1.** $U_G = 0$

Virtually no current flows, as the conduction band of the p-doped semiconductor is empty. The doping states (acceptor levels) are occupied by thermal excitations.

**2.** $0 < \frac{eU_G}{E_g} < 1$

In this case, the energy of the Si bands is lowered, such that in a narrow region within the p-doped Si the acceptor levels drops below the chemical potential and the states are filled with electrons (or, equivalently, holes are removed). This depletion layer has the extension $d$ measured from the Si-Sio$_2$ interface. The negative charge of the acceptors leads to a position-dependent potential $\Phi(z)$, where $z$ is the distance from the boundary between $SiO_2$ and Si. This potential $\Phi(z)$ satisfies the simple one-dimensional Poisson equation

$$\frac{d^2}{dz^2}\Phi(z) = \frac{4\pi\rho(z)}{\varepsilon}, \tag{2.52}$$

where the charge density originates in the occupied acceptor levels,

$$\rho(z) = \begin{cases} -en_A, & z < d, \\ 0, & z > d, \end{cases} \tag{2.53}$$

and $n_A$ is the density of acceptors. The boundary conditions are given by

$$\Phi(z=0) = U_G \qquad \text{and} \qquad \Phi(z=d) = 0. \tag{2.54}$$

The solution for $0 \le z \le d$ then reads

$$\Phi(z) = \frac{2\pi en_A}{\varepsilon}(z-d)^2, \qquad \text{with} \qquad d^2 = \frac{\varepsilon U_G}{2\pi en_A}. \tag{2.55}$$

The thickness of the depletion layer increases with increasing gate voltage $d^2 \propto U_G$.

**3.** $1 < \frac{eU_G}{E_g}$

When the applied gate voltage is sufficiently large, a so-called inversion layer is created (cf. Figure 2.15). Close to the boundary, the conduction band is bent down so that its lower edge lies below the chemical potential. The electrons accumulating in this inversion layer providing carriers connecting the n-type source and drain electrodes and producing a large, nearly metallic, current between source and drain. Conduction band electrons accumulating in the inversion layer behave like a two-dimensional electron gas. In such a system, the quantum Hall effect (QHE), which is characterized by highly unusual charge transport properties in the presence of a large magnetic field, can occur.

# Chapter 3

# Metals

The electronic states in a periodic atomic lattice are extended and have an energy spectrum forming energy bands. In the ground state these energy states are filled successively starting at the bottom of the electronic spectrum until the number of electrons is exhausted. Metallic behavior occurs whenever in this way a band is only partially filled. The fundamental difference that distinguishes metals from insulators and semiconductors is the absence of a gap for electron-hole excitations. In metals, the ground state can be excited at arbitrarily small energies which has profound phenomenological consequences.
We will consider a basic model suitable for the description of simple metals like the Alkali metals Li, Na, or K, where the (atomic) electron configuration consists of closed shell cores and one single valence electron in an ns-orbital. Neglecting the core electrons (completely filled bands), we consider the valence electrons only and apply the approximation of nearly free electrons. The lowest band around the Γ-point is then half-filled. First, we will also neglect the influence of the periodic lattice potential and consider the problem of a free electron gas subject to mutual (repulsive) Coulomb interaction.

## 3.1 The Jellium model of the metallic state

The Jellium[1] model is the probably simplest possible model of a metal that is able to describe qualitative and to some extend even quantitative aspects of simple metals. The main simplification made is to replace the ionic lattice by a homogeneous positively charged background (Jellium). The uniform charge density $en_{\text{ion}}$ is chosen such that the whole system – electrons and ionic background – is charge neutral, i.e. $n_{\text{ion}} = n$, where $n$ is the electron density. In this fully translational invariant system, the plane waves

$$\psi_{\boldsymbol{k},s}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{3.1}$$

represent the single-particle wave functions of the free electrons. Here $\Omega$ is the volume of the system, $\boldsymbol{k}$ and $s \in \{\uparrow, \downarrow\}$ denote the wave vector and spin, respectively. Assuming a cubic system of side length $L$ and volume $\Omega = L^3$ we impose periodic boundary conditions for the wave function

$$\psi_{\boldsymbol{k},s}(\boldsymbol{r} + (L,0,0)) = \psi_{\boldsymbol{k},s}(\boldsymbol{r} + (0,L,0)) = \psi_{\boldsymbol{k},s}(\boldsymbol{r} + (0,0,L)) = \psi_{\boldsymbol{k},s}(\boldsymbol{r}) \tag{3.2}$$

such that the reciprocal space is discretized as

$$\boldsymbol{k} = \frac{2\pi}{L}(n_x, n_y, n_z) \tag{3.3}$$

[1] Jellium originates form the word jelly (gelatin) and was first introduced by Conyers Herring.

where $(n_x, n_y, n_z) \in \mathbb{Z}^3$. The energy of a single-electron state is given by $\epsilon_{\boldsymbol{k}} = \hbar^2 \boldsymbol{k}^2/2m$ (free particle). The ground state of non-interacting electrons is obtained by filling all single particle states up to the Fermi energy with two electrons. In the language of second quantization the ground state is, thus, given by

$$|\Psi_0\rangle = \prod_{|\boldsymbol{k}|\leq k_F} \prod_s \widehat{c}^\dagger_{\boldsymbol{k},s}|0\rangle \tag{3.4}$$

where the operators $\widehat{c}^\dagger_{\boldsymbol{k},s}$ ($\widehat{c}_{\boldsymbol{k},s}$) create (annihilate) an electron with wave vector $\boldsymbol{k}$ and spin $s$. The Fermi wave vector $k_F$ with the corresponding Fermi energy $\epsilon_F = \hbar^2 k_F^2/2m$ is determined by equating the filled electronic states with the electron density $n$. We have

$$n = \frac{1}{\Omega}\sum_{|\boldsymbol{k}|\leq k_F, s} 1 = 2\int \frac{d^3k}{(2\pi)^3} 1 = 2\frac{4\pi}{3}\frac{k_F^3}{(2\pi)^3}, \tag{3.5}$$

which results in

$$k_F = (3\pi^2 n)^{1/3} . \tag{3.6}$$

Note $k_F$ is the radius of the Fermi sphere in $k$-space around $\boldsymbol{k} = 0$.[2] We define here also the Next, we compute the ground state energy of the Jellium system variationally, using the density $n$ as a variational parameter, which is equivalent to the variation of the lattice constant. In this way, we will obtain an understanding of the stability of a metal, i.e. the cohesion of the ion lattice through the itinerant electrons (in contrast to semiconductors where the stability was due to covalent chemical bonding). The variational ground state shall be $|\Psi_0\rangle$ from Eq.(3.4) for given $k_F$. The Hamiltonian splits into four terms

$$\mathcal{H} = \mathcal{H}_{\text{kin}} + \mathcal{H}_{ee} + \mathcal{H}_{ei} + \mathcal{H}_{ii} \tag{3.7}$$

with

$$\mathcal{H}_{\text{kin}} = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s} \tag{3.8}$$

$$\mathcal{H}_{\text{ee}} = \frac{1}{2}\sum_{s,s'}\int d^3r\, d^3r'\, \widehat{\Psi}^\dagger_s(\boldsymbol{r})\widehat{\Psi}^\dagger_{s'}(\boldsymbol{r}')\frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\widehat{\Psi}_{s'}(\boldsymbol{r}')\widehat{\Psi}_s(\boldsymbol{r}) \tag{3.9}$$

$$\mathcal{H}_{\text{ei}} = -\sum_s \int d^3r\, d^3r'\, \frac{ne^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\widehat{\Psi}^\dagger_s(\boldsymbol{r})\widehat{\Psi}_s(\boldsymbol{r}) \tag{3.10}$$

$$\mathcal{H}_{\text{ii}} = \frac{1}{2}\int d^3r\, d^3r'\, \frac{n^2e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{3.11}$$

where we have used in second quantization language the electron field operators

$$\widehat{\Psi}^\dagger_s(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}}\sum_{\boldsymbol{k}} \widehat{c}^\dagger_{\boldsymbol{k},s} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{3.12}$$

$$\widehat{\Psi}_s(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}}\sum_{\boldsymbol{k}} \widehat{c}_{\boldsymbol{k},s} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{3.13}$$

The variational energy – which we want to minimize with respect to $n$ – can be computed from $E_g = \langle\Psi_0|\mathcal{H}|\Psi_0\rangle$ and consists of four different contributions:

[2] Note that the function $k_F(n) \propto n^{(1/d)}$ depends on the dimensionality $d$ of the system.

First we have the kinetic energy

$$E_{\text{kin}} = \langle \Psi_0 | \mathcal{H}_{\text{kin}} | \Psi_0 \rangle = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \underbrace{\langle \Psi_0 | \widehat{c}^\dagger_{\boldsymbol{k}s} \widehat{c}_{\boldsymbol{k}s} | \Psi_0 \rangle}_{=\ n_{\boldsymbol{k}s}} \tag{3.14}$$

$$= 2\Omega \int \frac{d^3k}{(2\pi)^3}\, \epsilon_{\boldsymbol{k}}\, n_{\boldsymbol{k}s} = N \frac{3}{5} \epsilon_F \tag{3.15}$$

where we used $N = \Omega n$ the number of valence electrons and

$$n_{\boldsymbol{k}s} = \begin{cases} 1 & |\boldsymbol{k}| \le k_F \\ 0 & |\boldsymbol{k}| > k_F \end{cases} . \tag{3.16}$$

Secondly, there is the energy resulting from the Coulomb repulsion between the electrons,

$$E_{\text{ee}} = \frac{1}{2} \int d^3r\; d^3r' \; \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \sum_{s,s'} \langle \Psi_0 | \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}^\dagger_{s'}(\boldsymbol{r}') \widehat{\Psi}_{s'}(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}) | \Psi_0 \rangle \tag{3.17}$$

$$= \frac{1}{2} \int d^3r\; d^3r' \; \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \left( n^2 - G(\boldsymbol{r} - \boldsymbol{r}') \right) \tag{3.18}$$

$$= E_{\text{Hartree}} + E_{\text{Fock}}. \tag{3.19}$$

For this contribution we used the fact, that the two-particle correlation function from equation (3.17) may be expressed[3] as

$$\sum_{s,s'} \langle \Psi_0 | \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}^\dagger_{s'}(\boldsymbol{r}') \widehat{\Psi}_{s'}(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}) | \Psi_0 \rangle = n^2 - G(\boldsymbol{r} - \boldsymbol{r}') \tag{3.27}$$

[3] We shortly sketch the derivation of the pair correlation function. Using equations (3.12) and (3.13) we find

$$\langle \Psi_0 | \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}^\dagger_{s'}(\boldsymbol{r}') \widehat{\Psi}_{s'}(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}) | \Psi_0 \rangle = \frac{1}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q},\boldsymbol{q}'} e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}} e^{-i(\boldsymbol{q}-\boldsymbol{q}')\cdot\boldsymbol{r}'} \langle \Phi_0 | \widehat{c}^\dagger_{\boldsymbol{k}s} \widehat{c}^\dagger_{\boldsymbol{q}s'} \widehat{c}_{\boldsymbol{q}'s'} \widehat{c}_{\boldsymbol{k}'s} | \Phi_0 \rangle. \tag{3.20}$$

We distinguish two cases:
First, consider $s \neq s'$,

$$\langle \Phi_0 | \widehat{c}^\dagger_{\boldsymbol{k}s} \widehat{c}^\dagger_{\boldsymbol{q}s'} \widehat{c}_{\boldsymbol{q}'s'} \widehat{c}_{\boldsymbol{k}'s} | \Phi_0 \rangle = \delta_{\boldsymbol{k}\boldsymbol{k}'} \delta_{\boldsymbol{q}\boldsymbol{q}'} n_{\boldsymbol{k}s} n_{\boldsymbol{q}s'} \tag{3.21}$$

leading to

$$\langle \Psi_0 | \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}^\dagger_{s'}(\boldsymbol{r}') \widehat{\Psi}_{s'}(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}) | \Psi_0 \rangle = \frac{1}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{q}} n_{\boldsymbol{k}s} n_{\boldsymbol{q},s'} = \frac{n^2}{4}. \tag{3.22}$$

Secondly, assume $s = s'$ such that

$$\langle \Phi_0 | \widehat{c}^\dagger_{\boldsymbol{k}s} \widehat{c}^\dagger_{\boldsymbol{q}s} \widehat{c}_{\boldsymbol{q}'s} \widehat{c}_{\boldsymbol{k}'s} | \Phi_0 \rangle = (\delta_{\boldsymbol{k}\boldsymbol{k}'} \delta_{\boldsymbol{q}\boldsymbol{q}'} - \delta_{\boldsymbol{k}\boldsymbol{q}'} \delta_{\boldsymbol{q}\boldsymbol{k}'}) n_{\boldsymbol{k}s} n_{\boldsymbol{q}s} \,, \tag{3.23}$$

which in turn leads to

$$\langle \Psi_0 | \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}^\dagger_s(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}') \widehat{\Psi}_s(\boldsymbol{r}) | \Psi_0 \rangle = \frac{1}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{q}} \left( 1 - e^{i(\boldsymbol{q}-\boldsymbol{k})\cdot(\boldsymbol{r}-\boldsymbol{r}')} \right) n_{\boldsymbol{k}s} n_{\boldsymbol{q},s}. \tag{3.24}$$

Both cases eventually lead to the result in equation (3.27) with

$$G(\boldsymbol{r}) = 2 \left( \frac{1}{\Omega} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} n_{\boldsymbol{k}s} \right)^2 = 2 \left( \int_{|\boldsymbol{k}| \le k_F} \frac{d^3k}{(2\pi)^3} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \right)^2 \tag{3.25}$$

$$= 2 \left( \frac{1}{2\pi^2 r} \int_0^{k_F} dk\; k\; \sin kr \right)^2 = 2 \left( \frac{1}{2\pi^2} \frac{\sin k_F r - k_F r \cos k_F r}{r^3} \right)^2 \tag{3.26}$$

and $n = k_F^3/3\pi^2$ ($k = |\boldsymbol{k}|$ and $r = |\boldsymbol{r}|$).

where

$$G(\boldsymbol{r}) = \frac{9n^2}{2}\left(\frac{k_F|\boldsymbol{r}|\cos k_F|\boldsymbol{r}| - \sin k_F|\boldsymbol{r}|}{(k_F|\boldsymbol{r}|)^3}\right)^2. \tag{3.28}$$

The Coulomb repulsion $\mathcal{H}_{\text{ee}}$ between the electrons leads to two terms, called the *direct* or Hartree term describing the Coulomb energy of a uniformly spread charge distribution, and the *exchange* or Fock term resulting from the exchange hole that follows from the Fermi-Dirac statistics (Pauli exclusion principle).
The third contribution originates in the attractive interaction between the (uniform) ionic background and the electrons,

$$E_{\text{ei}} = -\int d^3r\ d^3r'\ \frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|} n \sum_s \langle\Psi_0|\widehat{\Psi}_s^\dagger(\boldsymbol{r})\Psi_s(\boldsymbol{r})|\Psi_0\rangle \tag{3.29}$$

$$= -\int d^3r\ d^3r'\ \frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|} n^2. \tag{3.30}$$

Were the expectation value $\langle\Psi_0|\widehat{\Psi}_s^\dagger(\boldsymbol{r})\Psi_s(\boldsymbol{r})|\Psi_0\rangle$ corresponds to the uniform density $n$, as is easily calculated from the definitions (3.12) and (3.13).
Finally we have the repulsive ion-ion interaction

$$E_{\text{ii}} = \langle\Psi_0|\mathcal{H}_{\text{ii}}|\Psi_0\rangle = \frac{1}{2}\int d^3r\ d^3r' \frac{n^2e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{3.31}$$

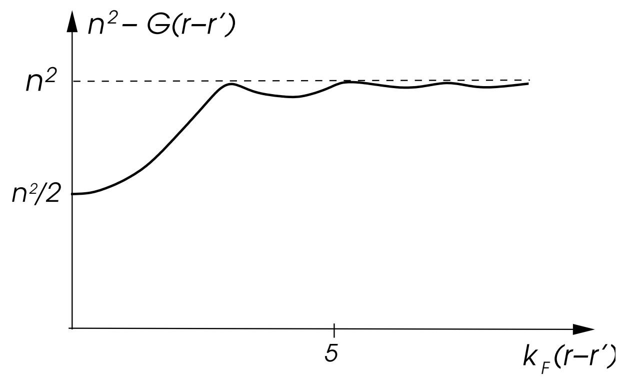



Figure 3.1: Pair correlation function.

It is easy to verify that the three contributions $E_{\text{Hartree}}$, $E_{\text{ei}}$, and $E_{\text{ii}}$ compensate each other to exactly zero. Note that these three terms are the only ones that would arise in a classical electrostatic calculation, implying that the stability of metals relies purely on quantum effect. The remaining terms are the kinetic energy and the Fock term. The latter is negative and reads

$$E_{\text{Fock}} = -\Omega\frac{9n^2}{4}\int d^3r \frac{e^2}{|\boldsymbol{r}|}\left(\frac{\sin k_F|\boldsymbol{r}| - k_F|\boldsymbol{r}|\cos k_F|\boldsymbol{r}|}{(k_F|\boldsymbol{r}|)^3}\right)^2 = -N\frac{3e^2}{4\pi}k_F. \tag{3.32}$$

Eventually, the total energy per electron is given by

$$\frac{E_g}{N} = \frac{3}{5}\frac{\hbar^2k_F^2}{2m} - \frac{3e^2}{4\pi}k_F = \left(\frac{2.21}{r_s^2} - \frac{0.916}{r_s}\right)\text{Ry} \tag{3.33}$$

where $1\text{Ry} = e^2/2a_B$ and the dimensionless quantity $r_s$ is defined via

$$n = \frac{3}{4\pi d^3} \tag{3.34}$$

and

$$r_s = \frac{d}{a_B} = \left(\frac{9\pi}{4}\right)^{1/3} \frac{me^2}{\hbar^2 k_F}. \tag{3.35}$$

The length $d$ is the average radius of the volume occupied by one electron. Minimizing the energy per electron with respect to $n$ is equivalent to minimize it with respect to $r_s$, yielding $r_{s,min} = 4.83\ d \approx 2.41$Å. This corresponds to a lattice constant of $a = (4\pi/3)^{1/3}d \approx 3.9$Å. This estimate is roughly in agreement with the lattice constants of the Alkali metals : $r_{s,\mathrm{Li}} = 3.22$, $r_{s,\mathrm{Na}} = 3.96$, $r_{s,\mathrm{K}} = 4.86$. Note that in metals the delocalized electrons are responsible for the cohesion of the positive background yielding a stable solid.
The good agreement of this simple estimate with the experimental values is due to the fact that the Alkali metals have only one valence electron in an s-orbital that is delocalized, whereas the the core electrons are in a noble gas configuration and, thus, relatively inert. In the variational approach outlined above correlation effects among the electrons due to the Coulomb repulsion have been neglected. In particular, electrons can be expected to 'avoid' each other not just because of the Pauli principle, but also as a result of the repulsive interaction. However, for the problem under consideration the correlation corrections turn out to be small for $r_s \sim r_{s,\mathrm{min}}$:

$$\frac{E_{\mathrm{tot}}}{N} = \left(\frac{2.21}{r_s^2} - \frac{0.916}{r_s} + 0.062\ln r_s - 0.096\ +\ \ldots\right)\mathrm{Ry} \tag{3.36}$$

which can be obtained from a more sophisticated quantum field theoretical analysis.

## 3.2 Charge excitations and the dielectric function

In analogy to semiconductors, the elementary excitations of metallic systems are the electron-hole excitations, which for metals, however, can have arbitrarily small energies. One particularly drastic consequence of this behavior is the strong screening of the long-ranged Coulomb potential. As we will see, a negative test charge in a metal reduces the electron density in its vicinity, and the induced cloud of positive charges, relative to the uniform charge density, weaken the Coulomb potential as,

$$V(r) \propto \frac{1}{r} \qquad \longrightarrow \qquad V'(r) \propto \frac{e^{-r/l}}{r} \tag{3.37}$$

i.e. the Coulomb potential is modified into the short-ranged Yukawa potential with screening length $l$. In contrast to metals, the finite energy gap for electron-hole excitations the charge distribution in semiconductors reduces the adaption of the system to perturbations, so that the screened Coulomb potential remains long-ranged,

$$V(r) \propto \frac{1}{r} \qquad \longrightarrow \qquad V'(r) \propto \frac{1}{\varepsilon r}. \tag{3.38}$$

As mentioned earlier, the semiconductor acts as a dielectric medium and its screening effects are accounted for by the polarization of localized electric dipoles, i.e., the Coulomb potential inside a semiconductor is renormalized by the dielectric constant $\varepsilon$.

### 3.2.1 Dielectric response and Lindhard function

We will now investigate the response of an electron gas to a time- and position-dependent weak external potential $V_a(\boldsymbol{r}, t)$ in more detail based on the equation of motion. We introduce the Hamiltonian

$$\mathcal{H} = \mathcal{H}_{\mathrm{kin}} + \mathcal{H}_V = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}^\dagger_{\boldsymbol{k}s} \widehat{c}_{\boldsymbol{k}s} + \sum_s \int d^3r\ V_a(\boldsymbol{r}, t) \widehat{\Psi}^\dagger_s(\boldsymbol{r}) \widehat{\Psi}_s(\boldsymbol{r}) \tag{3.39}$$

where the second term is considered as a small perturbation. In a first step we consider the linear response of the system to the external potential. On this level we restrict ourself to one Fourier component in the spatial and time dependence of the potential,

$$V_a(\boldsymbol{r},t) = V_a(\boldsymbol{q},\omega)e^{i\boldsymbol{q}\cdot\boldsymbol{r}-i\omega t}e^{\eta t}, \tag{3.40}$$

where $\eta \to 0_+$ includes the adiabatic switching on of the potential. To linear response this potential induces a small modulation of the electron density of the form $n_{\text{ind}}(\boldsymbol{r},t) = n_0 + \delta n_{\text{ind}}(\boldsymbol{r},t)$ with

$$\delta n_{\text{ind}}(\boldsymbol{r},t) = \delta n_{\text{ind}}(\boldsymbol{q},\omega)e^{i\boldsymbol{q}\cdot\boldsymbol{r}-i\omega t}. \tag{3.41}$$

Using equations (3.12) and (3.13) we obtain for the density operator in momentum space,

$$\widehat{\rho}_{\boldsymbol{q}} = \sum_s \int d^3r \widehat{\Psi}_s^\dagger(\boldsymbol{r})\widehat{\Psi}_s(\boldsymbol{r})e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\widehat{c}^\dagger_{\boldsymbol{k}+qs}\widehat{c}_{\boldsymbol{k}s} = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}, \tag{3.42}$$

where we define $\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s} = \widehat{c}^\dagger_{\boldsymbol{k}+qs}\widehat{c}_{\boldsymbol{k}s}$. The perturbation term $\mathcal{H}_V$ now reads

$$\mathcal{H}_V = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\widehat{\rho}_{\boldsymbol{k},-\boldsymbol{q},s}V_a(\boldsymbol{q},\omega)e^{-i\omega t}e^{\eta t}. \tag{3.43}$$

The density operator $\widehat{\rho}_{\boldsymbol{q}}(t)$ in Heisenberg representation is the relevant quantity needed to describe the electron density in the metal. We introduce the equation of motion for $\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}(t)$:

$$i\hbar\frac{d}{dt}\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s} = \left[\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s},\mathcal{H}\right] = \left[\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s},\mathcal{H}_{\text{kin}}+\mathcal{H}_V\right] \tag{3.44}$$

$$= \left(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}\right)\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s} + \left(\widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s} - \widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q}s}\widehat{c}_{\boldsymbol{k}+\boldsymbol{q}s}\right)V_a(\boldsymbol{q},\omega)e^{-i\omega t}e^{\eta t}. \tag{3.45}$$

We now take the thermal average $\langle\widehat{A}\rangle = \text{Tr}[\widehat{A}e^{-\beta\mathcal{H}}]/\text{Tr}[e^{-\beta\mathcal{H}}]$ and follow the linear response scheme by assuming the same time dependence for $\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}(t)$ as for the potential , so that the equation of motion reads,

$$(\hbar\omega + i\hbar\eta)\langle\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}\rangle = \left(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}\right)\langle\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}\rangle + \left(n_{0\boldsymbol{k},s}-n_{0\boldsymbol{k}+\boldsymbol{q},s}\right)V_a(\boldsymbol{q},\omega) \tag{3.46}$$

where $n_{0\boldsymbol{k},s} = \langle\widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s}\rangle$ and, therefore,

$$\delta n_{\text{ind}}(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\langle\widehat{\rho}_{\boldsymbol{k},\boldsymbol{q},s}\rangle = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\frac{n_{0\boldsymbol{k}+\boldsymbol{q},s}-n_{0\boldsymbol{k},s}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}-\hbar\omega-i\hbar\eta}V_a(\boldsymbol{q},\omega). \tag{3.47}$$

With this, we define the dynamical linear response function as

$$\chi_0(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\frac{n_{0\boldsymbol{k}+\boldsymbol{q},s}-n_{0\boldsymbol{k},s}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}-\hbar\omega-i\hbar\eta} \tag{3.48}$$

such that $\delta n_{\text{ind}}(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)V_a(\boldsymbol{q},\omega)$, where $\chi_0(\boldsymbol{q},\omega)$ is known to be the Lindhard function. So far we treated the linear response of the system to an external perturbation without considering "feedback effects" due to the interaction among electrons. In fact, the density fluctuation $\delta n(\boldsymbol{r},t)$ can be thought as a source for an additional Coulomb potential $V_{\delta n}$ which can be determined by means of the Poisson equation,

$$\boldsymbol{\nabla}^2 V_{\delta n}(\boldsymbol{r},t) = -4\pi e^2\delta n(\boldsymbol{r},t) \tag{3.49}$$

or in Fourier space

$$V_{\delta n}(\boldsymbol{q},\omega) = \frac{4\pi e^2}{q^2}\delta n(\boldsymbol{q},\omega). \tag{3.50}$$

If we allow feedback effects in our system with external perturbation $V_a(\boldsymbol{q},\omega)$, the effective potential $V$ felt by the electrons is determined self-consistently via

$$V(\boldsymbol{q},\omega) = V_a(\boldsymbol{q},\omega) + V_{\delta n}(\boldsymbol{q},\omega) \tag{3.51}$$

$$= V_a(\boldsymbol{q},\omega) + \frac{4\pi e^2}{q^2}\delta n(\boldsymbol{q},\omega), \tag{3.52}$$

where

$$\delta n(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)V(\boldsymbol{q},\omega). \tag{3.53}$$

The relation between $V$ and $V_a$ may then be written as

$$V(\boldsymbol{q},\omega) = \frac{V_a(\boldsymbol{q},\omega)}{\varepsilon(\boldsymbol{q},\omega)} \tag{3.54}$$

with

$$\varepsilon(\boldsymbol{q},\omega) = 1 - \frac{4\pi e^2}{q^2}\chi_0(\boldsymbol{q},\omega), \tag{3.55}$$

where $\varepsilon(\boldsymbol{q},\omega)$ is termed the dynamical dielectric function and describes the renormalization of the external potential due to the dynamical response of the electrons in the metal. Extending Eq.(3.53) to

$$\delta n(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)V(\boldsymbol{q},\omega) = \chi(\boldsymbol{q},\omega)V_a(\boldsymbol{q},\omega). \tag{3.56}$$

we define the response function $\chi(\boldsymbol{q},\omega)$ within ”random phase approximation” [4] to be

$$\chi(\boldsymbol{q},\omega) = \frac{\chi_0(\boldsymbol{q},\omega)}{\varepsilon(\boldsymbol{q},\omega)} = \frac{\chi_0(\boldsymbol{q},\omega)}{1 - \frac{4\pi e^2}{q^2}\chi_0(\boldsymbol{q},\omega)} > \tag{3.58}$$

This response function $\chi(\boldsymbol{q},\omega)$ contains also effects of electron-electron interaction and comprises information not only about the renormalization of potentials, but also on the excitation spectrum of the metal.

[4] The equation (3.58) can be written in the form of a geometric series,

$$\chi(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)\left[1 + \frac{4\pi e^2}{q^2}\chi_0(\boldsymbol{q},\omega) + \left(\frac{4\pi e^2}{q^2}\chi_0(\boldsymbol{q},\omega)\right)^2 + \cdots\right]. \tag{3.57}$$

From the point of view of perturbation theory, this series corresponds to summing a limited subset of perturbative terms to infinite order. This approximation is called Random Phase Approximation (RPA) and is based on the assumption the phase relation between different particle-hole excitations entering the perturbation series are random such that interference terms vanish on the average. This approximation is used quite frequently, in particular, in the discussion of instabilities of a system towards an ordered phase.

### 3.2.2 Electron-hole excitation

For simple particle-hole excitations in metals, neglecting Coulomb interaction between the electrons, it is sufficient to study the bare response function $\chi_0(\boldsymbol{q},\omega)$. We may separate $\chi_0$ into its real and imaginary part, $\chi_0(\boldsymbol{q},\omega) = \chi_{01}(\boldsymbol{q},\omega) + i\chi_{02}(\boldsymbol{q},\omega)$. Using the relation

$$\lim_{\eta \to 0_+} \frac{1}{z - i\eta} = \mathcal{P}\left(\frac{1}{z}\right) + i\pi\delta(z) \tag{3.59}$$

where the Cauchy principal value $\mathcal{P}$ of the first term has to be taken, we separate the Lindhard function (3.48) into

$$\chi_{01}(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s} \mathcal{P}\left(\frac{n_{0,\boldsymbol{k}+\boldsymbol{q}} - n_{0,\boldsymbol{k}}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega}\right) \tag{3.60}$$

$$\chi_{02}(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s} (n_{0,\boldsymbol{k}+\boldsymbol{q}} - n_{0,\boldsymbol{k}})\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega) \tag{3.61}$$

The real part will be important later in the context of instabilities of metals. The excitation spectrum is visible in the imaginary part which relates to the absorption of energy by the electrons subject to a time-dependent external perturbation.[5] Note that $\chi_{02}(\boldsymbol{q},\omega)$ corresponds to Fermi's golden rule known from time-dependent perturbation theory, i.e. the transition rate from the ground state to an excited state of energy $\hbar\omega$ and momentum $\boldsymbol{q}$.

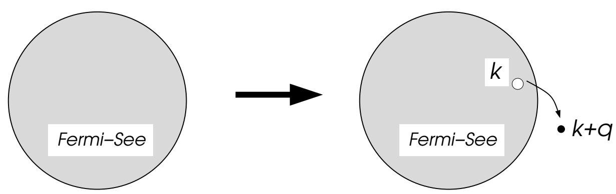



Figure 3.2: Schematic view of an particle-hole pair creation (electron-hole excitation).

The relevant excitations originating from the Lindhard function are particle-hole excitations. Starting from the ground state of a completely filled Fermi sea, one electron with momentum $\boldsymbol{k}$ is removed and inserted again outside the Fermi sea in some state with momentum $\boldsymbol{k}+\boldsymbol{q}$ (see Figure 3.2). The energy difference is then given by

$$E_{\boldsymbol{k},\boldsymbol{q}} = \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} > 0. \tag{3.62}$$

In analogy to the semiconducting case, there is a continuum of particle-hole excitation spectrum in the energy-momentum plane – sketched in Figure 3.3. Note the absence of an energy gap for excitations.

### 3.2.3 Collective excitation

For the electron-hole excitations the Coulomb interaction was ignored (by using $\chi_0(\boldsymbol{q},\omega)$ instead of $\chi(\boldsymbol{q},\omega)$), such that the bare Lindhard function provides information about the single particle spectrum. Including the Coulomb interaction a new collective excitation will arise, the so-called plasma resonance. For a long-ranged interaction like the Coulomb interaction this resonance appears at finite frequency for small momenta $\boldsymbol{q}$. We derive it here using the response function

[5] See Chapter 6 "Linear response theory" of the course "Statistical Physics" FS09.

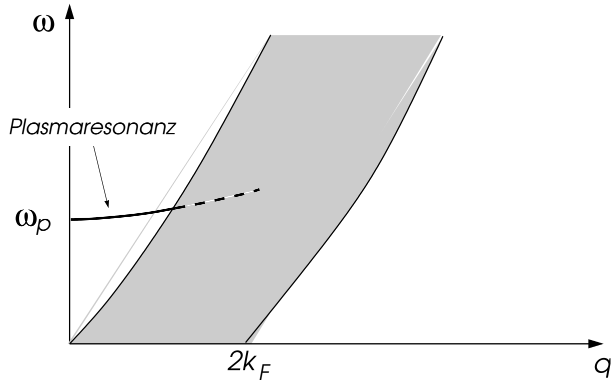



Figure 3.3: Excitation spectrum in the $\omega$-$q$-plane. The large shaded region corresponds to the electron-hole continuum and the sharp line outside the continuum represents the plasma resonance which is damped when entering the continuum.

$\chi(\boldsymbol{q},\omega)$. Assuming $|\boldsymbol{q}| \ll k_F$ we expand $\chi_0(\boldsymbol{q},\omega)$ in $\boldsymbol{q}$, starting with

$$\epsilon_{\boldsymbol{k}+\boldsymbol{q}} = \epsilon_{\boldsymbol{k}} + \boldsymbol{q}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\,\epsilon_{\boldsymbol{k}} + \cdots \tag{3.63}$$

$$n_{0,\boldsymbol{k}+\boldsymbol{q}} = n_{0,\boldsymbol{k}} + \frac{\partial n_0}{\partial\epsilon}\boldsymbol{q}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\,\epsilon_{\boldsymbol{k}} + \cdots \tag{3.64}$$

Note that $\partial n_0/\partial\epsilon_{\boldsymbol{k}} = -\delta(\epsilon_{\boldsymbol{k}} - \epsilon_F)$ at $T=0$ and $\boldsymbol{\nabla}_{\boldsymbol{k}}\,\epsilon_{\boldsymbol{k}} = \hbar\boldsymbol{v}_{\boldsymbol{k}}$ is the velocity. Since we will deal with states located at the Fermi energy here, $\boldsymbol{v}_{\boldsymbol{k}} = v_F\boldsymbol{k}/|\boldsymbol{k}|$ is the Fermi velocity. This leads to the approximation

$$\begin{aligned}\chi_0(\boldsymbol{q},\omega) &\approx -2\int\frac{d^3k}{(2\pi)^3}\frac{\boldsymbol{q}\cdot\boldsymbol{v}_F\,\delta(\epsilon_{\boldsymbol{k}}-\mu)}{\boldsymbol{q}\cdot\boldsymbol{v}_F - \omega - i\eta}\\ &= \frac{2}{(2\pi)^2}\int d\cos\theta\,\frac{k_F^2}{\hbar v_F}\left[\frac{qv_F\cos\theta}{\omega+i\eta} + \left(\frac{qv_F\cos\theta}{\omega+i\eta}\right)^2 + \left(\frac{qv_F\cos\theta}{\omega+i\eta}\right)^3 + \cdots\right] &(3.65)\\ &\approx \frac{k_F^3q^2}{3\pi^2m(\omega+i\eta)^2}\left(1+\frac{3}{5}\frac{v_F^2q^2}{(\omega+i\eta)^2}\right) &(3.66)\\ &= \frac{n_0q^2}{m(\omega+i\eta)^2}\left(1+\frac{3}{5}\frac{v_F^2q^2}{(\omega+i\eta)^2}\right). &(3.67)\end{aligned}$$

According to equation (3.55), we find

$$\lim_{|\boldsymbol{q}|\to 0}\varepsilon(\boldsymbol{q},\omega) = 1 - \frac{\omega_p^2}{\omega^2} \tag{3.68}$$

for the dielectric function in the long wavelength limit ($|\boldsymbol{q}| \to 0$), with

$$\omega_p^2 = \frac{4\pi e^2 n_0}{m}. \tag{3.69}$$

We now use the result in Eq.(3.67) to approximate $\chi(\boldsymbol{q},\omega)$,

$$\chi(\boldsymbol{q},\omega) \approx \frac{n_0q^2R(q,\omega)^2}{m(\omega+i\eta)^2 - 4\pi e^2n_0^2R(q,\omega)^2} \tag{3.70}$$

$$= \frac{n_0q^2R(q,\omega)}{2m\omega_p}\left\{\frac{1}{\omega+i\eta-\omega_pR(q,\omega)} - \frac{1}{\omega+i\eta+\omega_pR(q,\omega)}\right\} \tag{3.71}$$

where we introduced

$$R(q,\omega)^2 = \left(1 + \frac{3v_F^2 q^2}{5\omega^2}\right). \tag{3.72}$$

Applying the relation (3.59) in Eq.(3.67) we obtain

$$\mathrm{Im}\big[\chi(\boldsymbol{q},\omega)\big] \approx \frac{\pi n_0 q^2 R(q,\omega_p)}{\omega_p}\Big[\delta(\omega - \omega_p R(q,\omega_p)) - \delta(\omega + \omega_p R(q,\omega_p))\Big] \tag{3.73}$$

which yields a sharp excitation mode,

$$\omega(\boldsymbol{q}) = \omega_p R(q,\omega_p) = \omega_p \left\{1 + \frac{3v_F^2 q^2}{10\omega_p^2} + \cdots\right\}, \tag{3.74}$$

which is called *plasma resonance* with $\omega_p$ as the *plasma frequency.* Similar to the exciton, the plasma excitation has a well-defined energy-momentum relation and may consequently be viewed as a quasiparticle (*plasmon*) which has bosonic character. When the plasmon dispersion merges with the electron-hole continuum it is damped (Landau damping) because of the allowed decay into electron-hole excitations. This results in a finite life-time of the plasmons within the electron-hole continuum corresponding to a finite width of the resonance of the collective excitation.

| metal | $\omega_p^{(\mathrm{exp})}$ [eV] | $\omega_p^{(\mathrm{theo})}$ [eV] |
|---|---|---|
| Li | 7.1 | 8.5 |
| Na | 5.7 | 6.2 |
| K | 3.7 | 4.6 |
| Mg | 10.6 | - |
| Al | 15.3 | - |

Table 3.1: Experimental values of the plasma frequency for different compounds. For the alkali metals a theoretically determined $\omega_p$ is given for comparison, using equation (3.69) with $m$ the free electron mass and $n$ determined through $r_{s,\mathrm{Li}} = 3.22$, $r_{s,\mathrm{Na}} = 3.96$ and $R_{s,\mathrm{K}} = 4.86$.

It is possible to understand the plasma excitation in a classical picture. Consider negatively charged electrons in a positively charged ionic background. When the electrons are shifted uniformly by $\boldsymbol{r}$ with respect to the ions, a polarization $\boldsymbol{P} = -n_0 e\boldsymbol{r}$ results. The polarization causes an electric field $\boldsymbol{E} = -4\pi\boldsymbol{P}$ which acts as a restoring force. The equation of motion for an individual electron describes harmonic oscillations

$$m\frac{d^2}{dt^2}\boldsymbol{r} = -e\boldsymbol{E} = -4\pi e^2 n_0 \boldsymbol{r}. \tag{3.75}$$

with the same oscillation frequency as in Eq.(3.69), the plasma frequency,

$$\omega_p^2 = \frac{4\pi e^2 n_0}{m}. \tag{3.76}$$

Classically, the plasma resonance can therefore be thought as an oscillation of the whole electron gas cloud on top of a positively charged background.

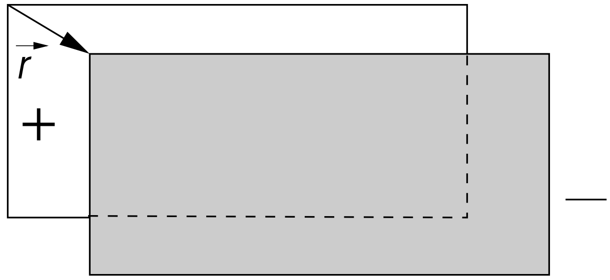



Figure 3.4: Classical understanding of the plasma excitation.

### 3.2.4 Screening

**Thomas-Fermi screening**

Next, we analyze the potential $V$ felt by the electrons exposed to a static field ($\omega \to 0$). Using the expansion (3.64) we obtain

$$\chi_0(\boldsymbol{q}, 0) = -\frac{1}{\Omega}\sum_{\boldsymbol{k},s}\delta(\epsilon_{\boldsymbol{k}} - \epsilon_F) = -\frac{1}{\pi^2}\frac{k_F^2}{\hbar v_F} = -\frac{3n_0}{2\epsilon_F} \tag{3.77}$$

and thus

$$\varepsilon(\boldsymbol{q}, 0) = 1 + \frac{k_{TF}^2}{q^2} \tag{3.78}$$

with the so-called Thomas-Fermi wave vector $k_{TF}^2 = 6\pi e^2 n_0/\epsilon_F$. The effect of the renormalized $\boldsymbol{q}$-dependence of the dielectric function can best be understood by considering a bare point charge $V_a(\boldsymbol{r}) = e^2/r$ (or $V_a(\boldsymbol{q}) = 4\pi e^2/q^2$) and its renormalization in momentum space

$$V(\boldsymbol{q}) = \frac{V_a(\boldsymbol{q})}{\epsilon(\boldsymbol{q}, 0)} = \frac{4\pi e^2}{q^2 + k_{TF}^2} \tag{3.79}$$

or in real space

$$V(\boldsymbol{r}) = \frac{e^2}{r}e^{-k_{TF}r}. \tag{3.80}$$

The potential is screened by a rearrangement of the electrons and this turns the long-ranged Coulomb potential into a Yukawa potential with exponential decay. The new length scale is $k_{TF}^{-1}$, the so-called Thomas-Fermi screening length. In ordinary metals $k_{TF}$ is typically of the same order of magnitude as $k_F$, i.e. the screening length is of order 5Å comparable to the distance between neighboring atoms.[6] As a consequence also external electric fields cannot penetrate a metal, but are screened on this length $1/k_{TF}$. This legitimates one of the basic assumptions used in electrostatics with metals.

[6]The Thomas-Fermi approach for electron gas is sketched in the following. The Thomas-Fermi theory for the charge distributions slowly varying in space is based on the approximation that locally the electrons form a Fermi gas with Fermi energy $\epsilon_F$ and electron density $n_e(\epsilon_F)$ neutralizing the ionic background. The electrostatic potential $\Phi(\boldsymbol{r})$ of an external charge distribution $\rho_{\rm ex}(\boldsymbol{r})$ induces a charge redistribution $\rho_{\rm ind}(\boldsymbol{r})$ relative to $n_e(\epsilon_F)$. Within Thomas-Fermi approximation the induced charge distribution can then be written as

$$\rho_{\rm ind}(\boldsymbol{r}) = -e\big[n_e(\epsilon_F + e\Phi(\vec{r})) - n_e(\epsilon_F)\big] \tag{3.81}$$

with

$$n_e(\epsilon_F) = \frac{k_F^3}{3\pi^2} = \frac{1}{3\pi^2\hbar^2}(2m\epsilon_F)^{3/2} \tag{3.82}$$

where $\epsilon_F = \hbar^2 k_F^2/2m$. This approach is justified, if the spacial change of the potential $\Phi(\boldsymbol{r})$ is slow compared to $k_F^{-1}$, so that locally we may describe the electron gas as a filled Fermi sphere of corresponding electron density.

**Friedel oscillations**

The static dielectric function can be evaluated exactly for a system of free electrons, resulting for 3 dimensions in

$$\varepsilon(\boldsymbol{q},0) = 1 + \frac{4e^2 m k_F}{\pi q^2}\left\{\frac{1}{2} + \frac{4k_F^2 - q^2}{8k_F q}\ln\left|\frac{2k_F + q}{2k_F - q}\right|\right\}. \tag{3.87}$$

Noticeably the dielectric function varies little for small $\boldsymbol{q} \ll k_F$. At $q = \pm 2k_F$ there is, however, a logarithmic singularity. This is a consequence of the sharpness of the Fermi surface in $k$-space. Consider the induced charge of a point charge at the origin: $n_a(r) = e\delta(\boldsymbol{r})$.[7]

$$\delta n(\boldsymbol{r}) = e\int \frac{d^3q}{(2\pi)^3}\left\{\frac{1}{\varepsilon(q)} - 1\right\} n_a(\boldsymbol{q},0)e^{i\boldsymbol{q}\cdot\boldsymbol{r}} = -\frac{e}{r}\int_0^\infty g(q)n_a(\boldsymbol{q},0)\sin qr\, dq \tag{3.89}$$

with

$$g(q) = \frac{q}{2\pi^2}\frac{\varepsilon(q) - 1}{\varepsilon(q)}. \tag{3.90}$$

Note that $g(q)$ vanishes for both $q \to 0$ and $q \to \infty$. Using partial integration twice, we find

$$\delta n(\boldsymbol{r}) = \frac{e}{r^3}\int_0^\infty g''(q)\sin qr dq \tag{3.91}$$

where

$$g'(q) \approx A\,\ln|q - 2k_F| \tag{3.92}$$

and

$$g''(q) \approx \frac{A}{q - 2k_F} \tag{3.93}$$

---

The Poisson equation may now be formulated as

$$\boldsymbol{\nabla}^2\Phi(\boldsymbol{r}) = -4\pi[\rho_{\text{ind}}(\boldsymbol{r}) + \rho_{\text{ex}}(\boldsymbol{r})] \approx 4\pi\left[e^2\Phi(\boldsymbol{r})\left.\frac{\partial n_e(\epsilon)}{\partial\epsilon}\right|_{\epsilon=\epsilon_F} - \rho_{\text{ex}}(\boldsymbol{r})\right] \tag{3.83}$$

$$= k_{TF}^2\Phi(\boldsymbol{r}) - 4\pi\rho_{\text{ex}}(\boldsymbol{r}) \tag{3.84}$$

with the Thomas-Fermi momentum $k_{TF}$ defined as,

$$k_{TF}^2 = 4\pi e^2\left.\frac{\partial n_e(\epsilon)}{\partial\epsilon}\right|_{\epsilon=\epsilon_F} = \frac{6\pi e^2 n_e}{\epsilon_F}. \tag{3.85}$$

and $n_e = n_e(\epsilon_F)$. For a point charge $Q$ located at the origin we obtain,

$$\Phi(\boldsymbol{r}) = Q\,\frac{e^{-rk_{TF}}}{r}. \tag{3.86}$$

This is the Yukawa potential as obtained above.

[7] The charge distribution can be deduced from the Poisson equation (3.50):

$$\delta n(\boldsymbol{q}) = \frac{q^2}{4\pi e^2}V_{\delta n}(\boldsymbol{q}) = \chi_0(\boldsymbol{q},0)V(\boldsymbol{q}) = \chi_0(\boldsymbol{q},0)\frac{V_a(\boldsymbol{q})}{\epsilon(\boldsymbol{q},0)} = \frac{1-\epsilon(\boldsymbol{q},0)}{\epsilon(\boldsymbol{q},0)}n_a(\boldsymbol{q},0) \tag{3.88}$$

The charge distribution in real space can be obtained by Fourier transformation.

dominate around $q \sim 2k_F$. Hence, for $k_F r \ll 1$,

$$\delta n(r) \approx \frac{eA}{r^3} \int_{2k_F-\Lambda}^{2k_F+\Lambda} \frac{\sin[(q-2k_F)r]\cos 2k_F r + \cos[(q-2k_F)r] sin 2k_F r}{q-2k_F} dq \tag{3.94}$$

$$\longrightarrow \pi e A \frac{\cos 2k_F r}{r^3}. \tag{3.95}$$

with a cutoff $\Lambda \to \infty$. The induced charge distribution exhibits so-called *Friedel oscillations*.

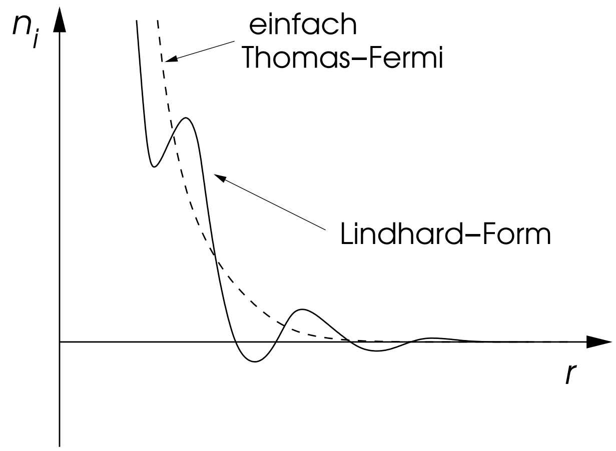



Figure 3.5: Friedel oscillations of the charge distribution.

**Dielectric function in various dimensions**

Above we have treated the dielectric function for a three-dimensional parabolic band. Similar calculations can be performed for one- and two-dimensional systems. In general, the static susceptibility is given by

$$\chi_0(q, \omega = 0) = \begin{cases} -\frac{1}{2\pi q} \ln \left| \frac{s+2}{s-2} \right|, & \text{1D} \\ -\frac{1}{2\pi} \left\{ 1 - \left(1 - \frac{4}{s^2}\right) \theta(s-2) \right\}, & \text{2D} \\ -\frac{k_F}{2\pi^2} \left\{ 1 - \frac{s}{4} \left(1 - \frac{4}{s^2}\right) \ln \left| \frac{s+2}{s-2} \right| \right\}, & \text{3D} \end{cases} \tag{3.96}$$

where $s = q/k_F$. Interestingly $\chi_0(q, 0)$ has a singularity at $q = 2k_F$ in all dimensions. The singularity becomes weaker as the dimensionality is increased. In one dimension, there is a logarithmic divergence, in two dimensions there is a kink, and in three dimensions only the derivative diverges. Later we will see that these singularities may lead to instabilities of the metallic state, in particular for the one-dimensional case.

## 3.3 Lattice vibrations - Phonons

The atoms in a lattice of a solid are not immobile but vibrate around their equilibrium positions. We will describe this new degree of freedom by treating the lattice as a continuous elastic medium (Jellium with elastic modulus $\lambda$). This approximation is sufficient to obtain some essential

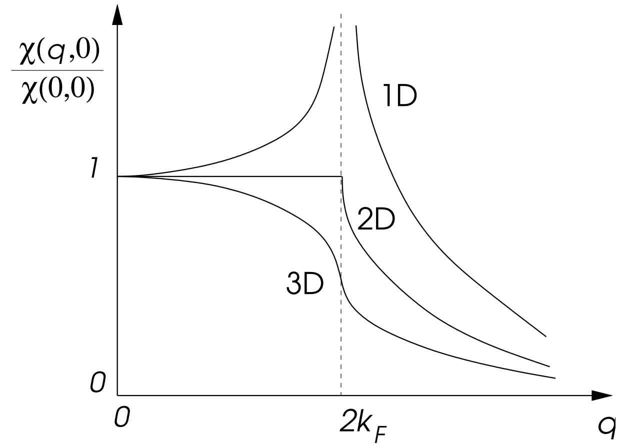



Figure 3.6: Lindhard functions for different dimensions. The lower the dimension the stronger the singularity at $q = 2k_F$.

features of the interaction between lattice vibrations and electrons. In particular, renormalized screening effects will be found. Our approach here is, however, limited to mono-atomic unit cells because the internal structure of a unit cell is neglected.

### 3.3.1 Vibration of a isotropic continuous medium

The deformation of an elastic medium can be described by the displacement of the infinitesimal volume element $d^3r$ around a point $\boldsymbol{r}$ to a different point $\boldsymbol{r}'(\boldsymbol{r})$. We can introduce here the so-called *displacement field* $\boldsymbol{u}(\boldsymbol{r}) = \boldsymbol{r}'(\boldsymbol{r}) - \boldsymbol{r}$ as function of $\boldsymbol{r}$. In general, $\boldsymbol{u}$ is also a function of time. In the simplest form of an isotropic medium the elastic energy for small deformations is given by

$$E_{\text{el}} = \frac{\lambda}{2} \int d^3r \; (\boldsymbol{\nabla} \cdot \boldsymbol{u}(\boldsymbol{r}, t))^2 \tag{3.97}$$

where $\lambda$ is the elastic modulus (note that there is no deformation energy, if the medium is just shifted uniformly). This energy term produces a restoring force trying to bring the system back to the undeformed state. In this model we are neglecting the shear contributions.[8] The continuum form above is valid for deformation wavelengths that are much longer than the lattice constant, so that details of the arrangement of atoms in the lattice can be neglected. The kinetic energy of the motion of the medium is given by

$$E_{\text{kin}} = \frac{\rho_0}{2} \int d^3r \; \left( \frac{\partial \boldsymbol{u}(\boldsymbol{r}, t)}{\partial t} \right)^2 \tag{3.99}$$

where $\rho_0 = M_i n_i$ is the mass density with the ionic mass $M_i$ and the ionic density $n_i$. Variation of the Lagrangian functional $L[\boldsymbol{u}] = E_{\text{kin}} - E_{\text{el}}$ with respect to $\boldsymbol{u}(\boldsymbol{r}, t)$ leads to the equation of motion

$$\frac{1}{c_s^2} \frac{\partial^2}{\partial t^2} \boldsymbol{u}(\boldsymbol{r}, t) - \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \boldsymbol{u}(\boldsymbol{r}, t)) = 0, \tag{3.100}$$

[8]Note that the most general form of the elastic energy of an isotropic medium takes the form

$$E_{\text{el}} = \int d^3r \sum_{\alpha,\beta=x,y,z} \left[ \frac{\lambda}{2} (\partial_\alpha u_\alpha)(\partial_\beta u_\beta) + \mu (\partial_\alpha u_\beta)(\partial_\alpha u_\beta) \right], \tag{3.98}$$

where $\partial_\alpha = \partial / \partial r_\alpha$. The *Lamé coefficients* $\lambda$ and $\mu$ characterize the elastic properties. The elastic constant $\lambda$ describes density modulations leading to longitudinal elastic waves, whereas $\mu$ corresponds to shear deformations connected with transversely polarized elastic waves. Note that transverse elastic waves are not important for the coupling of electrons and lattice vibrations.

which is a wave equation with sound velocity $c_s^2 = \lambda/\rho_0$. The resulting displacement field can be expanded into normal modes,

$$\boldsymbol{u}(\boldsymbol{r},t) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}} \left( q_{\boldsymbol{k}}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} + q_{\boldsymbol{k}}(t)^* e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \right) \tag{3.101}$$

where every $q_{\boldsymbol{k}}(t)$ satisfies the equation

$$\frac{d^2}{dt^2} q_{\boldsymbol{k}} + \omega_{\boldsymbol{k}}^2 q_{\boldsymbol{k}} = 0, \tag{3.102}$$

with the frequency $\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}| = c_s k$ and the polarization vector $\boldsymbol{e}_{\boldsymbol{k}}$ has unit length. Note that within our simplification for the elastic energy (3.98), all modes correspond to longitudinal waves, i.e. $\boldsymbol{\nabla} \times \boldsymbol{u}(\boldsymbol{r},t) = 0$ and $\boldsymbol{e}_{\boldsymbol{k}} \parallel \boldsymbol{k}$. The total energy expressed in terms of the normal modes reads

$$E = \sum_{\boldsymbol{k}} \rho_0 \omega_{\boldsymbol{k}}^2 \left[ q_{\boldsymbol{k}}(t) q_{\boldsymbol{k}}^*(t) + q_{\boldsymbol{k}}^*(t) q_{\boldsymbol{k}}(t) \right]. \tag{3.103}$$

Next, we switch from a Lagrangian to a Hamiltonian description by defining the new variables

$$Q_{\boldsymbol{k}} = \sqrt{\rho_0}(q_{\boldsymbol{k}} + q_{\boldsymbol{k}}^*) \tag{3.104}$$

$$P_{\boldsymbol{k}} = \frac{d}{dt} Q_{\boldsymbol{k}} = -i\omega_{\boldsymbol{k}} \sqrt{\rho_0}(q_{\boldsymbol{k}} - q_{\boldsymbol{k}}^*) \tag{3.105}$$

in terms of which the energy is given by

$$E = \frac{1}{2} \sum_{\boldsymbol{k}} \left( P_{\boldsymbol{k}}^2 + \omega_{\boldsymbol{k}}^2 Q_{\boldsymbol{k}}^2 \right). \tag{3.106}$$

Thus, the system is equivalent to an ensemble of independent harmonic oscillators, one for each normal mode $\boldsymbol{k}$. Consequently, the system may be quantized by defining the canonical conjugate operators $P_{\boldsymbol{k}} \to \widehat{P}_{\boldsymbol{k}}$ and $Q_{\boldsymbol{k}} \to \widehat{Q}_{\boldsymbol{k}}$ which obey, by definition, the commutation relation,

$$[\widehat{Q}_{\boldsymbol{k}}, \widehat{P}_{\boldsymbol{k}'}] = i\hbar \delta_{\boldsymbol{k},\boldsymbol{k}'}. \tag{3.107}$$

As it is usually done for quantum harmonic oscillators, we define the raising and lowering operators

$$\widehat{b}_{\boldsymbol{k}} = \frac{1}{\sqrt{2\hbar\omega_{\boldsymbol{k}}}} \left( \omega_{\boldsymbol{k}} \widehat{Q}_{\boldsymbol{k}} + i\widehat{P}_{\boldsymbol{k}} \right) \tag{3.108}$$

$$\widehat{b}_{\boldsymbol{k}}^\dagger = \frac{1}{\sqrt{2\hbar\omega_{\boldsymbol{k}}}} \left( \omega_{\boldsymbol{k}} \widehat{Q}_{\boldsymbol{k}} - i\widehat{P}_{\boldsymbol{k}} \right), \tag{3.109}$$

satisfying the commutation relations

$$[\widehat{b}_{\boldsymbol{k}}, \widehat{b}_{\boldsymbol{k}'}^\dagger] = \delta_{\boldsymbol{k},\boldsymbol{k}'}, \tag{3.110}$$

$$[\widehat{b}_{\boldsymbol{k}}, \widehat{b}_{\boldsymbol{k}'}] = 0, \tag{3.111}$$

$$[\widehat{b}_{\boldsymbol{k}}^\dagger, \widehat{b}_{\boldsymbol{k}'}^\dagger] = 0. \tag{3.112}$$

These relations can be interpreted in a way that these operators create and annihilate quasi-particles following the Bose-Einstein statistics. According to the correspondence principle, the quantum mechanical Hamiltonian corresponding to the energy (3.106) is

$$\mathcal{H} = \sum_{\boldsymbol{k}} \hbar\omega_{\boldsymbol{k}} \left( \widehat{b}_{\boldsymbol{k}}^\dagger \widehat{b}_{\boldsymbol{k}} + \frac{1}{2} \right). \tag{3.113}$$

In analogy to the treatment of the electrons in second quantization we say that the operators $\widehat{b}_{\boldsymbol{k}}^{\dagger}$ ($\widehat{b}_{\boldsymbol{k}}$) create (annihilate) a *phonon*, a quasiparticle with well-defined energy-momentum relation, $\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}|$. Using Eqs.(3.102, 3.105, and 3.109) the displacement field operator $\widehat{\boldsymbol{u}}(\boldsymbol{r})$ can now be defined as

$$\widehat{\boldsymbol{u}}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}} \sqrt{\frac{\hbar}{2\rho_0 \omega_{\boldsymbol{k}}}} \left[ \widehat{b}_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} + \widehat{b}_{\boldsymbol{k}}^{\dagger} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \right]. \tag{3.114}$$

As mentioned above, the continuum approximation is valid for long wavelengths (small $\boldsymbol{k}$) only. For wavevectors with $\boldsymbol{k} \approx \pi/a$ the discreteness of the lattice appears in the form of corrections to the linear dispersion $\omega_{\boldsymbol{k}} \propto |\boldsymbol{k}|$. Since the number of degrees of freedom is limited to $3N_i$ ($N_i$ number of atoms), there is a maximal wave vector called the Debye wavevector[9] $k_D$. We can now define the corresponding Debye frequency $\omega_D = c_s k_D$ and the Debye temperature $\Theta_D = \hbar\omega_D/k_B$. In the continuous medium approximation there are only acoustic phonons. For the inclusion of optical phonons, the arrangement of the atoms within a unit cell has to be considered, which goes beyond this simple picture.

### 3.3.2 Phonons in metals

The consideration above is certainly valid for semiconductors, where ionic interactions are mediated via covalent chemical bonds and oscillations around the equilibrium position may be approximated by a harmonic potential, so that the form of the elastic energy above is well motivated. The situation is more subtle for metals, where the ions interact through the long-ranged Coulomb interaction and are held to together through an intricate interplay with the mobile conduction electrons.

First, neglecting the gluey effect of the electrons, the positively charged background can itself be treated as an ionic gas. Similar to the electronic gas (3.69), the background exhibits a well-defined collective plasma excitation at the ionic plasma frequency

$$\Omega_p^2 = \frac{4\pi n_i (Z_i e)^2}{M_i}, \tag{3.115}$$

For equation (3.115) we used the formula (3.69) with $n_0 \to n_i = n_0/Z_i$ the density of ions with charge number $Z_i$, $e \to Z_i e$, and $m \to M_i$ the atomic mass. Apparently the excitation energy does not vanish as $\boldsymbol{k} \to 0$. So far, the background of the metallic system can not be described as an elastic medium where the excitation spectrum is expected to be linear in $k$, $\omega_{\boldsymbol{k}} \propto |\boldsymbol{k}|$.

The shortcoming in this discussion is that we neglected the feedback effects of the electrons that react nearly instantaneously to the slow ionic motion, due to their much smaller mass. The finite plasma frequency is a consequence of the long-range nature of the Coulomb potential (as mentioned earlier), but as we have seen above the electrons tend to screen these potentials, in particular for small wavevectors $\boldsymbol{k}$. The "bare" ionic plasma frequency $\Omega_p$ is thus renormalized to

$$\omega_{\boldsymbol{k}}^2 = \frac{\Omega_p^2}{\varepsilon(\boldsymbol{k}, 0)} = \frac{k^2 \Omega_p^2}{k^2 + k_{TF}^2} \approx (c_s k)^2, \tag{3.116}$$

where the presence of the electrons leads to a renormalization of the Coulomb potential by a factor $1/\varepsilon(\boldsymbol{k}, \omega)$. Having included the back-reaction of the electrons, a linear dispersion of a sound wave ($\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}|$) is finally recovered, and the renormalized velocity of sound $c_s$ reads

$$c_s^2 \approx \frac{\Omega_p^2}{k_{TF}^2} = \frac{Z m \omega_p^2}{M_i k_{TF}^2} = \frac{1}{3} Z \frac{m}{M_i} v_F^2. \tag{3.117}$$

[9] See course of Statistical Physics HS09.

For the comparison of the energy scales we find,

$$\frac{\Theta_D}{T_F} \sim \frac{c_s}{v_F} = \sqrt{\frac{1}{3} Z \frac{m}{M_i}} \ll 1, \tag{3.118}$$

where we used $k_B T_F = \epsilon_F$ and $k_D \approx k_F$.

**Kohn anomaly**

Notice that phonon frequencies are much smaller than the (electronic) plasma frequency, so that the approximation

$$\omega_{\boldsymbol{k}}^2 = \frac{\Omega_p^2}{\varepsilon(\boldsymbol{k}, 0)} \tag{3.119}$$

is valid even for larger wavevectors. Employing the Lindhard form of $\varepsilon(\boldsymbol{k}, 0)$, we deduce that the phonon frequency is singular at $|\boldsymbol{k}| = 2k_F$. More explicitly we find

$$\frac{\partial \omega_{\boldsymbol{k}}}{\partial \boldsymbol{k}} \to \infty \tag{3.120}$$

in the limit $k \to 2k_F$. This behavior is called the *Kohn anomaly* and results from the interaction between electrons and phonons. This effect is not contained in the previous elastic medium model that neglected ion-electron interactions.

### 3.3.3 Peierls instability in one dimension

The Kohn anomaly has particularly drastic effects in (quasi) one-dimensional electron systems, where the electron-phonon coupling leads to an instability of the metallic state. We consider a one-dimensional Jellium model where the ionic background is treated as an elastic medium with a displacement field $u$ along the extended direction ($x$-axis). We neglect both the electron-electron interaction and the slow time evolution of the background modulation so that the Hamiltonian reads,

$$\mathcal{H} = \mathcal{H}_{\text{isol}} + \mathcal{H}_{\text{int}}, \tag{3.121}$$

where contributions of the isolated electronic and ionic systems are included in

$$\mathcal{H}_{\text{isol}} = \sum_{k,s} \frac{\hbar^2 k^2}{2m} c_{k\,s}^\dagger c_{k\,s} + \frac{\lambda}{2} \int dx \left( \frac{du}{dx}(x) \right)^2 \tag{3.122}$$

whereas the interactions between the system comes in via the coupling

$$\mathcal{H}_{\text{int}} = -n_0 \sum_s \int dx\, dx'\, V(x - x') \frac{d}{dx} u(x) \widehat{\Psi}_s^\dagger(x') \widehat{\Psi}_s(x') \tag{3.123}$$

In the general theory of elastic media $\boldsymbol{\nabla} \cdot \boldsymbol{u} = -\delta n / n_0$ describes density modulations, so that the second term in (3.121) models the coupling of the electrons to charge density fluctuations of the positively charged background[10] mediated by the screened Coulomb interaction $V(x - x')$. Consider the ground state of $N$ electrons in a system of length $L$, leading to an electronic density

[10] Note that only phonon modes with a finite value of $\boldsymbol{\nabla} \cdot \boldsymbol{u}$ couple in lowest order to the electrons. This is only possible of longitudinal modes. Transverse modes are defined by the condition $\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0$ and do not couple to electrons in lowest order.

$n = N/L$. For a uniform background $u(x) = \text{const}$, the Fermi wavevector of free electrons is readily determined to be

$$N = \sum_s \int_{-k_F}^{+k_F} dk\, 1 = 2\frac{L}{2\pi}2k_F \tag{3.124}$$

leading to

$$k_F = \frac{\pi}{2}n. \tag{3.125}$$

**Emersion of the instability**

Now we consider the Kohn anomaly of this system. For a small background modulation $u(x) \neq \text{const}$, the interaction term $\mathcal{H}_{\text{int}}$ can be treated perturbatively and will lead to a renormalization of the elastic modulus $\lambda$ in (3.121). For that it will be useful to express the full Hamiltonian in momentum space,

$$\mathcal{H}_{\text{isol}} = \sum_{k,s} \frac{\hbar^2 k^2}{2m} c^\dagger_{k\,s} c_{k\,s} + \frac{\Omega\rho_0}{2} \sum_q \omega_q^2 u_q u_{-q} \tag{3.126}$$

$$\mathcal{H}_{\text{int}} = i \sum_{k,q,s} q \big[\tilde{V}_{-q} u_q \hat{c}^\dagger_{k+q,s} \hat{c}_{k,s} - \tilde{V}_q u_{-q} \hat{c}^\dagger_{k,s} \hat{c}_{k+q,s}\big], \tag{3.127}$$

where we used from previous considerations $\lambda q^2 = \rho_0 \omega_q^2$. Furthermore we defined

$$u(x) = \frac{1}{\sqrt{L}} \sum_q u_q e^{-iqx}, \tag{3.128}$$

$$V(x) = \frac{1}{\sqrt{L}} \sum_q \tilde{V}_q e^{iqx}, \tag{3.129}$$

with $\tilde{V}_q = 4\pi e^2/q^2\epsilon(q,0)$. We compute the second order correction to the ground state energy using Rayleigh-Schrödinger perturbation theory (note that the linear energy shift vanishes)

$$\Delta E^{(2)} = \sum_{k,q,s} q^2 |\tilde{V}_q|^2 u_q u_{-q} \sum_n \frac{|\langle\Psi_0|\hat{c}^\dagger_{k,s}\hat{c}_{k+q,s}|n\rangle|^2 + |\langle\Psi_0|\hat{c}^\dagger_{k+q,s}\hat{c}_{k,s}|n\rangle|^2}{E_0 - E_n} \tag{3.130}$$

$$= \sum_q |\tilde{V}_q|^2 q^2 u_q u_{-q} \sum_k \frac{n_{k+q} - n_k}{\epsilon_{k+q} - \epsilon_k} \tag{3.131}$$

$$= \Omega \sum_q |\tilde{V}_q|^2 q^2 \chi_0(q,0) u_q u_{-q} \tag{3.132}$$

where the virtual states $|n\rangle$ are electron-hole excitations of the filled Fermi sea. This term gives a correction to the elastic term in (3.126). In other words, the elastic modulus $\lambda$ and, thus, the phonon frequency ${\omega_q}^2 = \lambda/\rho_0 q^2$ is renormalized according to

$$\left(\omega_q^{\text{ren}}\right)^2 \approx \omega_q^2 + \frac{|\tilde{V}_q|^2 q^2}{\rho_0}\chi_0(q,0) = \omega_q^2 - \frac{|\tilde{V}_q|^2 q}{2\pi\rho_0} \ln\left|\frac{q+2k_F}{q-2k_F}\right| \tag{3.133}$$

From the behavior for $q \to 0$ we infer that the velocity of sound is renormalized. However, a much more drastic modification occurs at $q = 2k_F$. Here the phonon spectrum is 'softened', i.e. the frequency vanishes and even becomes negative. The latter effect is an artifact of the perturbation

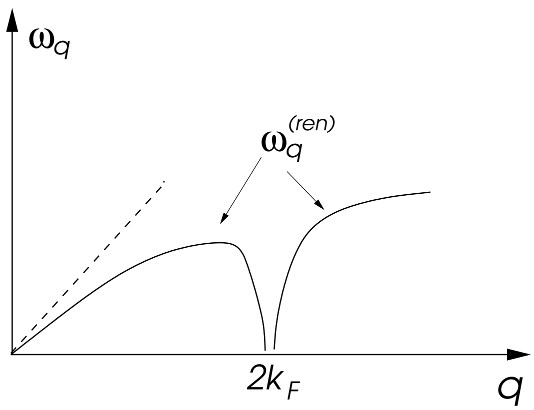



Figure 3.7: Kohn anomaly for the one-dimensional system with electron-phonon coupling. The renormalization of the phonon frequency is divergent at $q = 2k_F$.

theory.[11] This hints at an instability triggered by the Bose-Einstein condensation of phonons with a wave vector of $q = 2k_F$. This coherent superposition[12] of many phonons corresponds classically to a static periodic deformation of the ionic background with wave vector $2k_F$. The unphysical behavior of the frequency $\omega_q$ indicates that in the vicinity of $2k_F$, the current problem can not be treated with the help of perturbation theory around the uniform state.

### Peierls instability at $\mathbf{Q = 2k_F}$

Instead of the perturbative approach, we assume that the background shows a periodic density modulation (coherent phonon state)

$$u(x) = u_0 \cos(Qx) \tag{3.138}$$

where $Q = 2k_F$ and $u_0$ remains to be determined variationally. We investigate the effect of this modulation on the electron-phonon system. To this end we show that such a modulation lowers the energy of the electrons. Assuming that $u_0$ is small we can evaluate the electronic energy using the approximation of nearly free electrons, where $Q$ appears as a reciprocal lattice vector. The electronic spectrum for $0 \le k \le Q$ is then approximately determined by the secular

[11] Note that indeed the expression

$$\omega_q^2 = \frac{\Omega_p^2}{\varepsilon(q,0)} \tag{3.134}$$

in (3.119) does not yield negative energies but gives a zero of $\omega_q$ at $q = 2k_F$.

[12] We introduce the *coherent state*

$$|\Phi_Q^{\mathrm{coh}}\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{(\widehat{b}_Q^\dagger)^n}{n!} \alpha^n |0\rangle \tag{3.135}$$

which does not have a definite phonon number for the mode of wave vector $Q$. On the other hand, this mode is macroscopically occupied, since

$$n_Q = \langle \Phi_Q^{\mathrm{coh}} | \widehat{b}_Q^\dagger \widehat{b}_Q | \Phi_Q^{\mathrm{coh}} \rangle = |\alpha|^2 \tag{3.136}$$

and, moreover, we find

$$\langle \Phi_Q^{\mathrm{coh}} | \widehat{u}(x) | \Phi_Q^{\mathrm{coh}} \rangle = \frac{1}{L} \frac{\hbar}{2\rho_0 \omega_Q} \left[ \alpha e^{iQx} + \alpha^* e^{-iQx} \right] = u_0 \cos(Qx) \tag{3.137}$$

with $u_0 = \hbar\alpha/\rho_0 L \omega_Q$, assuming $\alpha$ being real.

equation

$$\det\begin{pmatrix} \frac{\hbar^2 k^2}{2m} - E & \Delta \\ \Delta^* & \frac{\hbar^2 (k-Q)^2}{2m} - E \end{pmatrix} = 0 \tag{3.139}$$

where $\Delta$ follows from the Fourier transform of the potential $V(x)$,

$$\Delta = -iQu_0 n\tilde{V}_Q \tag{3.140}$$

with

$$\tilde{V}_Q = \int dx\; e^{iQx} V(x). \tag{3.141}$$

The equation (3.139) leads to the energy eigenstates

$$E_k^{\pm} = \frac{\hbar^2}{4m}\left[(k-Q)^2 + k^2 \pm \sqrt{\{(k-Q)^2 - k^2\}^2 + 16m^2|\Delta|^2/\hbar^4}\right]. \tag{3.142}$$

The total energy of the electronic and ionic system is then given by

$$E_{\text{tot}}(u_0) = 2\sum_{0\le k<Q} E_{k-} + \frac{\lambda L Q^2}{4} u_0^2 \tag{3.143}$$

where all electronic states of the lower band ($E_{k-}$) are occupied and all states of the upper band ($E_{k+}$) are empty. The amplitude $u_0$ of the modulation is found by minimizing $E_{\text{tot}}$ with respect to $u_0$:

$$0 = \frac{1}{L}\frac{dE_{\text{tot}}}{du_0} \tag{3.144}$$

$$= -\frac{\hbar^2}{2m}\frac{32Q^2m^2n^2\tilde{V}_Q^2}{\hbar^4}u_0\int_0^Q \frac{dk}{2\pi}\frac{1}{\sqrt{\{(k-Q)^2-k^2\}^2 + 16m^2Q^2n^2\tilde{V}_Q^2u_0^2/\hbar^4}} + \frac{\lambda}{2}Q^2u_0 \tag{3.145}$$

$$= -u_0\frac{4Qmn^2\tilde{V}_Q^2}{\hbar^2\pi}\int_{-k_F}^{+k_F} dq\frac{1}{\sqrt{q^2 + 4m^2n^2\tilde{V}_Q^2u_0^2/\hbar^4}} + \frac{\lambda}{2}Q^2u_0 \tag{3.146}$$

$$= -u_0\frac{8Qmn^2\tilde{V}_Q^2}{\hbar^2\pi}\operatorname{arsinh}\left(\frac{\hbar^2 k_F}{2mn\tilde{V}_Q u_0}\right) + \frac{\lambda}{2}Q^2u_0. \tag{3.147}$$

We solve this equation for $u_0$ using $\operatorname{arsinh}(x) \approx \ln(2x)$ when $x \gg 1$.

$$u_0 = \frac{\hbar^2 k_F}{mn\tilde{V}_Q}\exp\left[-\frac{\hbar^2 k_F \pi\lambda}{8mn^2\tilde{V}_Q^2}\right] = \frac{2}{k_F}\frac{\epsilon_F}{n\tilde{V}_Q}e^{-1/N(0)g} \tag{3.148}$$

where $\epsilon_F = \hbar^2 k_F^2/2m$ is the Fermi energy and $N(0) = 2m/\pi\hbar^2 k_F$ is the density of states at the Fermi energy. We introduced the coupling constant $g = 4n^2\tilde{V}_Q^2/\lambda$ that describes the phonon-induced effective electron-electron interaction. The coupling is the stronger the more polarizable (softer) ionic background, i.e. when the elastic modulus $\lambda$ is small. Note that the static displacement $u_0$ depends exponentially on the coupling and on the density of states. The underlying reason for this so-called *Peierls instability* to happen lies in the opening of an energy gap,

$$\Delta E = E_{k_F}^{+} - E_{k_F}^{-} = 2|\Delta| = 8\epsilon_F \exp\left(-\frac{1}{N(0)g}\right), \tag{3.149}$$

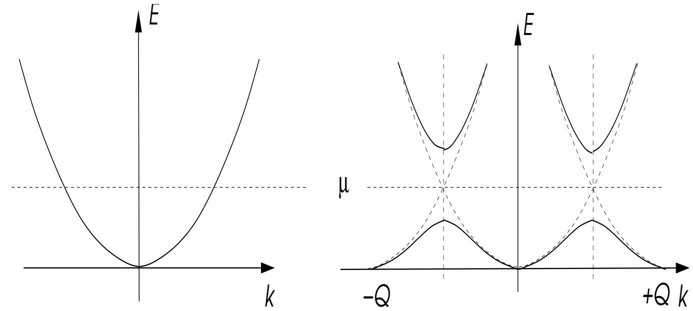



Figure 3.8: Change of the electron spectrum. The modulation of the ionic background yields gaps at the Fermi points and the system becomes an insulator.

at $k = \pm k_F$, i.e. at the Fermi energy. The gap is associated with a lowering of the energy of the electron states in the lower band in the vicinity of the Fermi energy. For this reason this kind of instability is called a Fermi surface instability. Due to the gap the metal has turned into a semiconductor with a finite energy gap for all electron-hole excitations.
The modulation of the electron density follows the charge modulation due to the ionic lattice deformation, which can be seen by expressing the wave function of the electronic states,

$$\psi'_k(x) = \frac{1}{\sqrt{\Omega}} \frac{\Delta e^{ikx} + (E_k - \epsilon_k) e^{i(k-Q)x}}{\sqrt{(E_k - \epsilon_k)^2 + |\Delta|^2}}, \tag{3.150}$$

which is a superposition of two plane waves with wave vectors $k$ and $k - Q$, respectively. Hence the charge density reads

$$\rho_k(x) = -e|\psi'_k(x)|^2 = -\frac{e}{\Omega}\left[1 - \frac{2(\epsilon_k - E_k)|\Delta|}{(E_k - \epsilon_k)^2 + |\Delta|^2} \sin Qx\right] \tag{3.151}$$

and its modulation from the homogeneous distribution $-en$ is given by

$$\delta\rho(x) = \sum_k \rho_k(x) - (-en) = \frac{e}{2}\int_0^{k_F} \frac{dk'}{2\pi} \frac{m|\Delta| \sin Qx}{\sqrt{\hbar^4 k_F^2 k'^2 + m^2|\Delta|^2}} \tag{3.152}$$

$$= \frac{en|\Delta|}{16\epsilon_F} \ln\left|\frac{2\epsilon_F}{|\Delta|}\right| \sin(2k_F x). \tag{3.153}$$

Such a state, with a spatially modulated electronic charge density, is called a *charge density wave* (CDW) state. This instability is important in quasi-one-dimensional metals which are for example realized in organic conductors such as TTF·TCNQ (tetrathiafulvalene tetracyanoquinomethane). In higher dimensions the effect of the Kohn anomaly is generally less pronounced, so that in this case spontaneous deformations rarely occur. As we will see later, a charge density wave instability can nevertheless be observed in multi-dimensional ($d > 1$) systems with a so-called nested Fermi surface. These systems resemble in some respects one-dimensional systems. Finally, notice that the electron-phonon interaction strongly contributes to another kind of Fermi surface instability, when metals exhibit superconductivity.

### 3.3.4 Dynamics of phonons and the dielectric function

We have seen that an external potential $V_a$ is screened by the polarization of the electrons. As the positively charged ionic background is also polarizable, it should be included in the renormalization of the external potential. In general, the fully renormalized potential $V_{\text{ren}}$ may be expressed via

$$\varepsilon V_{\text{ren}} = V_a, \tag{3.154}$$

with the full dielectric function $\varepsilon$. In order to determine $V_{\text{ren}}$ and $\varepsilon$, we define the 'bare' (unrenormalized) electronic (ionic) dielectric function $\varepsilon^{\text{el}}$ ($\varepsilon^{\text{ion}}$). The renormalized potential in (3.154) can be expressed considering three other points of view. First, if the ionic potential $V_{\text{ion}}$ is added to the external potential $V_a$, the remaining screening is due to the electrons only, i.e.,

$$\varepsilon^{\text{el}} V_{\text{ren}} = V_a + V_{\text{ion}}. \tag{3.155}$$

Secondly, the electronic potential $V_{\text{el}}$ may be added to the external potential $V_a$, so that the ions exclusively renormalize the new potential $V_{\text{el}} + V_a$, resulting in

$$\varepsilon^{\text{ion}} V_{\text{ren}} = V_a + V_{\text{el}}. \tag{3.156}$$

Note that in (3.156) all effects of electron polarization are included in $V_{\text{el}}$, so that the dielectric function results from the 'bare' ions. Finally we use the fact that $V_{\text{ren}}$ may be expressed as

$$V_{\text{ren}} = V_a + V_{\text{el}} + V_{\text{ion}}. \tag{3.157}$$

Adding (3.155) to (3.156) and subtracting (3.154), we obtain

$$(\varepsilon^{\text{el}} + \varepsilon^{\text{ion}} - \varepsilon) V_{\text{ren}} = V_a + V_{\text{el}} + V_{\text{ion}} \tag{3.158}$$

which simplifies with (3.157) to

$$\varepsilon = \varepsilon^{\text{el}} + \varepsilon^{\text{ion}} - 1 \tag{3.159}$$

In order to find an alternative expression relating the renormalized potential $V_{\text{ren}}$ to the external potential $V_a$, we make the Ansatz

$$V_{\text{ren}} = \frac{1}{\varepsilon} V_a = \frac{1}{\varepsilon^{\text{ion}}_{\text{eff}}} \frac{1}{\varepsilon^{\text{el}}} V_a \tag{3.160}$$

i.e. the potential $V_a/\varepsilon^{\text{el}}$ that results from bare screening of the polarizable electrons is additionally screened by an effective ionic dielectric function $\varepsilon^{\text{ion}}_{\text{eff}}$ which includes electron-phonon interactions. Using equation (3.159) and the definition of $\varepsilon^{\text{ion}}_{\text{eff}}$ via (3.160) we obtain

$$\varepsilon^{\text{ion}}_{\text{eff}} = 1 + \frac{1}{\varepsilon^{\text{el}}} (\varepsilon^{\text{ion}} - 1), \tag{3.161}$$

or using the definition (3.55)

$$\chi^{\text{ion}}_{0,\text{eff}} = \frac{\chi^{\text{ion}}_0}{\varepsilon^{\text{el}}}. \tag{3.162}$$

Taking into account the discussion of the plasma excitation of the bare ions in Eqs.(3.68, 3.69, and 3.115), and considering the long wave-length excitations ($\boldsymbol{k} \to 0$), we approximate

$$\varepsilon^{\text{ion}} = 1 - \frac{\Omega_p^2}{\omega^2}, \tag{3.163}$$

$$\varepsilon^{\text{el}} = 1 + \frac{k_{TF}^2}{k^2}. \tag{3.164}$$

For the electrons we used the result from the quasi-static limit in (3.78). The full dielectric function now reads

$$\varepsilon = 1 + \frac{k_{TF}^2}{k^2} - \frac{\Omega_p^2}{\omega^2} = \left(1 + \frac{k_{TF}^2}{k^2}\right)\left(1 - \frac{\omega_{\boldsymbol{k}}^2}{\omega^2}\right). \tag{3.165}$$

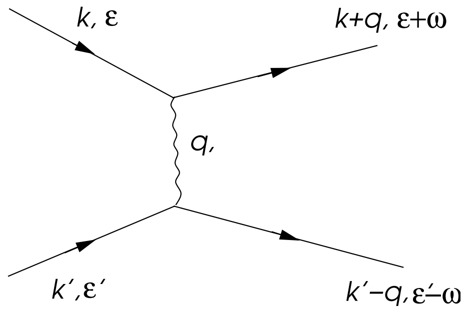



Figure 3.9: Diagram for the electron-electron interaction involving also electron-phonon coupling.

The time-independent Coulomb interaction

$$V_a = \frac{4\pi e^2}{q^2} \tag{3.166}$$

between the electrons is replaced in a metal by an effective interaction

$$V_{\rm ren}(\boldsymbol{q},\omega) = \frac{4\pi e^2}{q^2 \varepsilon(\boldsymbol{q},\omega)} \tag{3.167}$$

$$= \frac{4\pi e^2}{k_{TF}^2 + q^2}\left(\frac{\omega^2}{\omega^2 - \omega_{\boldsymbol{q}}^2}\right). \tag{3.168}$$

This interaction corresponds to the matrix element for a scattering process of two electrons with momentum exchange $\boldsymbol{q}$ and energy exchange $\omega$. The phonon frequency $\omega_{\boldsymbol{q}}$ is always less than the Debye frequency $\omega_D$. Hence the effect of the phonons is almost irrelevant for energy exchanges $\omega$ that are much larger than $\omega_D$. The time scale for such energies would be too short for the slow ions to move and influence the interaction. Interestingly, the repulsive bare Coulomb potential is renormalized to an interaction with an attractive channel for $\omega < \omega_D$ because of overcompensation by the ions. This aspect of the electron-phonon interaction is most important for superconductivity.

# Chapter 4

# Itinerant electrons in a magnetic field

Electrons couple through their orbital motion and their spin to external magnetic fields. In this chapter we focus on the case of orbital coupling which can be also used as a diagnostic tool to observe the presence of a Fermi surface in a metallic system and to map out the Fermi surface topology. A further most intriguing feature of electrons moving in a magnetic field is the Quantum Hall effect of a two-dimensional electronic system. In both case the Landau levels with play an important role and will be introduced here in a first step.

## 4.1 The de Haas-van Alphen effect

The ground state of a metal is characterized by the existence of a discontinuity of the occupation number in momentum space - the Fermi surface. The de Haas-van Alphen experiment is one of the best methods to verify its existence and to determine the shape of a Fermi surface. It is based on the behavior of electrons at low temperatures in a strong magnetic field.

### 4.1.1 Landau levels

Consider a free electron gas subject to a uniform magnetic field $\boldsymbol{B} = (0, 0, B)$. The one-particle Hamiltonian for an electron is given by

$$\mathcal{H} = \frac{1}{2m}\left(-i\hbar\boldsymbol{\nabla} - \frac{e}{c}\boldsymbol{A}\right)^2 - \frac{g\mu_B}{\hbar}\widehat{S}_z B. \tag{4.1}$$

We fix the gauge freedom of the vector potential $\boldsymbol{A}$ by working in the Landau gauge, $\boldsymbol{A} = (0, Bx, 0)$, satisfying $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$. Hence the Hamiltonian (4.1) simplifies to

$$\mathcal{H} = \frac{1}{2m}\left[-\hbar^2\frac{\partial^2}{\partial x^2} + \left(-i\hbar\frac{\partial}{\partial y} - \frac{e}{c}Bx\right)^2 - \hbar^2\frac{\partial^2}{\partial z^2}\right] - \frac{g\mu_B}{\hbar}\widehat{S}_z B. \tag{4.2}$$

In this gauge, the vector potential acts like a confining harmonic potential along the $x$-axis. As translational invariance in the $y$- and $z$-directions is preserved, the eigenfunctions separate in the three spacial components and take the form

$$\psi(\boldsymbol{r}) = e^{ik_z z}e^{ik_y y}\phi(x)\xi_s \tag{4.3}$$

where $\xi_s$ is the spin wave function. The states $\phi(x)$ are found to be the eigenstates of the harmonic oscillator problem, so that we have

$$\phi_{n,k_y}(x) = \frac{1}{\sqrt{2^n\, n!\, 2\pi\, \ell^2}}H_n[(x - k_y\ell^2)/\ell]e^{-(x-k_y\ell^2)^2/2\ell^2} \tag{4.4}$$

where $H_n(x)$ are the Hermite polynomials and $\ell$ represents the *magnetic length* defined via $\ell^2 = \hbar c/|eB|$. The eigenenergies of the Hamiltonian (4.2) read

$$E_{n,k_z,s} = \frac{\hbar^2 k_z^2}{2m} + \hbar\omega_c\left(n + \frac{1}{2}\right) - \frac{g\mu_B}{\hbar}Bs \tag{4.5}$$

where $s = \pm\hbar/2$, $n \in \mathbb{N}_0$ and we have introduced the cyclotron frequency $\omega_c = |eB|/mc$. Note that the energy (4.5) does not depend on $k_y$. The apparent differences in the spatial dependence of the wave functions for the $x$- and $y$-directions are merely a consequence of the chosen gauge.[1] The fact that the energy does not depend on $k_y$ in the chosen gauge indicates a huge degeneracy of the eigenstates. To obtain the number of degenerate states we concentrate for simplicity on $k_z = 0$ and neglect the electron spin. We take the electrons to be confined to a cube of volume $L^3$ with periodic boundary conditions, i.e., $k_y = 2\pi n_y/L$ with $n_y \in \mathbb{N}_0$. As the wave function $\phi_{n,k_y}(x)$ is centered around $k_y\ell^2$, the condition

$$0 < k_y\ell^2 < L \tag{4.8}$$

fixes the maximal number $N_{\rm deg}$ of degenerate states

$$0 < n_y < N_{\rm deg} = \frac{L^2}{2\pi\ell^2}. \tag{4.9}$$

The energies correspond to a discrete set of one-dimensional systems, so that the density of states is determined by the structure of the one-dimensional dispersion (with square root singularities at the band edges) along the $z$-direction:

$$N_0(E,n,s) = \frac{N_{\rm deg}}{\Omega}\sum_{k_z}\delta(E - E_{n,k_z,s}) \tag{4.10}$$

$$= \frac{1}{2\pi\ell^2}\int\frac{dk_z}{2\pi}\delta\left(E - \frac{\hbar^2k_z^2}{2m} - \hbar\omega_c\big(n+\frac{1}{2}\big) + \frac{g\mu_B}{\hbar}Bs\right) \tag{4.11}$$

$$= \frac{(2m)^{3/2}\omega_c}{8\pi^2\hbar^2}\frac{1}{\sqrt{E - \hbar\omega_c(n+1/2) + g\mu_B Bs/\hbar}} \tag{4.12}$$

The total density of states $N_0(E)$ for a given energy $E$ is obtained by summing over $n \in \mathbb{N}_0$ and $s = \pm\hbar/2$. This should be compared to the density of states without the magnetic field,

$$N_0(E, B=0) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\delta\left(E - \frac{\hbar^2\boldsymbol{k}^2}{2m}\right) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3}\sqrt{E}\,. \tag{4.13}$$

The density of states for finite applied field is shown in Fig. 4.1 for one spin-component.

### 4.1.2 Oscillatory behavior of the magnetization

In the presence of a magnetic field, the smooth density of states of the three-dimensional metal is replaced by a discontinuous form dominated by square root singularities. The position of the

[1] Like the vector potential, the wave function is a gauge dependent quantity. To see this, observe that under a gauge transformation

$$\boldsymbol{A}(\boldsymbol{r},t) \to \boldsymbol{A}'(\boldsymbol{r},t) = \boldsymbol{A}(\boldsymbol{r},t) + \boldsymbol{\nabla}\chi(\boldsymbol{r},t) \tag{4.6}$$

the wave function undergoes a position dependent phase shift

$$\psi(\boldsymbol{r},t) \to \psi'(\boldsymbol{r},t) = \psi(\boldsymbol{r},t)e^{i\hbar c\chi(\boldsymbol{r},t)/e}. \tag{4.7}$$

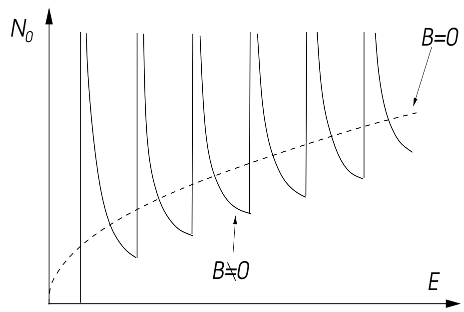



Figure 4.1: Density of states for electrons in a magnetic field due to Landau levels. The dashed line shows the density of states in the absence of a magnetic field.

singularities depends on the strength of the magnetic field. In order to understand the resulting effect on the magnetization, we consider the free energy

$$F = N\mu - TS = N\mu - k_B T \sum_{k_z,k_y,n,s} \ln\left(1 + e^{-(E_{n,k_z,s}-\mu)/k_BT}\right) \tag{4.14}$$

and use the general thermodynamic relation $M = -\partial F/\partial B$. For the details of the somewhat tedious calculation, we refer to J. M. Ziman, *Principles of the Theory of Solids*[2] and merely present the result

$$M = N\chi_P B\left[1 + \frac{\chi_L}{\chi_P} + \frac{\pi k_B T}{\mu_B B}\sqrt{\frac{\epsilon_F}{\mu_B B}}\sum_{\nu=1}^{\infty}\frac{1}{\sqrt{\nu}}\frac{\sin\left(\frac{\pi}{4} - \frac{\pi\nu\epsilon_F}{\mu_B B}\right)}{\sinh\left(\frac{\pi^2\nu k_B T}{\mu_B B}\right)}\right]. \tag{4.15}$$

Here $\chi_P$ is the Pauli-spin susceptibility originating from the Zeeman-term and the second term $\chi_L = -\chi_P/3$ is the diamagnetic Landau susceptibility which is due to induced orbital currents (the Landau levels). For sufficiently low temperatures, $k_B T < \mu_B B$, the magnetization as a function of the applied field exhibits oscillatory behavior. The dominant contribution comes from the summand with $\nu = 1$. The oscillations are a consequence of the singularities in the density of states that influence the magnetic moment upon successively passing through the Fermi energy as the magnetic field is varied. The period in $1/B$ of the oscillations of the term $\nu = 1$ is easily found to be

$$\frac{\pi\epsilon_F}{\mu_B}\Delta\left(\frac{1}{B}\right) = 2\pi \tag{4.16}$$

or

$$\Delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar c}\frac{1}{A(k_F)} \tag{4.17}$$

where we used that $\mu_B = \hbar e/2mc$ and defined the cross sectional area $A(k_F) = \pi k_F^2$ of the Fermi sphere perpendicular to the magnetic field.

### 4.1.3 Onsager equation

The behavior we have found above for free electrons, generalizes to systems with arbitrary band structures. In these cases there are usually no exact solutions available. Instead of generalizing the above treatment to such band systems, we discuss the behavior of electrons within the quasiclassical approximation (see Section 1.7) and consider the closed orbits of a wave packet

[2] German title : *Prinzipien der Festkörpertheorie*

subject to a magnetic field. The quasi-classical equations of motion for the center of mass of the wave packet (1.96, 1.97) simplify in the absence of an electric field to

$$\boldsymbol{v}_{\boldsymbol{k}} = \frac{\partial \epsilon_{\boldsymbol{k}}}{\partial \hbar \boldsymbol{k}} \tag{4.18}$$

$$\hbar \dot{\boldsymbol{k}} = -\frac{e}{c} \boldsymbol{v}_{\boldsymbol{k}} \times \boldsymbol{B}. \tag{4.19}$$

The wave vector $\boldsymbol{k}$ is restricted to the plane perpendicular to the applied field $\boldsymbol{B}$. The time needed for travelling along a path between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ is given by

$$t_2 - t_1 = \int_{\boldsymbol{k}_1}^{\boldsymbol{k}_2} dk \frac{1}{|\dot{\boldsymbol{k}}|} = \frac{\hbar c}{eB} \int_{\boldsymbol{k}_1}^{\boldsymbol{k}_2} \frac{dk}{|\boldsymbol{v}_{\boldsymbol{k},\perp}|} \tag{4.20}$$

where $\boldsymbol{v}_{\boldsymbol{k},\perp}$ denotes the component of the velocity that is perpendicular to $\boldsymbol{B}$. For fixed positive energies $\epsilon$ and $\Delta\epsilon$, let $\Delta\boldsymbol{k}$ be the vector in the plane of the motion that is perpendicular to both $\dot{\boldsymbol{k}}$ and $\boldsymbol{B}$, and that points from an orbit of energy $\epsilon$ to one with energy $\epsilon + \Delta\epsilon$. Then, we have

$$\Delta\epsilon = \frac{\partial \epsilon_{\boldsymbol{k}}}{\partial \boldsymbol{k}} \cdot \Delta\boldsymbol{k} = \frac{\partial \epsilon_{\boldsymbol{k}}}{\partial \boldsymbol{k}}_{\perp} \cdot \Delta\boldsymbol{k} = \left| \frac{\partial \epsilon_{\boldsymbol{k}}}{\partial \boldsymbol{k}}_{\perp} \right| |\Delta\boldsymbol{k}| = \hbar |\boldsymbol{v}_{\boldsymbol{k},\perp}| |\Delta\boldsymbol{k}| \,, \tag{4.21}$$

because $\partial\epsilon_{\boldsymbol{k}}/\partial\boldsymbol{k}_{\perp}$ and $\Delta\boldsymbol{k}$ are perpendicular to orbits of constant energy. Therefore, we infer that

$$\frac{1}{|\boldsymbol{v}_{\boldsymbol{k},\perp}|} = \frac{\hbar |\Delta\boldsymbol{k}|}{\Delta\epsilon}, \tag{4.22}$$

hence

$$t_2 - t_1 = \frac{\hbar^2 c}{eB} \frac{1}{\Delta\epsilon} \int_{\boldsymbol{k}_1}^{\boldsymbol{k}_2} |\Delta\boldsymbol{k}| dk = \frac{\hbar^2 c}{eB} \frac{\Delta A_{1,2}}{\Delta\epsilon} \,. \tag{4.23}$$

Here $\Delta A_{1,2}$ is the ($k$-space) area swept by $\Delta\boldsymbol{k}$ when going from $\boldsymbol{k}_1$ to $\boldsymbol{k}_2$ (see Fig. 4.2). Passing to infinitesimal calculus ($\Delta\boldsymbol{k} \to \delta\boldsymbol{k}$), we can write

$$t_2 - t_1 = \frac{\hbar^2 c}{eB} \frac{\partial A_{1,2}}{\partial \epsilon} \,. \tag{4.24}$$

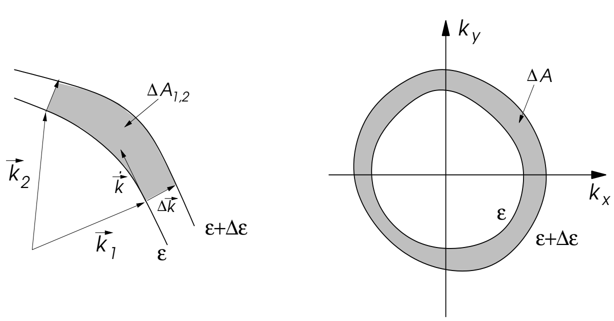



Figure 4.2: Motion of electrons in $\boldsymbol{k}$-space. The shaded area shows the area covered by the displacement vector $\Delta\boldsymbol{k}$ during the motion.

When $\boldsymbol{k}_2 \to \boldsymbol{k}_1$ (sweeping the whole circle) with $A_{1,2}(\epsilon) \to A(\epsilon) \neq 0$ and $A(\epsilon)$ is the $k$-space area enclosed by the curve $\epsilon_{\boldsymbol{k}} = \epsilon$ the wave packet has returned to the initial point in $k$-space. The time $T(\epsilon)$ it takes for the wave packet for such a period is

$$T(\epsilon) = \frac{\hbar^2 c}{eB}\frac{\partial A(\epsilon)}{\partial \epsilon}. \tag{4.25}$$

Using the discrete Landau levels with energies $E_{n,k_z}$, we conclude from Bohr's correspondence principle the relation

$$\Delta E = E_{n+1,k_z} - E_{n,k_z} = \frac{h}{T(E_{n,k_z}, k_z)}, \tag{4.26}$$

when the number of the Landau levels involved is large. This result states that the difference between the energies of adjacent energy levels is given by the inverse period of classical closed orbits. As we are interested in the energy levels close to the Fermi energy, $E_{n,k_z} \sim \epsilon_F$, we find

$$n \sim \frac{\epsilon_F}{\hbar\omega_c} \gg 1 \tag{4.27}$$

Note that $\hbar\omega_c = 1.25K \times B$ [Tesla] and $\epsilon_F \sim 10^4 K$ typically. Invoking the results from (4.25) and (4.26) we can show that

$$\Delta A = A(E_{n+1,k_z}) - A(E_{n,k_z}) = \frac{2\pi eB}{\hbar c} \tag{4.28}$$

where we have used that

$$\frac{\partial A(\epsilon)}{\partial \epsilon} = \frac{\Delta A}{\Delta E} = \frac{eB}{\hbar^2 c} T(E_{n,k_z}, k_z) \tag{4.29}$$

is a good approximation to the equation (4.25).

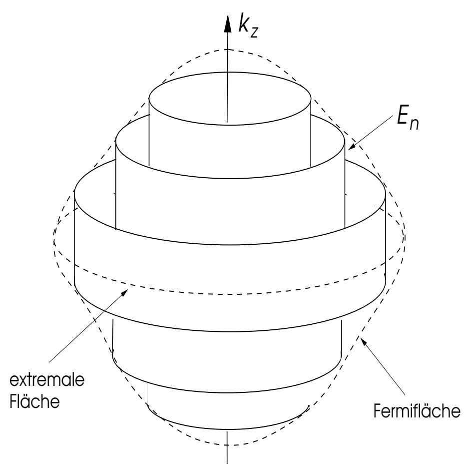



Figure 4.3: Tubes of quantized electronic states in a magnetic field along the $z$-axis. A maximum of the magnetization occurs every time a tube crosses the extremal Fermi surface area as the magnetic field is increased.

The area bounded by two neighboring classical orbits with quantum mechanically allowed energies is $\Delta A$, irrespective of the quantum number $n$. It follows that the area enclosed by one classical orbit with given quantum numbers $n$ and $k_z$ is

$$A(E_{n,k_z}, k_z) = (n + \gamma)\Delta A \tag{4.30}$$

where $\gamma$ is an integration constant. This equation is called the *Onsager equation*. The area corresponding to an extremal density of states at the Fermi surface belongs to the quantum number $n = n_F$ and the orbit with $E_{n_F,k_z} = \epsilon_F$

$$A(\epsilon_F, k_z = 0) = \Delta A(n_F + \gamma) = \frac{2\pi eB}{\hbar c}(n_F + \gamma). \tag{4.31}$$

Notice, that $n_F = n_F(B)$, and the period of the oscillatory behavior is induced by a jump of $n_F(B)$ by an integer with a period

$$\Delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar c}\frac{1}{A(\epsilon_F)} \tag{4.32}$$

The oscillations in the magnetization thus allow to measure the cross sectional area of the Fermi 'sphere'. By varying the orientation of the field the topology of the Fermi surface can be mapped. As an alternative to the measurement of magnetization oscillations one can also measure resistivity oscillations known under the name Schubnikov-de Haas effect. For both methods it is crucial that the Landau levels are sufficiently clearly recognizable. Apart from low temperatures this necessitates sufficiently clean samples. In this context, sufficiently clean means that the average life-time $\tau$ (average time between two scattering events) has to be much larger than the period of the cyclotron orbits, i.e. $\omega_c\tau \gg 1$. This condition follows from the uncertainty relation

$$\Delta\epsilon \sim \frac{\hbar}{\tau} \ll \hbar\omega_c. \tag{4.33}$$

## 4.2 Quantum Hall Effect

### Classical Hall Effect

The Hall effect, discovered by Edwin Hall in 1879, originates from the Lorentz force exerted by a magnetic field on a moving charge. This force is perpendicular both to the velocity of the charged particle and to the magnetic field. In the presence of an electrical current, a Lorentz force produces a transverse voltage, whenever the applied magnetic field points in a direction non-collinear to the current in the conductor. This so-called *Hall effect* can be used to investigate some properties of the charge carriers. Before treating the quantum version, we briefly review the original Hall effect. To this end we consider the classical equation of motion of an electron, subject to an electric and a magnetic field

$$m^* \frac{d\boldsymbol{v}}{dt} = -e\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c} \times \boldsymbol{B}\right), \tag{4.34}$$

where $m^*$ is the effective electron mass. For this classical system, the steady state equation reads

$$\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c} \times \boldsymbol{B}\right) = 0. \tag{4.35}$$

For the Hall geometry shown in Fig. 4.4 with fixed current $\boldsymbol{j} = (0, j_y, 0) = (0, -n_0 ev, 0)$ and magnetic field $\boldsymbol{B} = B_z$, the steady state condition (4.35) simplifies to

$$E_x + \frac{vB_z}{c} = 0. \tag{4.36}$$

The solution $E_x = -vB_z/c$ yields the Hall voltage that compensates the Lorentz force. The Hall conductivity $\sigma_H$ is defined as the ratio between the longitudinal current $j_y$ and the transverse electric field $E_x$, leading to

$$\sigma_H = \frac{j_y}{E_x} = \frac{n_0 ec}{B_z} = \nu\frac{e^2}{h}, \tag{4.37}$$

where $\nu = n_0 hc/Be$. We infer from Eq.(4.37), that the measurements of the Hall conductivity can be used to determine both the charge density $n_0$ and the sign of the charge carriers, i.e. whether the Fermi surfaces encloses the Γ-point for electron-like, negative charges or a point on the boundary of the Brillouin zone for hole-like, positive charges.

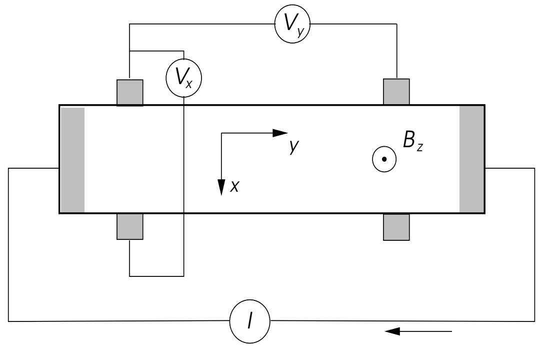



Figure 4.4: Schematic view of a Hall bar. The current runs a long the $y$-direction and the magnetic field is applied along $z$-direction. The voltage $V_y$ determines the conductance along the Hall bar, while $V_x$ corresponds to the transverse Hall voltage.

### Discovery of the Quantum Hall Effects

When measuring the Hall effect in a special two-dimensional electron system, Klaus von Klitzing and his collaborators made 1980 an astonishing discovery.[3] The system was the inversion layer of GaAs-MOSFET device with a sufficiently high gate voltage (see Section 2.4.3) which behaves like a two-dimensional electron gas with a high mobility $e\tau/m^*$ due to the mean free path $l \sim 10$Å and low density ($n_0 \sim 10^{11}\text{cm}^{-2}$. The two extended dimensions correspond to the interface of the MOSFET, whereas the electrons are confined in the third dimension like in a potential well (cf. Section 2.4.3). In high magnetic fields between $1 - 30$T and at sufficiently low temperatures ($T < 4$K), von Klitzing and coworkers observed a quantization of the Hall conductivity corresponding to exact integer multiples of $e^2/h$

$$\sigma_H = \nu_n \frac{e^2}{h} \tag{4.38}$$

where $\nu_n \in \mathbb{N}$. By now, the integer quantization is so widely verified, that the von Klitzing constant (resistance quantum named after the discoverer of the Quantum Hall effect) $R_K = h/e^2 = 25812.807557\Omega$ is used in resistance calibrations. In the field range where the transverse conductivity shows integer plateaus in $\nu \propto 1/B$, the longitudinal conductivity $\sigma_{yy}$ vanishes and takes finite values only when $\sigma_H$ crossed over from one quantized value to the next (see Fig. 4.5).

In 1982, Tsui, Störmer, and Gossard[4] discovered an additional quantization of $\sigma_H$, corresponding to certain rational multiples of $e^2/h$. Correspondingly, one now distinguishes between the integer quantum Hall effect (IQHE) and the fractional quantum Hall effect (FQHE). These discoveries marked the beginning of a whole new field in solid state physics that continues to produce interesting results.

[3] See [von Klitzing, Dorda, and Pepper, Phys. Rev. Lett. **45**, 494 (1980)] for the original paper.

[4] See [Tsui, Störmer and Gossard Phys. Rev. Lett. **48**, 1559 (1982)] for the original paper

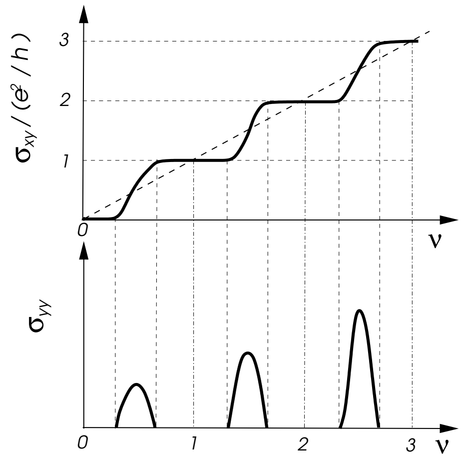



Figure 4.5: Integer Quantum Hall effect: As a function of the filling factor $\nu$ plateaus in $\sigma_{xy}$ appear at multiples of $e^2/h$. The longitudinal conductance $\sigma_{yy}$ is only finite for fillings where $\sigma_{xy}$ changes between plateaus.

### 4.2.1 Hall effect of the two-dimensional electron gas

Here we first discuss the Hall effect in the quantum mechanical treatment. For this purpose we start with the Hamilton operator (4.1) and neglect the electron spin. Working again in the Landau gauge, $\boldsymbol{A} = (0, Bx, 0)$, and confining the electronic system to two dimensions, the Hamiltonian reduces to

$$\mathcal{H} = \frac{1}{2m}\left[-\hbar^2\frac{\partial^2}{\partial x^2} + \left(-i\hbar\frac{\partial}{\partial y} - \frac{e}{c}Bx\right)^2\right]. \tag{4.39}$$

For the two-dimensional gas there is no motion in the $z$-direction, so that the highly degenerate energy eigenvalues are given by the spectrum of a one-dimensional harmonic oscillator $E_n = \hbar\omega_c(n+1/2)$, where again $\omega_c = |eB|/m^*c$. Here, we will concentrate on the lowest Landau level ($n = 0$) with the wave function

$$\phi_{0,k_y} = \frac{1}{\sqrt{2\pi\ell^2}}e^{-(x-x_0)^2/2\ell^2}e^{ik_y y}. \tag{4.40}$$

where the magnetic length $\ell = \sqrt{\hbar c/|eB|}$ gives the extension of the wave function in the presence of the magnetic field. In $x$-direction, the wave function is localized around $x_0 = k_y\ell^2$, whereas it takes the form of a plane wave in $y$-direction. As discussed previously, the energy does not depend on $k_y$.

We now introduce an electric field $E_x$ along the $x$-direction. The Hamilton operator (4.39) is then modified by an additional potential $U(\boldsymbol{r}) = -eE_x x$. This term can easily be absorbed into the harmonic potential and leads to a shift of the center of the wave function,

$$x_0(k_y) \to x_0'(k_y) = k_y\ell^2 - \frac{eE_x}{m^*\omega_c^2}. \tag{4.41}$$

Moreover the degeneracy of the Landau level is lifted since the energy becomes $k_y$-dependent and (after completing the square) takes the form

$$E_{n=0}(k_y) = \frac{\hbar\omega_c}{2} - eE_x x_0'(k_y) + \frac{m^*}{2}\left(\frac{cE_x}{B}\right)^2. \tag{4.42}$$

The energy (4.42) corresponds to the wave function $\phi_{0k_y}$ from (4.40) where $x_0$ is replaced by $x'_0$. The velocity of the electrons is then given by

$$v_y(k_y) = \frac{1}{\hbar}\frac{dE_{n=0}(k_y)}{dk_y} = -\frac{eE_x\ell^2}{\hbar} = -\frac{cE_x}{B\hbar}, \tag{4.43}$$

and from this, we determine the current density,

$$j_y = -en_0 v_y(k_y) = en_0\frac{cE_x}{B\hbar} = \frac{e\nu}{2\pi\ell^2}\frac{cE_x}{B\hbar} = \nu\frac{e^2}{h}E_x = \sigma_H E_x \tag{4.44}$$

where $\nu = n_0 2\pi\ell^2$ is the filling of the Landau level.[5] The Hall conductivity is then identical to the result (4.37) derived previously based on the quasiclassical approximation. There is a linear relation between the Hall conductivity $\sigma_H$ and the index $\nu \propto B^{-1}$.

### 4.2.2 Integer Quantum Hall Effect

The plateaus observed by von Klitzing in the Hall conductivity $\sigma_H$ of the two-dimensional electron gas as a function of the magnetic field correspond to the values $\sigma_H = \nu_n e^2/h$, as if $\nu = \nu_n \in \mathbb{N}$ was restricted to be an integer. Meanwhile, the longitudinal conductivity of the electron gas vanishes when a plateau of $\sigma_H$ is realized

$$\sigma_{yy} = \frac{j_y}{E_y} = 0, \tag{4.45}$$

and only becomes finite at the transition points of $\sigma_H$ between two plateaus (cf. Fig. 4.5). This fact seems to be in contradiction with the results from the consideration above. The solution to this mysterious behavior lies in the fact that disorder, which is always present in a real inversion layer, plays a crucial role and should not be neglected. In fact, due to the disorder, the electrons move in a randomly modulated potential landscape $U(x,y)$. As we will find out, even small amounts of disorder lead to the localization of electronic states in this two-dimensional system. To illustrate this new aspect we focus on the lowest Landau level in the symmetric gauge $\boldsymbol{A} = (-y, x, 0)B/2$. The Schrödinger equation in polar coordinates is given by

$$\frac{\hbar^2}{2m^*}\left[-\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} - \left(\frac{1}{r}\frac{\partial}{\partial\varphi} - i\frac{e}{2\hbar c}Br\right)^2\right]\psi(r,\varphi) + U(x,y)\psi(r,\varphi) = E\psi(r,\varphi)\ . \tag{4.46}$$

Without the external potential $U(x,y)$ we find the ground state solutions

$$\psi_{n=0,m}(r,\varphi) = \frac{1}{\sqrt{2\pi\ell^2 2^m m!}}\left(\frac{r}{\ell}\right)^m e^{-im\varphi}e^{-r^2/4\ell^2} \tag{4.47}$$

where all values of $m \in \mathbb{N}_0$ correspond to the same energy $E_{n=0} = \hbar\omega_c/2$. One easily verifies, that the wave functions $|\psi_{n=0,m}(r,\varphi)|$ are peaked on circles of radius $r_m = \sqrt{2m}\,\ell$ (see Fig.4.6). Note that the magnetic flux threading such a circle is given by

$$\pi B r_m^2 = \pi B 2m\ell^2 = 2\pi m B\frac{\hbar c}{eB} = m\frac{hc}{e} = m\Phi_0, \tag{4.48}$$

which is an integer multiple of the flux quantum $\Phi_0 = hc/e$.
Now we consider the effect of the disorder potential. The gauge can be adjusted to the potential landscape. If, for simplicity, we assume the potential to be rotationally invariant around the origin, the symmetric gauge is already optimal. For the potential

$$U(x,y) = U(r) = \frac{C_1}{r^2} + C_2 r^2 + C_3, \tag{4.49}$$

[5] Note that $\nu^{-1} = B/n_0\Phi_0$ where $\Phi_0 = hc/e$ represents the flux quantum, i.e. $\nu^{-1} \propto B$ is the number of flux quanta $\Phi_0$ per electron.

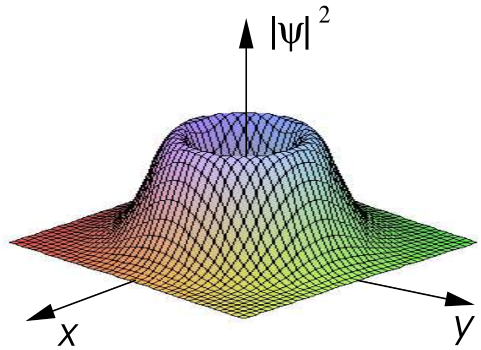



Figure 4.6: Wavefunction of a Landau level state in the symmetric gauge.

the exact expression of all eigenstates of equation (4.46) in the lowest Landau level is obtained using the Ansatz

$$\tilde{\psi}_{0,m}(r,\varphi) = \frac{1}{\sqrt{2\pi\ell^{*2}2^{\alpha}\Gamma(\alpha+1)}}\left(\frac{r}{\ell^*}\right)^{\alpha} e^{-im\varphi}e^{-r^2/4\ell^{*2}}. \tag{4.50}$$

After introducing the dimensionless parameters $C_1^* = 2m^*C_1/\hbar^2$ and $C_2^* = 8\ell^4 m^* C_2/\hbar^2$, the quantities $\alpha$ and $\ell^*$ from equation (4.50) can be expressed via

$$\alpha^2 = m^2 + C_1^*\ , \tag{4.51}$$

$$\left(\ell^*\right)^{-2} = \left(\ell\right)^{-2}\sqrt{1+C_2^*}\ . \tag{4.52}$$

Indeed, the Ansatz (4.50) describes eigenstates of the disordered problem (4.46). The degeneracy of the ground state energy (the lowest Landau level) is now lifted,

$$E_{0,m} = \frac{\hbar\omega_c}{2}\left[\frac{\ell^2}{\ell^{*2}}(\alpha+1) - m\right] + C_3. \tag{4.53}$$

The wave functions are concentrated around the radii $r_m = \sqrt{2\alpha}\ell^*$. For weak potentials $C_1^*, C_2^* \ll 1$ and $m \gg 1$ the energy is approximatively given by

$$E_{0,m} \approx \frac{\hbar\omega_c}{2} + \frac{C_1}{r_m^2} + C_2 r_m^2 + C_3 + \ldots, \tag{4.54}$$

i.e. the wave function adjusts itself to the potential landscape. It turns out that the same is true for arbitrarily structured weak potential landscapes. The wave function describes electrons on quasi-classical trajectories that trace the equipotential lines of the underlying disorder potential. Consequently the states described here are localized in the sense that they are attached to the structure of the potential. The application of an electric field cannot set the electrons in the concentric rings in motion. Therefore, the electrons are localized and do not contribute to electric transport.

**Picture of the potential landscape**

When the magnetic field is varied the filling $\nu = n_0 2\pi\ell^2$ of the Landau level is adjusted accordingly. While all states of a given level are degenerate in the transitionally invariant case, now, these states are spread over a certain energy range due to the disorder. In the quasi-classical approximation, these states correspond to equipotential trajectories that are either filled or empty depending on the strength of the magnetic field, i.e. they are either below or above the chemical potential. These considerations lead to an intuitive picture on localized (closed trajectories) and extended (percolating trajectories) states. We may consider the potential landscape like a real landscape where the the trajectories are contour lines. Assume that we fill now water into such a landscape. The trajectories of the particles is restricted to the shore line. For small filling, we find lakes whose shores are closed and correspond to contour lines. They correspond to closed electron trajectories and represent localized electronic states. At very high water level, only the large mountains of the potential landscape would reach out of the water, forming islands in the sea. The coastlines again represent closed trajectories corresponding to localized electronic states. At some intermediate filling, a boundary between the lake and the island topology, there is a water level at which the coast lines become arbitrarily long and percolate through the whole landscape. Only these contour lines correspond to extended (non-localized) electron states. From this picture we conclude that when a Landau level of a system subject to a random potential is gradually filled, first all occupied state are localized (low filling). At some special intermediate filling level, the extended states are filled and contribute to the current transport. At higher chemical potential (filling) the states would be localized again. In the following argument, going back to Robert B. Laughlin, the presence of filled extended states plays an important role.

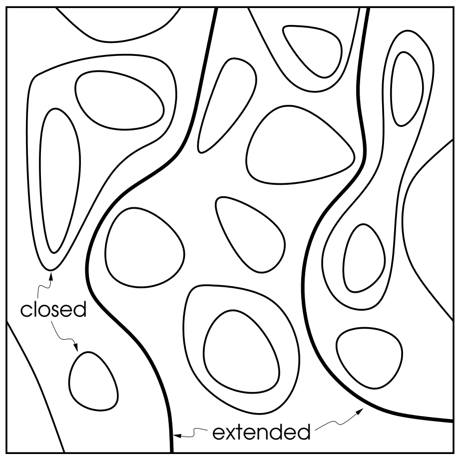



Figure 4.7: Contour plot of potential landscape. There are closed trajectories and extended percolating trajectories.

**Laughlin's gauge argument**

We consider a long rectangular Hall element that is deformed into a so-called *Corbino disc*, i.e. a circular disc with a hole in the middle as shown in Fig.4.8. The Hall element is threaded by a constant and uniform magnetic field $B$ along the $z$-axis. In addition we can introduce an arbitrary flux $\Phi$ through the hole without influencing the uniform field in the disc. This flux is irrelevant for all localized electron trajectories since only extended (percolating) trajectories wind around the hole of the disc and by doing so receive an Aharonov-Bohm phase. When the flux is increased adiabatically by $\delta\Phi$, the vector potential is changed according to

$$\boldsymbol{A} \to \boldsymbol{A} + \delta\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{\nabla}\chi,$$

which in our case means

$$(\delta A)_\varphi = \frac{\delta\Phi}{2\pi r}. \tag{4.55}$$

At the same time, the wave function acquires a phase factor

$$\psi \to \psi\, e^{ie\chi/\hbar c} = \psi\, e^{i(\delta\Phi/\Phi_0)\,\varphi}. \tag{4.56}$$

If the disc was translationally invariant, meaning that disorder is neglected and only extended states exist, we could use the wave functions $\psi_{0,m}$ from (4.47), so that $B\pi r_m^2 = m\Phi_0 + \delta\Phi$. The single-valuedness of the wave function implies that $m$ has to be adjusted, $m \to m - \delta\Phi/\Phi_0$. This guarantees, that increasing $\Phi$ by one flux quantum leads to a decrease of $m$ by 1. Hence, gauge invariance implies that the wave functions are shifted in their radius. This argument is also applicable to higher Landau levels.

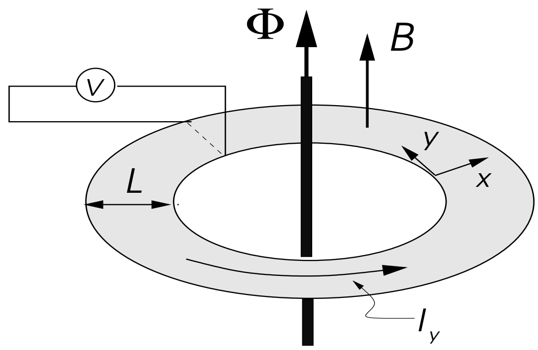



Figure 4.8: Corbino disk for Laughlin's argument. According to the Hall bar in Fig. 4.4, the radial (transverse) component of the Corbino disc is denoted by $x$, while the angular (longitudinal) component is termed as $y$. Both the homogeneous magnetic field $B$ and the flux $\Phi$ point along the $z$-axis perpendicular to the plane of the disc.

Since this argument is topological in nature, it will not break down for independent electrons when disorder is introduced. The transfer of one electron between neighboring extended states due to the change of $\Phi$ by $\Phi_0$ leads to a net shift of one electron from the outer to the inner boundary. If an electric field $E_x$ is applied in the radial direction (here denoted by $x$-direction, see Fig. 4.8), the transfer of this electron results in the energy change

$$\Delta\epsilon_V = -eE_x L \tag{4.57}$$

where $L$ is the distance between the inner and the outer boundary of the Corbino disc. A further change in the electromagnetic energy

$$\Delta\epsilon_I = \frac{I_y\,\delta\Phi}{c}, \tag{4.58}$$

is caused by the constant current $I_y$ (here the angular component is denoted by $y$, see Fig.4.8) in the disc when the magnetic flux is increased by $\delta\Phi$. Following the Aharanov-Bohm argument that the energy of the system is invariant under a flux change by integer multiples of $\Phi_0$, the two energies should compensate each other. Thus, setting $\delta\Phi = \Phi_0$ and demanding that $\Delta\epsilon_V + \Delta\epsilon_I = 0$ leads to

$$\sigma_H = \frac{j_y}{E_x} = \frac{I_y}{LE_x} = \frac{e^2}{h}. \tag{4.59}$$

We conclude from this argument, that each filled Landau level containing percolating states will contribute $e^2/h$ to the total Hall conductivity. Hence, for $\nu_n \in \mathbb{N}_0$ filled levels the Hall conductivity is given by $\sigma_H = \nu_n e^2/h$. Note the importance of the topological nature of the Hall conductivity ensuring the universal character of the quantization.

### Localized and extended states

The density of states of the two-dimensional electron gas (2DEG) in absence of an external magnetic field is given by

$$N_{\text{2DEG}}(E) = 2 \sum_{k_x,k_y} \delta\left(E - \frac{\hbar^2(k_x^2+k_y^2)}{2m}\right) = \frac{L_x L_y m}{2\pi}, \tag{4.60}$$

with twice degenerate energy states for the spins, whereas for the Landau levels in a clean sample, we have

$$N_L(E) = \frac{L_x L_y}{2\pi\ell^2} \sum_{n,s} \delta(E - E_{n,s}). \tag{4.61}$$

Here the prefactor is given by the large degeneracy (4.9) of each Landau level.

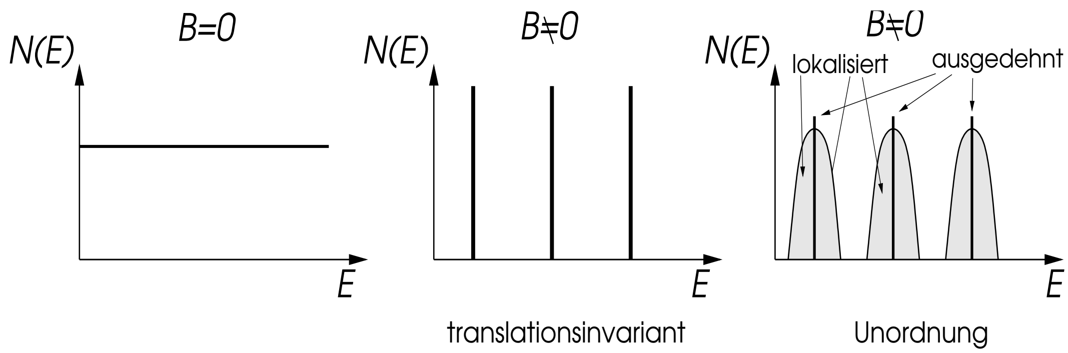



Figure 4.9: Density of states for the two-dimensional electron gas in three different cases. On the left panel, the system without applied magnetic field. On the middle panel, an external magnetic field is applied to a clean system showing sharp strongly degenerate Landau levels. The right panel visualizes the effect of disorder in the two-dimensional system with magnetic field; the Landau levels are spread and the density of state shows broadened peaks where most of the states are localized and only few states in the center percolate.

According to our previous discussion, the main effect of a potential is to lift the degeneracy of the states comprising a Landau level. This remains true for random potential landscapes. Most of the states are then localized and do not contribute to electric transport. Only the few extended states contribute to the transport if they are filled (see Fig. 4.9). For partially filled extended states the Hall conductivity $\sigma_H$ is not an integer multiple of $e^2/h$, since not all percolating states – necessary for transferring one electron from one edge to the other, when the flux is changed by $\Phi_0$ (in Laughlin's argument) – are occupied. Thus, the charge transferred does not amount to a complete $-e$. The appearance of partially filled extended states marks the transition from one plateau to the next and is accompanied by a finite longitudinal conductivity $\sigma_{yy}$. When all extended states of a Landau level are occupied, the contribution to the longitudinal transport stops, i.e., in the range of a plateau $\sigma_{yy}$ vanishes. Because of thermal occupation, the plateaus quickly shrink when the temperature of the system is increased. This is the reason why the Quantum Hall Effect is only observable for sufficiently low temperatures ($T < 4$K).

### Edge states and Büttiker's argument

A confining potential prevents the electrons from leaving the metal. This potential at the edges of the sample has also to be considered in the potential landscape. Interestingly, equi-potential trajectories of states close to the edge are always extended and percolate along the edge. The corresponding wave functions have already been discussed in Section 4.2.1. From Eq.(4.42), we find that the energy is not symmetric in $k_y$, the wave vector along the edge, i.e. $E(k_y) \neq E(-k_y)$. This implies that the states are chiral and can move in one direction only for a given energy. The edge states on the opposing edges move in opposite directions, a fact that can be readily verified by inspection of (4.42) based on Fig.4.10. The total current flowing along the edge for a given Landau level is

$$I = \sum_{k_y} \frac{e}{L_y} v_y, \tag{4.62}$$

meaning that one state per $k_y$ extends over the whole length $L_y$ of the Hall element. Thus, the density is given by $1/L_y$. The wave vector is quantized according to the periodic boundary conditions; $k_y = 2\pi n_y / L_y$ with $n_y \in \mathbb{Z}$. The velocity $v_y$ is given by equation (4.43). In summary, we have

$$I = \frac{e}{2\pi\hbar} \int_{\text{occupied}} dk_y \frac{dE_n(k_y)}{dk_y} = \frac{e}{h} \int_{\text{occupied}} dx_0 \frac{dE}{dx_0} = \frac{e}{h}(\mu - E_n^{(0)}) \tag{4.63}$$

where $x_0 = k_y \ell^2$ is the transversal position of the wave function and $\mu$ is the chemical potential. Sufficiently far away from the boundary $E_n$ is independent of $x_0$ and approaches the value $E_n^{(0)} = \hbar\omega_c(1/2 + n)$ of a translationally invariant electron gas. The potential difference between the two opposing edges leads to a net current along the edge direction of the Hall bar,

$$\mu_A - \mu_B = eV_H = eE_x L_x = \frac{h}{e}(I_A + I_B) = \frac{h}{e} I_H, \tag{4.64}$$

and with that

$$\sigma_H = \frac{I_H}{E_x L_x} = \frac{e^2}{h}, \tag{4.65}$$

where for $\mu_A = \mu_B$ we have $I_A = -I_B$. Note that $I_H = I_A + I_B$ is only valid when no currents are present in the bulk of the system. The latter condition is ensured by the localization of the states at the chemical potential. This argument by Büttiker leads to the same quantization as derived before, namely every Landau level contributes one edge state and thus $\sigma_H = \nu_n e^2/h$ ($\nu_n$ is the number of occupied Landau levels). Note that this argument is independent of the precise shape of the confining edge potential.
Within this argument we are able to understand why the longitudinal conductivity has to vanish at the Hall plateaus. For that we express Ohm's law However it is simpler to discuss the resistivity. Like the conductivity $\widehat{\sigma}$ the resistivity $\widehat{\rho}$ is a tensor:

$$\boldsymbol{j} = \widehat{\sigma}\boldsymbol{E} \tag{4.66}$$

$$\boldsymbol{E} = \widehat{\rho}\boldsymbol{j} \tag{4.67}$$

where the conductivity $\widehat{\sigma}$ and the resistivity $\widehat{\rho}$ are both a symmetric $2 \times 2$ tensor with $\sigma_{xx} = \sigma_{yy}$ and $\rho_{xx} = \rho_{yy}$. Therefore we find

$$\sigma_{yy} = \frac{\rho_{yy}}{\rho_{yy}^2 + \rho_{xy}^2}, \tag{4.68}$$

$$\sigma_{xy} = \frac{\rho_{xy}}{\rho_{yy}^2 + \rho_{xy}^2}. \tag{4.69}$$

In the following argument, we explain why the longitudinal resistivity $\rho_{yy}$ in two dimensions has to vanish in the presence of a finite Hall resistivity $\rho_{xy}$. Since the edge state electrons with a given energy can only move in one direction, there is no backward scattering by obstacles as long as the edges are far apart from each other. No scattering between the two edges implies $\rho_{yy} = 0$ and hence $\sigma_{yy} = 0$. A finite resistivity can only occur when extended states are present in the bulk, such that the edge states on opposite edges are no longer spatially separated from each other.

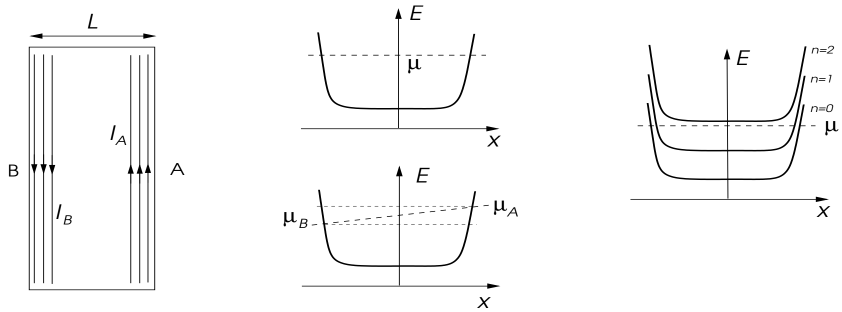



Figure 4.10: Edge state picture by Büttiker. The left panel gives a top view of the two dimensional system ($L = L_x$) where chiral edge states exist on both edges of the Hall bar with opposite chirality. On the middle panel, a single Landau level without and with transverse potential difference is shown, where the latter yields a finite net current due to current imbalance between left and right edge. The right panel visualizes the two lowest Landau levels which are occupied. All higher Landau levels are empty. This fixes the Hall conductance value $\nu_n$ to 2.

### 4.2.3 Fractional Quantum Hall Effect

Only two years after the discovery of the Integer Quantum Hall Effect, Störmer, Tsui and Gossard observed further series of plateaus of the Hall resistivity in a 2DEG realized with very high quality MOSFET inversion layers at low temperatures. The most pronounced of these plateaus is observed at a filling of $\nu = 1/3$ ($\sigma_{xy} = \nu e^2/h$). Later, an entire hierarchy of plateaus at fractional values $\nu = \nu_{p,m} = p/m$ with $p, m \in \mathbb{N}$ has been discovered,

$$\nu_{p,m} \in \left\{ \frac{1}{3}, \frac{2}{3}, \frac{2}{5}, \frac{3}{5}, \frac{3}{7}, \dots \right\}. \tag{4.70}$$

The emergence of these new plateaus is a clear evidence of the so-called Fractional Quantum Hall Effect (FQHE).

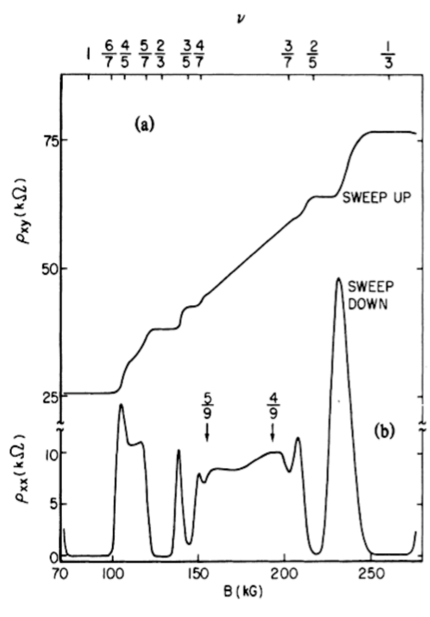



Figure 4.11: Experimental evidence of the Fractional Quantum Hall Effect.

It was again Laughlin who found the key concept to explain the FQHE. Unlike the IQHE, this new quantization feature can not be understood from a single-electron picture, but is based on the Coulomb repulsion among the electrons and the accompanying correlation effects. Laughlin specially investigated the case $\nu_{p,m} = 1/3$ and made the Ansatz

$$\Psi_{1/m}(z_1, \ldots, z_N) \propto \prod_{i<j} (z_i - z_j)^m \exp\left(-\sum_i \frac{|z_i|^2}{4\ell^2}\right) \tag{4.71}$$

for the $N$-body wave function, where $z = x - iy$ is a complex number representing the coordinates of the two-dimensional system. Limiting ourselves to the consideration of the lowest Landau level, this state gives a stable plateau with $\sigma_H = (1/3)e^2/h$, when $m = 3$.

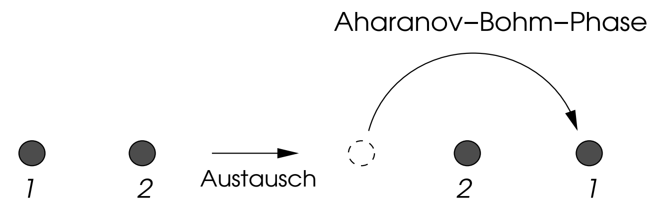



Figure 4.12: Exchange of two particles in two dimensions involves the motion of the particles around each other. There are two topologically distinct paths.

A heuristic interpretation of the Laughlin state was proposed by J. K. Jane and it is based on the concept of so-called composite fermions. In fact, Laughlin's state (4.71) can be written as

$$\Psi_{1/m} = \prod_{i<j} (z_i - z_j)^{m-1} \Psi_S \tag{4.72}$$

where $\Psi_S$ is the Slater determinant[6] describing the completely filled lowest Landau level. We see that the prefactor of $\Psi_S$ in equation (4.72) acts as a so-called *Jastro factor* that introduces

[6]The *Slater determinant* of the lowest Landau level is obtained from the states of the independent electrons. In symmetric gauge, the states are labelled by the quantum number $\tilde{m} \in \mathbb{N}_0$ and apart from the normalization (given in equation (4.47)) they are given by

$$\phi_{\tilde{m}}(z) = z^{\tilde{m}} e^{|z|^2/4\ell^2}, \tag{4.73}$$

where $z = x - iy$. The Slater determinant for $N$ independent electrons is

$$\Psi_S(z_1, \ldots, z_N) = \frac{1}{\sqrt{N!}} \det \begin{bmatrix} \phi_0(z_1) & \cdots & \phi_N(z_1) \\ \vdots & & \vdots \\ \phi_0(z_N) & \cdots & \phi_N(z_N) \end{bmatrix} \tag{4.74}$$

$$= \frac{1}{\sqrt{N!}} \det \begin{bmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^N \\ 1 & z_2 & z_2^2 & \cdots & z_2^N \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & z_N & z_N^2 & \cdots & z_N^N \end{bmatrix} \exp\left(-\sum_i \frac{|z_i|^2}{4\ell^2}\right). \tag{4.75}$$

The remaining determinant is a so-called *Vandermonde determinant*, which can be reexpressed in the form of a product, such that

$$\Psi_S = \prod_{i<j} (z_i - z_j) \exp\left(-\sum_i \frac{|z_i|^2}{4\ell^2}\right). \tag{4.76}$$

The prefactor is a homogenous polynomial with roots whenever $z_i = z_j$, which is a manifestation of the Pauli exclusion principle. We also see that the state $\Psi_S$ has a well defined total angular momentum $L_z = N\hbar$.

correlation effects into the wave function, since only the correlations due to the Pauli exclusion principle are contained in $\Psi_S$. The Jastro factor treats the Coulomb repulsion among the electrons and consequently leads to an additional suppression of the wave function whenever two electrons approach each other. In the form introduced above, it produces an phase factor for the electrons encircling each other. In particular, exchanging two electrons (see Fig. 4.12) leads to a phase

$$\exp(i(m-1)\pi) = \exp\left(i\frac{e}{\hbar c}\frac{m-1}{2}\Phi_0\right), \tag{4.77}$$

since $\Phi_0 = 2\pi\hbar c/e$. This phase has to be unity since the Slater state $\Psi_S$ is odd under exchange of two electrons. Therefore $m$ is restricted to odd integer values. This guarantees that the total wave function $\Psi_{1/m}$ still changes sign when two electrons are exchanged.

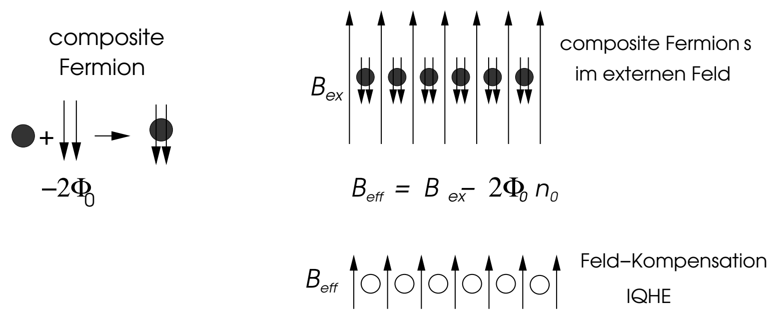



Figure 4.13: Sketch of the composite Fermion concept. Electrons with attached magnetic flux lines, here for the state of $\nu = 1/3$.

According to the Footnote 5 (see equation (4.44)), the case $\nu_{p,m} = 1/3$ implies that there are three flux quanta $\Phi_0$ per electron. In order to understand the FQHE, one constructs so-called *composite fermions* which do not interact with each other. Here, a composite fermion consists of an electron that has two (in fact $m-1$) negative flux quanta attached to it. These objects may be considered as independent fermions since the attached flux quanta compensate the Jastro factor in equation 4.72 through factors of the type $(z_i - z_j)^{-(m-1)}$. The exchange of two such composite fermions in two dimensions leads to an Aharanov-Bohm phase that is just opposed to that in equation (4.77). Due to the presence of the flux $-2\Phi_0$ per electron, the composite fermions are subject to an effective field composed of the external field and the attached flux quanta:

$$B_{\text{eff}} = B - \sum_i 2\Phi_0(z_i) \tag{4.78}$$

$$= \frac{1}{3}B - \left(\sum_i 2\Phi_0(z_i) - \frac{2}{3}B\right) \tag{4.79}$$

For an external field $B = 3n_0\Phi_0$, the expression in the brackets of equation (4.79) vanishes and the composite fermions feel an effective field $B_{\text{eff}} = n_0\Phi_0$ (Fig. (4.13)). Thus, these fermions form an Integer Quantum Hall state with $\nu = 1$ (for $B = 3n_0\Phi_0$), as discussed previously. This way of interpretation is applicable to other Fractional Quantum Hall states, too, since for $n$ filled Landau levels with composite fermions consisting of an electron with an attached flux of $-2k\Phi_0$, the effective field reads

$$B_{\text{eff}} = n_0\frac{\Phi_0}{\nu_{p,m}} - 2kn_0\Phi_0 = n_0\frac{\Phi_0}{n}. \tag{4.80}$$

From this we infer, that

$$\frac{1}{n} + 2k = \frac{1}{\nu_{p,m}} \tag{4.81}$$

or equivalently

$$\nu_{p,m} = \frac{p}{m} = \frac{n}{2kn+1}. \tag{4.82}$$

Despite the apparent simplicity of the treatment in terms of independent composite fermions, one should keep in mind that one is dealing with a strongly correlated electron system. The structure of the composite fermions is a manifestation of the fact that the fermions are not independent electrons. No composite fermions can exist in the vacuum, they can only arise within a certain many-body state. The Fractional Quantum Hall state also exhibits unconventional excitations with fractional charges. For example in the case $\nu_{p,m} = 1/3$, there are excitations with effective charge $e^* = e/3$. These are so-called 'topological' excitations, that can only exist in correlated systems. The Fractional Quantum Hall system is a very peculiar 'ordered' state of a two-dimensional electron system that has many interesting and complex properties.[7]

[7] Additional literature on the quantum Hall effects. For the Integer quantum Hall effect consult

- K. von Klitzing et al., Physik Journal **4** (6), 37 (2005)

while detailed literature on the fractional quantum Hall effect is found in

- R. Morf, Physik in unserer Zeit **33**, 21 (2002)
- J.K. Jain, Advances in Physics **41**, 105 (1992).

# Chapter 5

# Landau's Theory of Fermi Liquids

In the previous chapters, we considered the electrons of the system as more or less independent particles. The effect of their mutual interactions only entered via the renormalization of potentials and collective excitations. The underlying assumptions of our earlier discussions were that electrons in the presence of interactions can still be described as particles with a well-defined energy-momentum relation, and that their groundstate is a filled Fermi sea with a sharp Fermi surface. Since there is no guarantee that this hypothesis holds in general (and in fact they do not), we have to show that in metals the description of electrons as quasiparticles is justified. This quasiparticle picture will lead us to Landau's phenomenological theory of Fermi liquids.

## 5.1 Lifetime of quasiparticles

We first consider the lifetime of a state consisting of a filled Fermi sea to which one electron is added. Let $\boldsymbol{k}$ with $|\boldsymbol{k}| > k_F$ ($\epsilon_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ with $\epsilon_{\boldsymbol{k}} > \epsilon_F$) be the momentum (energy) of the additional electron. Due to interactions between the electrons, this state will decay into a many-body state. In momentum space such an interaction takes the form

$$\mathcal{H}_{ee} = \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}} \sum_{s,s'} V(\boldsymbol{q}) \widehat{c}^{\dagger}_{\boldsymbol{k}-\boldsymbol{q},s} \widehat{c}^{\dagger}_{\boldsymbol{k}'+\boldsymbol{q},s'} \widehat{c}_{\boldsymbol{k}',s'} \widehat{c}_{\boldsymbol{k},s}, \tag{5.1}$$

where $V(\boldsymbol{q})$ represents the electron-electron interaction in momentum space while $\boldsymbol{q}$ indicates the momentum transfer in the scattering process. Below, the short-ranged Yukawa potential

$$V(\boldsymbol{q}) = \frac{4\pi e^2}{q^2 \varepsilon(\boldsymbol{q},0)} = \frac{4\pi e^2}{q^2 + k_{TF}^2} \tag{5.2}$$

from equation (3.79) will be used. As we are only interested in very small energy transfers $\hbar\omega \ll \epsilon_F$, the static approximation is admissible.

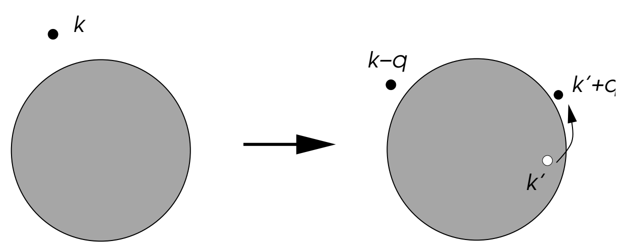



Figure 5.1: The decay of an electron state above the Fermi energy happens through scattering by creating particle-hole excitations.

In a perturbative treatment, the lowest order effect of the interaction is the creation of a particle-hole excitation in addition to the single electron above the Fermi energy. As the additional electron changes its momentum from $\boldsymbol{k}$ to $\boldsymbol{k}-\boldsymbol{q}$, a hole appears at $\boldsymbol{k}'$ and a second electron with wavevector $\boldsymbol{k}'+\boldsymbol{q}$ is created outside the Fermi sea. The transition is allowed whenever both energy and momentum are conserved, meaning

$$\boldsymbol{k} = (\boldsymbol{k}-\boldsymbol{q}) - \boldsymbol{k}' + (\boldsymbol{k}'+\boldsymbol{q}), \tag{5.3}$$

and

$$\epsilon_{\boldsymbol{k}} = \epsilon_{\boldsymbol{k}-\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'} + \epsilon_{\boldsymbol{k}'+\boldsymbol{q}}. \tag{5.4}$$

We calculate the lifetime $\tau_{\boldsymbol{k}}$ of the initial state with momentum $\boldsymbol{k}$ using Fermi's golden rule, yielding the transition rate from the initial state of a filled Fermi sea and one particle with momentum $\boldsymbol{k}$ to a state with two electrons above the Fermi sea, with momenta $\boldsymbol{k}-\boldsymbol{q}$ and $\boldsymbol{k}'+\boldsymbol{q}$, and a hole with $\boldsymbol{k}'$, as shown in Fig. 5.1. Since neither the momenta $\boldsymbol{k}'$ and $\boldsymbol{q}$, nor the spin of the created electron are fixed, a summation over the possible configuration has to be performed, leading to

$$\frac{1}{\tau_{\boldsymbol{k}}} = \frac{2\pi}{\hbar}\frac{1}{\Omega^2}\sum_{\boldsymbol{k}',\boldsymbol{q}}\sum_{s'}|V(\boldsymbol{q})|^2\, n_{0,\boldsymbol{k}'}(1-n_{0,\boldsymbol{k}-\boldsymbol{q}})(1-n_{0,\boldsymbol{k}'+\boldsymbol{q}})\delta(\epsilon_{\boldsymbol{k}-\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - (\epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}'+\boldsymbol{q}})). \tag{5.5}$$

Note that the term $n_{0,\boldsymbol{k}'}(1-n_{0,\boldsymbol{k}-\boldsymbol{q}})(1-n_{0,\boldsymbol{k}'+\boldsymbol{q}})$ takes care of the Pauli principle, by ensuring that the final state after the scattering process exists, i.e. the hole state $\boldsymbol{k}'$ lies inside and the two particle states $\boldsymbol{k}-\boldsymbol{q}$ and $\boldsymbol{k}'+\boldsymbol{q}$ lie outside the Fermi sea. First the integral over $\boldsymbol{k}'$ is performed under the condition that the energy $\epsilon_{\boldsymbol{k}'+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'}$ of the excitation is small. With that, the integral reduces to

$$S(\omega_{\boldsymbol{q},\boldsymbol{k}}, q) = \frac{1}{\Omega}\sum_{\boldsymbol{k}'} n_{0,\boldsymbol{k}'}(1-n_{0,\boldsymbol{k}'+\boldsymbol{q}})\delta(\epsilon_{\boldsymbol{k}-\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - (\epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}'+\boldsymbol{q}})) \tag{5.6}$$

$$= \frac{1}{(2\pi)^3}\int d^3k'\; n_{0,\boldsymbol{k}'}(1-n_{0,\boldsymbol{k}'+\boldsymbol{q}})\delta(\epsilon_{\boldsymbol{k}'+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'} - \hbar\omega_{\boldsymbol{q},\boldsymbol{k}}) \tag{5.7}$$

$$= \frac{N(\epsilon_F)}{4}\frac{\omega_{\boldsymbol{q},\boldsymbol{k}}}{qv_F} \tag{5.8}$$

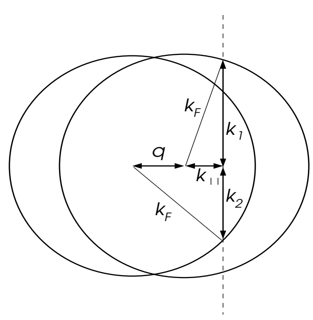
where $N(\epsilon_F) = mk_F/\pi^2\hbar^2$ is the density of states of the electrons at the Fermi surface and $\hbar\omega_{\boldsymbol{q},\boldsymbol{k}} = \hbar^2(2\boldsymbol{k}\cdot\boldsymbol{q} - \boldsymbol{q}^2)/2m > 0$ is the energy loss of the decaying electron.[1] In order to compute

[1] Small $\omega$ are justified, because $\hbar\omega \leq (2k_F q - q^2)/2m$ for most allowed $\omega$. The integral may be computed using cylindrical coordinates, where the vector $\boldsymbol{q}$ points along the axis of the cylinder. It results in

$$S(\boldsymbol{q},\omega) = \frac{1}{(2\pi)^2}\int_{k_2}^{k_1} dk'_\perp k'_\perp \int_0^{k_F} dk'_\parallel\; \delta\left(\frac{\hbar^2 q^2}{2m} + \frac{\hbar^2 q k'_\parallel}{m} - \hbar\omega\right) \tag{5.9}$$

$$= \frac{m}{4\pi^2\hbar^2 q}\left(k_1^2 - k_2^2\right), \tag{5.10}$$

with $k_1^2 = k_F^2 - k_{\parallel,0}^2$ and $k_2^2 = k_F^2 - (k_{\parallel,0} + q)^2$, where $k_{\parallel,0} = (2m\omega - \hbar q^2)/2\hbar q$ is enforced by the delta function.




The wave vectors $k_2$ and $k_1$ are the upper and lower limits of integration determined from the condition $n_{0,\boldsymbol{k}'}(1 - n_{0,\boldsymbol{k}'+\boldsymbol{q}}) > 0$ and can be obtained by simple geometric considerations. equation (5.8) follows immediately.

the remaining integral over $\boldsymbol{q}$, we assume that the matrix element $|V(\boldsymbol{q})|^2$ depends only weakly on $\boldsymbol{q}$ when $q \ll k_F$. This is especially true when the interaction is short-ranged. In spherical coordinates, the integral reads

$$\frac{1}{\tau_{\boldsymbol{k}}} = \frac{2\pi}{\hbar} \cdot \frac{N(\epsilon_F)}{4v_F\Omega} \sum_{\boldsymbol{q},s'} |V(\boldsymbol{q})|^2 \frac{\omega_{\boldsymbol{q},\boldsymbol{k}}}{q} \tag{5.11}$$

$$= \frac{N(\epsilon_F)}{(2\pi)^2 2\hbar v_F} \int d^3\boldsymbol{q}\, |V(\boldsymbol{q})|^2 \frac{\omega_{\boldsymbol{q},\boldsymbol{k}}}{q} \tag{5.12}$$

$$= \frac{N(\epsilon_F)}{(2\pi)4mv_F} \int dq\, |V(q)|^2 q^2 \int_{\theta_1}^{\theta_2} d\theta \sin\theta (2k\cos\theta - q) \tag{5.13}$$

$$= \frac{N(\epsilon_F)}{(2\pi)4mv_F} \int dq\, |V(q)|^2 q^2 \left[ -\frac{1}{4k} (2k\cos\theta - q)^2 \right]_{\theta_1}^{\theta_2} . \tag{5.14}$$

The restriction of the domain of integration of $\theta$ follows from the two conditions $k^2 \geq (\boldsymbol{k}-\boldsymbol{q})^2 \geq k_F^2$ and $(\boldsymbol{k}-\boldsymbol{q})^2 = k^2 - 2kq\cos\theta + q^2$. From the first condition, $\cos\theta_2 = q/2k$, and from the second, $\cos\theta_1 = (k^2 - k_F^2 + q^2)/2kq$. Thus,

$$\frac{1}{\tau_{\boldsymbol{k}}} = \frac{N(\epsilon_F)}{(2\pi)4mv_F} \int dq\, |V(q)|^2 \frac{1}{4k} \left(k^2 - k_F^2\right)^2 \tag{5.15}$$

$$\approx \frac{N(\epsilon_F)}{(2\pi)4v_F} \frac{m}{k_F} \frac{1}{\hbar^4} (\epsilon_{\boldsymbol{k}} - \epsilon_F)^2 \int dq\, |V(q)|^2 \tag{5.16}$$

$$= \frac{1}{8\pi\hbar^3} \frac{N(\epsilon_F)}{v_F^2} (\epsilon_{\boldsymbol{k}} - \epsilon_F)^2 \int dq\, |V(q)|^2 . \tag{5.17}$$

Note that convergence of the last integral over $q$ requires that the integrand does not diverge stronger than $q^\alpha$ ($\alpha < 1$) for $q \to 0$. With the dielectric constant obtained in the previous chapter, this condition is certainly fulfilled. Essentially, the result states that

$$\frac{1}{\tau_{\boldsymbol{k}}} \propto (\epsilon_{\boldsymbol{k}} - \epsilon_F)^2 \tag{5.18}$$

for $\boldsymbol{k}$ slightly above the Fermi surface. This implies that the state $|\boldsymbol{k}s\rangle$ occurs as a resonance of width $\hbar/\tau_{\boldsymbol{k}}$ and features a quasiparticle, which can be observed in the spectral function $A(E,\boldsymbol{k})$ as depicted in Fig.5.2.[2] The quasiparticle (coherent) part of the spectral function has a weight reduced from one (corresponding to the quasiparticle weight $Z_{\boldsymbol{k}}$). The remaining weight is shifted to higher energies as a so-called incoherent part (continuum without clear momentum-energy relation).
The resonance becomes arbitrarily sharp as the Fermi surface is approached

$$\lim_{k\to k_F} \frac{\hbar/\tau_{\boldsymbol{k}}}{\epsilon_{\boldsymbol{k}} - \epsilon_F} = 0, \tag{5.21}$$

[2]The spectral function is defined as

$$A(E,\boldsymbol{k}) \propto \sum_n |\langle\Psi_n|\hat{c}_{\boldsymbol{k}s}|\Psi_0\rangle|^2 \,\delta(E - E_n) \tag{5.19}$$

where $|\Psi_0\rangle$ is the exact (renormalized) ground state and $|\Psi_n\rangle$ are the corresponding exact excited states. The coherent part of the spectral function can be represented as a Lorentzian form

$$A_{\rm coh}(E,\boldsymbol{k}) = \frac{Z_{\boldsymbol{k}}\hbar}{\pi\tau_{\boldsymbol{k}}} \frac{1}{(E - \tilde{\epsilon}_{\boldsymbol{k}})^2 + \frac{\hbar^2}{\tau_{\boldsymbol{k}^2}}} , \tag{5.20}$$

where $Z_{\boldsymbol{k}}$ is the quasiparticle weight, smaller than 1.

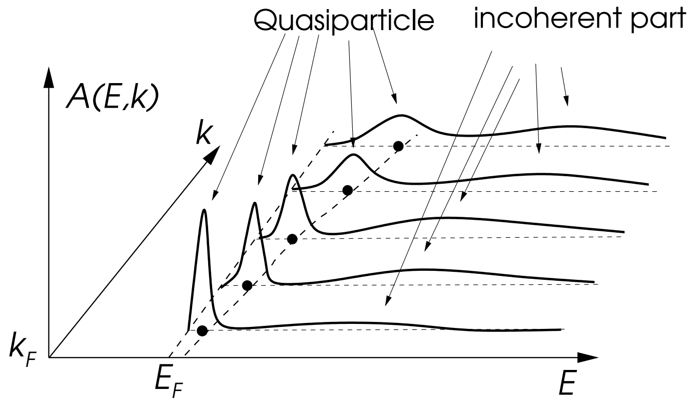



Figure 5.2: Quasiparticle spectrum: Quasiparticle peaks appear the sharper the closer the energy lies to the Fermi energy. The area under the "sharp" quasiparticle peak corresponds to the quasiparticle weight. The missing quasiparticle weight is transferred to higher energies.

so that the quasiparticle concept is asymptotically valid. The equation (5.21) can also be seen as a verification of Heisenberg's uncertainty principle. Consequently, the momentum of an electron is a good quantum number in the vicinity of the Fermi surface. Underlying this result is the Pauli exclusion principle, which restricts the phase space for decay processes of single particle states close to the Fermi surface. In addition, the assumption of short ranged interactions is crucial. Long ranged interactions can change the behavior drastically due to the larger number of decay channels.

## 5.2 Phenomenological Theory of Fermi Liquids

The existence of well-defined fermionic quasiparticles in spite of the underlying complex many-body physics inspired Landau to the following phenomenological theory. Just like the states of independent electrons, quasiparticle states shall be characterized by their momentum $\boldsymbol{k}$ and spin $\sigma$. In fact, there is a one-to-one mapping between the free electrons and the quasiparticles. Consequently, the number of quasiparticles and the number of electrons coincide. The momentum distribution function of quasiparticles, defined as $n_\sigma(\boldsymbol{k})$, is subject to the condition

$$N = \sum_{\boldsymbol{k},\sigma} n_\sigma(\boldsymbol{k}). \tag{5.22}$$

In analogy to the Fermi-Dirac distribution of free electrons, one demands, that the ground state distribution function $n_\sigma^{(0)}(\boldsymbol{k})$ for the quasiparticles is described by a simple step function

$$n_\sigma^{(0)}(\boldsymbol{k}) = \Theta(k_F - |\boldsymbol{k}|). \tag{5.23}$$

For a spherically symmetric electron system, the quasiparticle Fermi surface is a sphere with the same radius as the one for free electrons of the same density. For a general point group symmetry, the Fermi surface may be deformed by the interactions without changing the underlying symmetry. The volume enclosed by the Fermi surface is always conserved despite the deformation.[3] Note that the distribution $n_\sigma^{(0)}(\boldsymbol{k})$ of the quasiparticles in the ground state and that $n_{0\boldsymbol{k}s} = \langle \hat{c}_{\boldsymbol{k}\sigma}^\dagger \hat{c}_{\boldsymbol{k}\sigma} \rangle$ of the real electrons in the ground state are not identical (Figure 5.3). Interestingly, $n_{0\boldsymbol{k}s}$ is still discontinuous at the Fermi surface, but the height of the jump is, in general, smaller than unity. The modification of the electron distribution function from a step

[3]This is the content of the Luttinger theorem [J.M. Luttinger, Phys. Rev. **119**, 1153 (1960)].

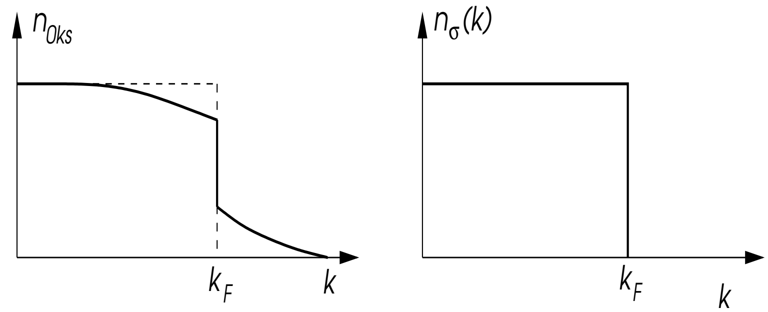



Figure 5.3: Schematic picture of the distribution function: Left panel: modified distribution function of the original electron states; right panel: distribution function of quasiparticle states making a simple step function.

function to a "smoother" Fermi surface indicates the involvement of electron-hole excitations and the renormalization of the electronic properties, which deplete the Fermi sea and populate the states above the Fermi level. The reduced jump in $n_{0\boldsymbol{k}s}$ is a measure for the quasiparticle weight at the Fermi surface, $Z_{\boldsymbol{k}_F}$, i.e. the amplitude of the corresponding free electron state in the quasiparticle state.

In Landau's theory of Fermi liquids, the essential information on the low-energy physics of the system shall be contained in the deviation of the quasiparticle distribution $n_\sigma(\boldsymbol{k})$ from its ground state distribution $n_\sigma^{(0)}(\boldsymbol{k})$,

$$\delta n_\sigma(\boldsymbol{k}) = n_\sigma(\boldsymbol{k}) - n_\sigma^{(0)}(\boldsymbol{k}). \tag{5.24}$$

The symbol $\delta$ is generally used in literature to denote this difference. Unfortunately this may suggest that the term $\delta n_\sigma(\boldsymbol{k})$ is small, which is not true in general. Indeed, $\delta n_\sigma(\boldsymbol{k})$ is concentrated on momenta $\boldsymbol{k}$ very close to the Fermi energy only, where the quasiparticle concept is valid. This distribution function, describing the deviation from the ground state, enters a phenomenological energy functional of the form

$$E = E_0 + \sum_{\boldsymbol{k},\sigma} \epsilon_\sigma(\boldsymbol{k})\delta n_\sigma(\boldsymbol{k}) + \frac{1}{2\Omega}\sum_{\boldsymbol{k},\boldsymbol{k}'}\sum_{\sigma,\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_\sigma(\boldsymbol{k})\delta n_{\sigma'}(\boldsymbol{k}') + O(\delta n^3) \tag{5.25}$$

where $E_0$ denotes the energy of the ground state. Moreover, the phenomenological parameters $\epsilon_\sigma(\boldsymbol{k})$ and $f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')$ have to be determined by experiments or by means of a microscopic theory. The variational derivative

$$\tilde{\epsilon}_\sigma(\boldsymbol{k}) = \frac{\delta E}{\delta n_\sigma(\boldsymbol{k})} = \epsilon_\sigma(\boldsymbol{k}) + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_{\sigma'}(\boldsymbol{k}') \tag{5.26}$$

yields an effective energy-momentum relation $\tilde{\epsilon}_\sigma(\boldsymbol{k})$, whose second term depends on the distribution of all quasiparticles. A quasiparticle moves in the "mean-field" of all other quasiparticles, so that changes $\delta n_\sigma(\boldsymbol{k})$ in the distribution affect $\tilde{\epsilon}_\sigma(\boldsymbol{k})$. The second variational derivative describes the coupling between the quasiparticles

$$\frac{\delta^2 E}{\delta n_\sigma(\boldsymbol{k})\delta n_{\sigma'}(\boldsymbol{k}')} = \frac{1}{\Omega} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}'). \tag{5.27}$$

We introduce a parametrization for these couplings $f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')$ by assuming spherical symmetry of the system. Furthermore, the radial dependence is ignored, as we only consider quasiparticles in the vicinity of the Fermi surface where $|\boldsymbol{k}|,|\boldsymbol{k}'| \approx k_F$. Therefore the dependence of $f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')$ on $\boldsymbol{k},\boldsymbol{k}'$ can be reduced to the relative angle $\theta_{\hat{k},\hat{k}'}$

$$f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}') = f^s(\hat{k},\hat{k}') + \sigma\sigma' f^a(\hat{k},\hat{k}') \tag{5.28}$$

where $\hat{k} = \boldsymbol{k}/|\boldsymbol{k}|$. The symmetric ($s$) and antisymmetric ($a$) part of $f_{\sigma\sigma'}(\boldsymbol{k}, \boldsymbol{k}')$ can be expanded in Legendre-polynomials $P_l(z)$, leading to

$$f^{s,a}(\hat{k}, \hat{k}') = \sum_{l=0}^{\infty} f_l^{s,a} P_l(\cos\theta_{\hat{k},\hat{k}'}). \tag{5.29}$$

The density of states at the Fermi surface is defined as

$$N(\epsilon_F) = \frac{2}{\Omega} \sum_{\boldsymbol{k}} \delta(\epsilon(\boldsymbol{k}) - \epsilon_F) = \frac{k_F^2}{\pi^2 \hbar v_F} = \frac{m^* k_F}{\pi^2 \hbar^2} \tag{5.30}$$

and follows from the dispersion $\epsilon(\boldsymbol{k})$ of the bare quasiparticle energy

$$\boldsymbol{\nabla}_{\boldsymbol{k}}\, \epsilon(\boldsymbol{k})|_{\boldsymbol{k}_F} = \boldsymbol{v_F} = \frac{\hbar \boldsymbol{k}_F}{m^*} \tag{5.31}$$

With this definition, we also introduce the so-called *Landau parameters*

$$F_l^s = N(\epsilon_F) f_l^s, \tag{5.32}$$
$$F_l^a = N(\epsilon_F) f_l^a, \tag{5.33}$$

commonly used in the literature.[4] In the following, we want to study the relation between the different phenomenological parameters of Landau's theory of Fermi liquids and the experimentally accessible quantities of a real system, such as specific heat, compressibility, spin susceptibility among others.

### 5.2.1 Specific heat - Density of states

Since the quasiparticles are fermions, they obey Fermi-Dirac statistics

$$n_\sigma(0)(T, \boldsymbol{k}) = \frac{1}{e^{[\tilde{\epsilon}(\boldsymbol{k}) - \mu]/k_B T} + 1} \tag{5.34}$$

with the chemical potential $\mu$. For low temperatures, we find

$$\delta n_\sigma(\boldsymbol{k}) = n_\sigma^{(0)}(T, \boldsymbol{k}) - n_\sigma^{(0)}(0, \boldsymbol{k}) \tag{5.35}$$

which leads to

$$\frac{1}{\Omega} \sum_{\boldsymbol{k}} \delta n_\sigma(\boldsymbol{k}) \propto T^2 + O(T^4). \tag{5.36}$$

In the limit $T \longrightarrow 0$, the left-hand side of equation (5.36) will vanish quadratically in $T$. Remember, that this does not mean that $\delta n_\sigma(\boldsymbol{k})$ is small. To leading order in $T$, the distribution (5.34) can be replaced by

$$n_\sigma(\boldsymbol{k}) = \frac{1}{e^{[\epsilon(\boldsymbol{k}) - \epsilon_F]/k_B T} + 1} \tag{5.37}$$

with the bare quasiparticle energy $\epsilon_\sigma(\boldsymbol{k})$ in place of the renormalized dispersion $\tilde{\epsilon}_\sigma(\boldsymbol{k})$, as $\tilde{\epsilon}$ differs from $\epsilon$ only by the order $T^2$. Similarly we replaced $\mu = \epsilon_F + O(T^2)$ by $\epsilon_F$. When focussing on leading terms, which are usually of the order $T^0$ and $T^1$, corrections of higher order may be neglected in the low-temperature regime. In order to discuss the specific heat, we employ the

[4] Another frequently used notation for the Landau parameters in the literature is $F_l = F_l^s$ and $Z_l = F_l^a$.

expression for the entropy of a fermion gas. For each quasiparticles with a given spin there is one state labelled by $\boldsymbol{k}$. The entropy density may be computed from the distribution function

$$S = -\frac{k_B}{\Omega}\sum_{\boldsymbol{k},\sigma}\left[n_\sigma(\boldsymbol{k})\ln\left(n_\sigma(\boldsymbol{k})\right) + \left(1 - n_\sigma(\boldsymbol{k})\right)\ln\left(1 - n_\sigma(\boldsymbol{k})\right)\right]. \tag{5.38}$$

Taking the derivative of the entropy $S$ with respect to $T$, the specific heat

$$C(T) = T\frac{\partial S}{\partial T} \tag{5.39}$$

$$= -\frac{k_B T}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{e^{\xi(\boldsymbol{k})/k_B T}}{(\,e^{\xi(\boldsymbol{k})/k_B T}+1\,)^2}\,\frac{\xi(\boldsymbol{k})}{k_B T^2}\ln\left(\frac{n_\sigma(\boldsymbol{k})}{1-n_\sigma(\boldsymbol{k})}\right) \tag{5.40}$$

$$= -\frac{k_B T}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{1}{4\cosh\left(\xi(\boldsymbol{k})/2k_B T\right)^2}\,\frac{\xi(\boldsymbol{k})}{k_B T^2}\,\frac{\xi(\boldsymbol{k})}{k_B T} \tag{5.41}$$

is obtained, where we introduced $\xi(\boldsymbol{k}) = \epsilon(\boldsymbol{k}) - \epsilon_F$. In the limit $T \to 0$ we find

$$\frac{C(T)}{T} \approx \frac{N(\epsilon_F)}{4k_B T^3}\int d\xi\frac{\xi^2}{\cosh^2(\xi/2k_B T)} \tag{5.42}$$

$$\approx \frac{k_B^2 N(\epsilon_F)}{4}\int_{-\infty}^{+\infty} dy\frac{y^2}{\cosh^2(y/2)} \tag{5.43}$$

$$= \frac{\pi^2 k_B^2 N(\epsilon_F)}{3}, \tag{5.44}$$

which is the well-known linear behavior $C(T) = \gamma T$ for the specific heat at low temperatures, with $\gamma = \pi^2 k_B^2 N(\epsilon_F)/3$. Since $N(\epsilon_F) = m^* k_F/\pi^2\hbar^2$, the effective mass $m^*$ of the quasiparticles can directly determined by measuring the specific heat of the system.

### 5.2.2 Compressibility - Landau parameter $\mathbf{F_0^s}$

A Fermi gas has a finite compressibility because each fermion occupies a finite amount of space due to the Pauli principle. The compressibility $\kappa$ is defined as[5]

$$\kappa = -\frac{1}{\Omega}\left(\frac{\partial\Omega}{\partial p}\right)_{T,N} \tag{5.46}$$

where $p$ is the uniform hydrostatic pressure. The indexed $T, N$ means, that the temperature $T$ and the particle number $N$ are kept fixed. We consider the response of the Fermi liquid upon application of uniform pressure $p$. The shift of the quasiparticle energies is given by

$$\delta\epsilon(\boldsymbol{k}) = \frac{\partial\epsilon(\boldsymbol{k})}{\partial p}\delta p = \frac{\partial\epsilon(\boldsymbol{k})}{\partial\boldsymbol{k}}\cdot\frac{\partial\boldsymbol{k}}{\partial\Omega}\frac{\partial\Omega}{\partial p}\delta p = \frac{\kappa^{(0)}}{3}\hbar\boldsymbol{v}_{\boldsymbol{k}}\cdot\boldsymbol{k}\,\delta p = \gamma_{\boldsymbol{k}}\kappa^{(0)}\delta p \tag{5.47}$$

[5] An alternative definition considers the change of particle number upon change of the chemical potential,

$$\kappa = \frac{1}{n^2}\left(\frac{\partial n}{\partial\mu}\right)_{T,\Omega} \tag{5.45}$$

with $n = N/\Omega$.

with $\gamma_{\boldsymbol{k}} = \hbar \boldsymbol{v}_{\boldsymbol{k}} \cdot \boldsymbol{k}/3 = 2\epsilon_\sigma(\boldsymbol{k})/3$. Analogous we introduce the shift of the renormalized quasi-particle energies,

$$\begin{aligned}
\delta\tilde{\epsilon}_\sigma(\boldsymbol{k}) &= \gamma_{\boldsymbol{k}}\kappa\delta p = \gamma_{\boldsymbol{k}}\kappa^{(0)}\delta p + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma,\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_{\sigma'}(\boldsymbol{k}') \\
&= \gamma_{\boldsymbol{k}}\kappa_\delta p + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma,\sigma'}(\boldsymbol{k},\boldsymbol{k}')\frac{\partial n_{\sigma'}(\boldsymbol{k}')}{\partial\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}')}\delta\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}') \qquad . \qquad (5.48) \\
&= \gamma_{\boldsymbol{k}}\kappa^{(0)}\delta p - \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma,\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta(\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}') - \epsilon_F)\gamma_{\boldsymbol{k}'}\kappa\ \delta p
\end{aligned}$$

Changes are concentrated on the Fermi surface such that we can replace $\gamma_{\boldsymbol{k}} = 2\epsilon_F/3$ so that

$$\kappa = \kappa^{(0)} - \kappa N(\epsilon_F)\int \frac{d\Omega_{\hat{k}'}}{4\pi} f^s(\hat{k},\hat{k}') = \kappa^{(0)} - \kappa F_0^s\ . \qquad (5.49)$$

Therefore we find

$$\kappa = \frac{\kappa^{(0)}}{1 + F_0^s}\ . \qquad (5.50)$$

Now we determine $\kappa^{(0)}$ from the volume dependence of the energy

$$E^{(0)} = \sum_{\boldsymbol{k},\sigma}\epsilon_\sigma(\boldsymbol{k}) = \frac{3}{5}N\epsilon_F = \frac{3}{5}N\frac{\hbar^2 k_F^2}{2m^*} = \frac{3}{10}\frac{\hbar^2 N}{m^*}\left(3\pi^2\frac{N}{\Omega}\right)^{2/3} \quad . \qquad (5.51)$$

Then we determine the pressure

$$p = -\left(\frac{\partial E^{(0)}}{\partial\Omega}\right)_N = \frac{1}{5}\frac{\hbar^2 N}{m^*}\left(3\pi^2\frac{N}{\Omega}\right)^{2/3}\frac{1}{\Omega} \qquad (5.52)$$

and

$$\frac{1}{\kappa^{(0)}} = -\Omega\frac{\partial p}{\partial\Omega} = \frac{1}{3}\frac{\hbar^2 N}{m^*\Omega}(3\pi^2 n)^{2/3} = \frac{2}{3}n\epsilon_F\ . \qquad (5.53)$$

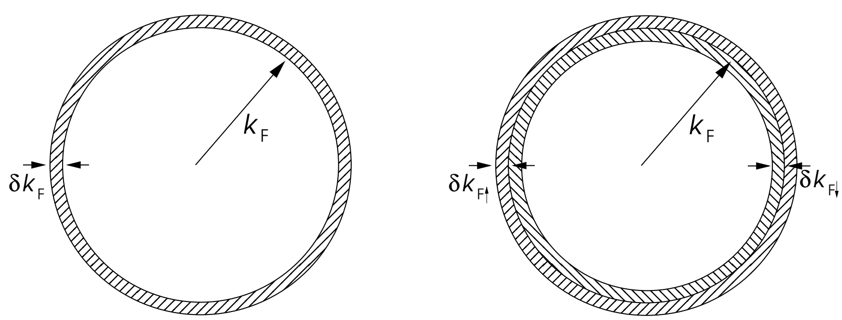



Figure 5.4: Deviations of the distribution functions: Left panel: isotropic increase of the Fermi surface as used for the uniform compressibility; right panel: spin dependent change of size of the Fermi surface as used for the uniform spin susceptibility.

### 5.2.3 Spin susceptibility - Landau parameter $\mathbf{F_0^a}$

If a magnetic field $H$ is applied to the system, the Zeeman coupling yields a shift of the quasiparticle energies

$$\begin{aligned}\delta\epsilon_\sigma(\boldsymbol{k}) &= \epsilon_\sigma(\boldsymbol{k}) - \epsilon(\boldsymbol{k}) \\ &= -g\mu_B H\frac{\sigma}{2},\end{aligned} \tag{5.54}$$

depending on their spin orientation. The term $\epsilon(\boldsymbol{k})$ represents the quasiparticle spectrum at $H = 0$. As we work with an isotropic system, we assumed without loss of generality that the magnetic field points along the $z$-direction. The energy shift due to the applied field is

$$\begin{aligned}\delta\tilde{\epsilon}_\sigma(\boldsymbol{k}) &= \tilde{\epsilon}_\sigma(\boldsymbol{k}) - \epsilon(\boldsymbol{k}) \\ &= -g\mu_B H\frac{\sigma}{2} + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma,\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_{\sigma'}(\boldsymbol{k}') \\ &= -\tilde{g}\mu_B H\frac{\sigma}{2}.\end{aligned} \tag{5.55}$$

Due to interactions, the renormalized gyromagnetic factor $\tilde{g}$ differs from the value of $g = 2$ for free electrons. We focus on the second term in Eq.(5.55), which can be reexpressed as

$$\begin{aligned}\frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_{\sigma'}(\boldsymbol{k}') &= \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\frac{\partial n_{\sigma'}(\boldsymbol{k}')}{\partial\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}')}\delta\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}') \\ &= \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta(\tilde{\epsilon}_{\sigma'}(\boldsymbol{k}') - \epsilon_F)\tilde{g}\mu_B H\frac{\sigma'}{2}\end{aligned} \tag{5.56}$$

Combining this result with the Eqs. (5.55) and (5.56), we derive

$$\tilde{g} = g - \tilde{g}N(\epsilon_F)\int\frac{d\Omega_{\hat{k}'}}{4\pi}f^a(\hat{k},\hat{k}') = g - \tilde{g}F_0^a, \tag{5.57}$$

or equivalently

$$\tilde{g} = \frac{g}{1 + F_0^a}. \tag{5.58}$$

The magnetization of the system can be computed from the distribution function,

$$\begin{aligned}M &= g\mu_B\sum_{\boldsymbol{k},\sigma}\frac{\sigma}{2}\delta n_\sigma(\boldsymbol{k}) = g\mu_B\sum_{\boldsymbol{k},\sigma}\frac{\sigma}{2}\frac{\partial n_\sigma(\boldsymbol{k})}{\partial\tilde{\epsilon}_\sigma(\boldsymbol{k})}\delta\tilde{\epsilon}_\sigma(\boldsymbol{k}) \\ &= g\mu_B\sum_{\boldsymbol{k},\sigma}\frac{\sigma}{2}\delta(\tilde{\epsilon}_\sigma(\boldsymbol{k}) - \epsilon_F)\tilde{g}\mu_B H\frac{\sigma}{2}\end{aligned} \tag{5.59}$$

from which the susceptibility is immediately found to be

$$\chi = \frac{M}{H\Omega} = \frac{\mu_B^2 N(\epsilon_F)}{1 + F_0^a}. \tag{5.60}$$

The changes in the distribution function induced by the magnetic field feed back into the susceptibility, so that the latter may be either weakened ($F_0^a > 0$) or enhanced ($F_0^a < 0$). For the magnetic susceptibility, the Landau parameter $F_0^a$ and the effective mass $m^*$ (through $N(\epsilon_F)$) lead to a renormalization compared to the free electron susceptibility.

### 5.2.4 Galilei invariance - effective mass and $\mathbf{F_1^s}$

We initially introduced by hand the effective mass of quasiparticles in $\epsilon_\sigma(\boldsymbol{k})$. In this section we show, that overall consistency of the phenomenological theory requires a relation between the effective mass and one Landau parameter ($F_1^s$). The reason is, that the effective mass is the result of the interactions among the electrons. This self-consistency is connected with the Galilean invariance of the system. When the momenta of all particles are shifted by $\hbar\boldsymbol{q}$ ($|\boldsymbol{q}|$ shall be very small compared to the Fermi momentum $k_F$ in order to remain within the assumption-range of the Fermi liquid theory) the distribution function given by

$$\delta n_\sigma(\boldsymbol{k}) = n_\sigma^{(0)}(\boldsymbol{k}+\boldsymbol{q}) - n_\sigma^{(0)}(\boldsymbol{k}) \approx \boldsymbol{q}\cdot\boldsymbol{\nabla_k}\, n_\sigma^{(0)}(\boldsymbol{k}). \tag{5.61}$$

This function is strongly concentrated around the Fermi energy (see Figure 5.5).

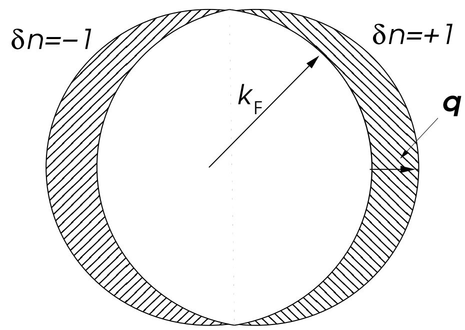



Figure 5.5: Distribution function due to a Fermi surface shift (Galilei transformation).

The current density can now be calculated, using the distribution function $n_\sigma(\boldsymbol{k}) = n_\sigma^{(0)}(\boldsymbol{k}) + \delta n_\sigma(\boldsymbol{k})$. Within the Fermi liquid theory this yields,

$$\boldsymbol{j_q} = \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma} \boldsymbol{v}(\boldsymbol{k}) n_\sigma(\boldsymbol{k}) \tag{5.62}$$

with

$$\begin{aligned}\boldsymbol{v}(\boldsymbol{k}) &= \frac{1}{\hbar}\boldsymbol{\nabla_k}\,\tilde{\epsilon}_\sigma(\boldsymbol{k}) \\ &= \frac{1}{\hbar}\left(\boldsymbol{\nabla_k}\,\epsilon_\sigma(\boldsymbol{k}) + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'}\boldsymbol{\nabla_k}\, f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_\sigma(\boldsymbol{k}')\right).\end{aligned} \tag{5.63}$$

Thus we obtain for the current density,

$$\begin{aligned}\boldsymbol{j_q} &= \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{\hbar\boldsymbol{k}}{m^*}n_\sigma(\boldsymbol{k}) + \frac{1}{\Omega^2}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'}[n_\sigma^{(0)}(\boldsymbol{k}) + \delta n_\sigma(\boldsymbol{k})]\frac{1}{\hbar}\boldsymbol{\nabla_k}\, f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_\sigma(\boldsymbol{k}') \\ &= \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{\hbar\boldsymbol{k}}{m^*}\delta n_\sigma(\boldsymbol{k}) - \frac{1}{\Omega^2}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'}\frac{1}{\hbar}[\boldsymbol{\nabla_k}\, n_\sigma^{(0)}(\boldsymbol{k})] f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta n_\sigma(\boldsymbol{k}') + O(q^2) \\ &= \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{\hbar\boldsymbol{k}}{m^*}\delta n_\sigma(\boldsymbol{k}) + \frac{1}{\Omega^2}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta(\epsilon_\sigma(\boldsymbol{k}') - \epsilon_F)\frac{\hbar\boldsymbol{k}'}{m^*}\delta n_\sigma(\boldsymbol{k}) + O(q^2) = \boldsymbol{j}_1 + \boldsymbol{j}_2.\end{aligned} \tag{5.64}$$

where, for the second line, we performed an integration by parts and neglect terms quadratic in $\delta n$ and, in the third line, used $f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}') = f_{\sigma'\sigma}(\boldsymbol{k}',\boldsymbol{k})$ and

$$\boldsymbol{\nabla_k}\, n_\sigma^{(0)}(\boldsymbol{k}) = \frac{\partial n_\sigma^{(0)}(\boldsymbol{k})}{\partial\epsilon_\sigma(\boldsymbol{k})}\boldsymbol{\nabla_k}\,\epsilon_\sigma(\boldsymbol{k}) = -\delta(\epsilon_\sigma(\boldsymbol{k}) - \epsilon_F)\boldsymbol{\nabla_k}\,\epsilon_\sigma(\boldsymbol{k}) = -\delta(\epsilon_\sigma(\boldsymbol{k}) - \epsilon_F)\frac{\hbar^2\boldsymbol{k}}{m^*}. \tag{5.65}$$

The first term of equation (5.65) denotes quasiparticle current, $\boldsymbol{j}_1$, while the second term can be interpreted as a drag current, $\boldsymbol{j}_2$, an induced motion (backflow) of the other particles due to interactions.

From a different viewpoint, we consider the system as being in the inertial frame with a velocity $\hbar\boldsymbol{q}/m$, as all particles received the same momentum. Here $m$ is the bare electron mass. The current density is then given by

$$\boldsymbol{j}_{\boldsymbol{q}} = \frac{N}{\Omega}\frac{\hbar\boldsymbol{q}}{m} = \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{\hbar\boldsymbol{k}}{m}n_\sigma(\boldsymbol{k}) = \frac{1}{\Omega}\sum_{\boldsymbol{k},\sigma}\frac{\hbar\boldsymbol{k}}{m}\delta n_\sigma(\boldsymbol{k}). \tag{5.66}$$

Since these two viewpoints have to be equivalent, the resulting currents should be the same. Thus, we compare equation (5.65) and (5.66) and obtain the equation,

$$\frac{\hbar\boldsymbol{k}}{m} = \frac{\hbar\boldsymbol{k}}{m^*} + \frac{1}{\Omega}\sum_{\boldsymbol{k}',\sigma'} f_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{k}')\delta(\epsilon_\sigma(\boldsymbol{k}') - \epsilon_F)\frac{\hbar\boldsymbol{k}'}{m^*} \tag{5.67}$$

which then leads to

$$\frac{1}{m} = \frac{1}{m^*} + N(\epsilon_F)\int\frac{d\Omega_{\hat{k}'}}{4\pi}f^s(\hat{k},\hat{k}')\frac{\hat{k}\cdot\hat{k}'}{m^*} \tag{5.68}$$

or by using the orthogonality of the Legendre polynomials,

$$\frac{m^*}{m} = 1 + \frac{1}{3}F_1^s. \tag{5.69}$$

The factor $1/3 = 1/(2l+1)$ for $l = 1$ originates in the term, as

$$\int_{-1}^{+1} dz\; P_l(z)\; P_{l'}(z) = \frac{2\delta_{ll'}}{2l+1} \tag{5.70}$$

Therefore, the relation (5.69) has to couple $m^*$ to $F_1^s$ in order for Landau's theory of Fermi liquids to be self-consistent. Generally, we find that $F_1^s > 0$ so that quasiparticles in a Fermi liquid are effectively heavier than bare electrons.

### 5.2.5 Stability of the Fermi liquid

Upon inspection of the renormalization of the quantities treated previously

$$\frac{\gamma}{\gamma_0} = \frac{m^*}{m}, \tag{5.71}$$

$$\frac{\kappa}{\kappa_0} = \frac{m^*}{m}\frac{1}{1+F_0^s}, \tag{5.72}$$

$$\frac{\chi}{\chi_0} = \frac{m^*}{m}\frac{1}{1+F_0^a} \tag{5.73}$$

with

$$\frac{m^*}{m} = 1 + \frac{1}{3}F_1^s, \quad \gamma_0 = \frac{k_B^2 m k_F}{3\hbar^2}, \quad \kappa_0 = \frac{3m}{n\hbar^2 k_F^2} \text{ and } \chi_0 = \mu_B^2\frac{mk_F}{\pi^2\hbar^2} \tag{5.74}$$

one notes that the compressibility $\kappa$ (susceptibility $\chi$) diverges for $F_0^s \to -1$ ($F_0^a \to -1$), indicating an instability of the system. A diverging spin susceptibility for example leads to a ferromagnetic state with a split Fermi surface, one for each spin direction. On the other hand, a diverging compressibility leads to a spontaneous contraction of the system. More generally, the deformation of the quasiparticle distribution function may vary over the Fermi surface, so that arbitrary deviations of the Fermi liquid ground state may be classified by the deformation

$$\delta n_\sigma(\hat{k}) = \sum_{l=0}^{\infty} \delta n_{\sigma,l} P_l(\cos\theta_{\hat{k}}) \tag{5.75}$$

For pure charge density deformations we have $\delta n_{+l}(\hat{k}) = \delta n_{-l}(\hat{k})$, while pure spin density deformations are described by $\delta n_{+l}(\hat{k}) = -\delta n_{-l}(\hat{k})$. Stability of the Fermi liquid against any of these deformations requires

$$1 + \frac{F_l^{s,a}}{2l+1} > 0. \tag{5.76}$$

If for any deformation channel $l$ this conditions is violated one talks about a "Pomeranchuck instability".[6] Generally, the renormalization of the Fermi liquid leads to a change in the Wilson ratio, defined as

$$\frac{R}{R_0} = \frac{\chi}{\chi_0}\frac{\gamma_0}{\gamma} = \frac{1}{1+F_0^a} \tag{5.77}$$

where $R_0 = \chi_0/\gamma_0 = 6\mu_B^2/\pi^2 k_B^2$. Note that the Wilson ratio does not depend on the effective mass. A remarkable feature of the Fermi liquid theory is that even very strongly interacting Fermions remain Fermi liquids, notably the quantum liquid $^3$He and so-called heavy Fermion systems, which are compounds of transition metals and rare earths. Both are strongly renormalized Fermi liquids. For $^3$He we give some of the parameters in Table 5.1 both for zero pressure and for pressures just below the critical pressure at which He solidifies ($p_c \approx 2.5$MPa = 25bar).

| pressure | $m^*/m$ | $F_0^s$ | $F_0^a$ | $F_1^s$ | $\kappa/\kappa_0$ | $\chi/\chi_0$ |
|---|---|---|---|---|---|---|
| $p = 0$ | 3.0 | 10.1 | -0.52 | 6.0 | 0.27 | 6.3 |
| $p < p_c$ | 6.2 | 94 | -0.74 | 15.7 | 0.065 | 24 |

Table 5.1: List of the parameters of the Fermi liquid theory for $^3$He at zero pressure and at a pressure just below solidification.

The trends show obviously, that the higher the applied pressure is, the denser the liquid becomes and the stronger the quasiparticles interact. Approaching the solidification the compressibility is reduced, the quasiparticles become heavier (slower) and the magnetic response increases drastically. Finally the heavy fermion systems are characterized by the extraordinary enhancements of the effective mass which for many of these compounds lie between 100 and 1000 times higher than the bare electron mass (e.g. $CeAl_3$, $UBe_{13}$, etc.). This large masses lead the notion of almost localized Fermi liquids, since the large effective mass is induced by the hybridization of itinerant conduction electrons with strongly interacting (localized) electron states in partially filled $4f$- or $5f$-orbitals of Lanthanide and Actinide atoms, respectively.

## 5.3 Microscopic considerations

A rigorous derivation of Landau's Fermi liquid theory requires methods of quantum field theory and would go beyond the scope of these lectures. However, plain Rayleigh-Schrödinger theory applied to a simple model allows to gain some insights into the microscopic fundament of this phenomenologically based theory. In the following, we consider a model of fermions with contact

[6] I.J. Pomeranchuk, JETP 8, 361 (1958)

interaction $U\delta(\boldsymbol{r}-\boldsymbol{r}')$, described by the Hamiltonian

$$\mathcal{H} = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \hat{c}^\dagger_{\boldsymbol{k}s} \hat{c}_{\boldsymbol{k}s} + \int d^3r\, d^3r'\, \hat{\Psi}_\uparrow(\boldsymbol{r})^\dagger \hat{\Psi}_\downarrow(\boldsymbol{r}')^\dagger U\delta(\boldsymbol{r}-\boldsymbol{r}') \hat{\Psi}_\downarrow(\boldsymbol{r}') \hat{\Psi}_\uparrow(\boldsymbol{r}) \tag{5.78}$$

$$= \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \hat{c}^\dagger_{\boldsymbol{k}s} \hat{c}_{\boldsymbol{k}s} + \frac{U}{\Omega} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}} \hat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q}\uparrow} \hat{c}^\dagger_{\boldsymbol{k}'-\boldsymbol{q}\downarrow} \hat{c}_{\boldsymbol{k}'\downarrow} \hat{c}_{\boldsymbol{k}\uparrow}. \tag{5.79}$$

where $\epsilon_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ is a parabolic dispersion of non-interacting electrons. We previously noticed that, in order to find well-defined quasiparticles, the interaction between the Fermions has to be short ranged. This specially holds for the contact interaction.

### 5.3.1 Landau parameters

Starting form the Hamiltonian (5.78), we will determine Landau parameters for a corresponding Fermi liquid theory. For a given momentum distribution $n_{\boldsymbol{k}s} = \langle c^\dagger_{\boldsymbol{k}s} c_{\boldsymbol{k}s}\rangle = n^{(0)}_{\boldsymbol{k}s} + \delta n_{\boldsymbol{k}s}$, we can expand the energy resulting form equation (5.79) following the Rayleigh-Schrödinger perturbation method,

$$E = E^{(0)} + E^{(1)} + E^{(2)} + \cdots \tag{5.80}$$

with

$$E^{(0)} = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} n_{\boldsymbol{k}s}, \tag{5.81}$$

$$E^{(1)} = \frac{U}{\Omega} \sum_{\boldsymbol{k},\boldsymbol{k}'} n_{\boldsymbol{k}\uparrow} n_{\boldsymbol{k}'\downarrow}, \tag{5.82}$$

$$E^{(2)} = \frac{U^2}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}} \frac{n_{\boldsymbol{k}\uparrow} n_{\boldsymbol{k}'\downarrow}(1-n_{\boldsymbol{k}+\boldsymbol{q}\uparrow})(1-n_{\boldsymbol{k}'-\boldsymbol{q}\downarrow})}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}. \tag{5.83}$$

The second order term $E^{(2)}$ describes virtual processes corresponding to a pair of particle-hole excitations. The numerator of this term can be split into four different contributions.

We first consider the term quadratic in $n_{\boldsymbol{k}}$ and combine it with the first order term $E^{(1)}$, which has the same structure,

$$\tilde{E}^{(1)} = E^{(1)} + \frac{U^2}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}} \frac{n_{\boldsymbol{k}\uparrow} n_{\boldsymbol{k}'\downarrow}}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}} \approx \frac{\tilde{U}}{\Omega} \sum_{\boldsymbol{k},\boldsymbol{k}'} n_{\boldsymbol{k}\uparrow} n_{\boldsymbol{k}'\downarrow}. \tag{5.84}$$

In the last step, we defined the renormalized interaction $\tilde{U}$ through,

$$\tilde{U} = U + \frac{U^2}{\Omega} \sum_{\boldsymbol{q}} \frac{1}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}. \tag{5.85}$$

In principle, $\tilde{U}$ depends on the wave vectors $\boldsymbol{k}$ and $\boldsymbol{k}'$. However, when the wave vectors are restricted to the Fermi surface ($|\boldsymbol{k}| = |\boldsymbol{k}'| = k_F$), and if the range of the interaction $\ell$ is small compared to the mean electron spacing, i.e., $k_F\ell \ll 1$,[7] this dependency may be neglected.

Since the term quartic in $n_{\boldsymbol{k}}$ vanishes due to symmetry, the remaining contribution to $E^{(2)}$ is cubic in $n_{\boldsymbol{k}}$ and reads

$$\tilde{E}^{(2)} = -\frac{\tilde{U}^2}{\Omega^2} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}} \frac{n_{\boldsymbol{k}\uparrow} n_{\boldsymbol{k}'\downarrow}(n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} + n_{\boldsymbol{k}'-\boldsymbol{q}\downarrow})}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}. \tag{5.90}$$

[7] We should be careful with our choice of a contact interaction, since it would lead to a divergence in the large-$q$ range. A cutoff for $q$ of order $Q_c \sim \ell^{-1}$ would regularize the integral which is dominated by the large-$q$ part.

We replaced $U^2$ by $\tilde{U}^2$, which is admissible at this order of the perturbative expansion. The variation of the energy $E$ in Eq.(5.80) with respect to $\delta n_{\boldsymbol{k}\uparrow}$ can easily be calculated,

$$\tilde{\epsilon}_{\uparrow}(\boldsymbol{k}) = \epsilon_{\boldsymbol{k}} + \frac{\tilde{U}}{\Omega}\sum_{\boldsymbol{k}'} n_{\boldsymbol{k}'\downarrow} - \frac{\tilde{U}^2}{\Omega^2}\sum_{\boldsymbol{k}',\boldsymbol{q}} \frac{n_{\boldsymbol{k}'\downarrow}(n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} + n_{\boldsymbol{k}'-\boldsymbol{q}\downarrow}) - n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} n_{\boldsymbol{k}'-\boldsymbol{q}\downarrow}}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}, \tag{5.91}$$

and an analogous expression is found for $\epsilon_{\downarrow}(\boldsymbol{k})$. The coupling parameters $f_{\sigma\sigma'}(\boldsymbol{k}, \boldsymbol{k}')$ may be determined using the definition (5.26). Starting with $f_{\uparrow\uparrow}(\boldsymbol{k}_F, \boldsymbol{k}'_F)$ with wave-vectors on the Fermi surface ($|\boldsymbol{k}_F| = |\boldsymbol{k}'_F| = k_F$), the terms contributing to the coupling can be written as

$$\frac{\tilde{U}^2}{\Omega^2}\sum_{\boldsymbol{k}',\boldsymbol{q}} n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} \frac{n_{\boldsymbol{k}'-\boldsymbol{q}\downarrow} - n_{\boldsymbol{k}'\downarrow}}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}} \overset{\boldsymbol{k}+\boldsymbol{q}\to\boldsymbol{k}'_F}{\longrightarrow} \frac{1}{\Omega}\sum_{\boldsymbol{k}'_F} n_{\boldsymbol{k}'_F\uparrow} \frac{\tilde{U}^2}{\Omega}\sum_{\boldsymbol{k}'} \left.\frac{n^{(0)}_{\boldsymbol{k}'-\boldsymbol{q}\downarrow} - n^{(0)}_{\boldsymbol{k}'\downarrow}}{\epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}\right|_{\boldsymbol{q}=\boldsymbol{k}'_F-\boldsymbol{k}_F} \tag{5.92}$$

$$= -\frac{1}{\Omega}\sum_{\boldsymbol{k}'_F} n_{\boldsymbol{k}'_F\uparrow} \frac{\tilde{U}^2}{2}\chi_0(\boldsymbol{k}'_F - \boldsymbol{k}_F), \tag{5.93}$$

where we consider $n_{\boldsymbol{k}'_F\uparrow} = n^{(0)}_{\boldsymbol{k}'_F\uparrow} + \delta n_{\boldsymbol{k}'_F\uparrow}$. Note that the part in this term which depends on $n^{(0)}_{\boldsymbol{k}'_F\uparrow}$ will contribute the ground state energy in Landau's energy functional. Here, $\chi_0(\boldsymbol{q})$ is the static Lindhard susceptibility as it was defined in (3.48). With the help of equation (5.26), it follows immediately, that

$$f_{\uparrow\uparrow}(\boldsymbol{k}_F, \boldsymbol{k}'_F) = f_{\downarrow\downarrow}(\boldsymbol{k}_F, \boldsymbol{k}'_F) = \frac{\tilde{U}^2}{2}\chi_0(\boldsymbol{k}_F - \boldsymbol{k}'_F). \tag{5.94}$$

The other couplings are obtained in a similar way, resulting in

$$f_{\uparrow\downarrow}(\boldsymbol{k}_F, \boldsymbol{k}'_F) = f_{\downarrow\uparrow}(\boldsymbol{k}_F, \boldsymbol{k}'_F) = \tilde{U} - \frac{\tilde{U}^2}{2}\left[2\tilde{\chi}_0(\boldsymbol{k}_F + \boldsymbol{k}'_F) - \chi_0(\boldsymbol{k}_F - \boldsymbol{k}'_F)\right], \tag{5.95}$$

where the function $\tilde{\chi}_0(\boldsymbol{q})$ is defined as

$$\tilde{\chi}_0(\boldsymbol{q}) = \frac{1}{\Omega}\sum_{\boldsymbol{k}'} \frac{n^{(0)}_{\boldsymbol{k}'+\boldsymbol{q}\uparrow} + n^{(0)}_{\boldsymbol{k}'\downarrow}}{2\epsilon_F - \epsilon_{\boldsymbol{k}'+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'}} \tag{5.96}$$

If the couplings are be parametrized by the angle $\theta$ between $\boldsymbol{k}_F$ and $\boldsymbol{k}'_F$, they can be expressed as

$$f_{\sigma\sigma'}(\theta) = \frac{\tilde{U}}{2}\left[\left(1 + \frac{\tilde{U}N(\epsilon_F)}{2}\left(2 + \frac{\cos\theta}{2\sin(\theta/2)}\ \ln\frac{1+\sin(\theta/2)}{1-\sin(\theta/2)}\right)\right)\delta_{\sigma\sigma'}\right. \tag{5.97}$$

$$\left. - \left(1 + \frac{\tilde{U}N(\epsilon_F)}{2}\left(1 - \frac{\sin(\theta/2)}{2}\ \ln\frac{1+\sin(\theta/2)}{1-\sin(\theta/2)}\right)\right)\sigma\sigma'\right]. \tag{5.98}$$

---

Thus we may use the following expansion,

$$\frac{1}{\Omega}\sum_{\boldsymbol{q}} \frac{1}{\epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}'-\boldsymbol{q}}} = \frac{1}{(2\pi)^3}\int_o^{Q_c} dq\, q^2 \int d\Omega_q \frac{m}{(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{q} - \boldsymbol{q}^2} \tag{5.86}$$

$$= \frac{m}{(2\pi)^2}\int dq\, q \int_{-1}^{+1} \frac{d\cos\theta}{K\cos\theta - q} = \frac{m}{(2\pi)^2}\int_0^{Q_c} dq\, q \ln\left|\frac{q-K}{q+K}\right| \tag{5.87}$$

$$= -\frac{m}{(2\pi)^2}\left(Q_c + \frac{K^2 - Q_c^2}{2K}\ln\left|\frac{Q_c - K}{Q_c + K}\right|\right) \tag{5.88}$$

$$\approx -\frac{2mQ_c}{(2\pi)^2}\left(1 - \frac{K^2}{Q_c^2} + O\left(\frac{K^4}{Q_c^4}\right)\right), \tag{5.89}$$

where we use $K = |\boldsymbol{k}' - \boldsymbol{k}| \leq 2k_F \ll Q_c$. From this we conclude that the momentum dependence of $\tilde{U}$ is indeed weak.

Finally, we are in the position to determine the most important Landau parameters by matching the expressions (5.94) and (5.95) to the parametrization (5.97),

$$F_0^s = \tilde{u}\Big[1+\tilde{u}\big[1+\frac{1}{6}(2+\ln(2))\big]\Big] = \tilde{u} + 1.449\,\tilde{u}^2, \tag{5.99}$$

$$F_0^a = -\tilde{u}\Big[1+\tilde{u}\big[1-\frac{2}{3}(1-\ln(2))\big]\Big] = -\tilde{u} - 0.895\,\tilde{u}^2, \tag{5.100}$$

$$F_1^s = \tilde{u}^2\frac{2}{15}(7\ln(2)-1) \approx 0.514\,\tilde{u}^2, \tag{5.101}$$

where $\tilde{u} = \tilde{U}N(\epsilon_F)$ has been introduced for better readability. Since the Landau parameter $F_1^s$ is responsible for the modification of the effective mass $m^*$ compared to the bare mass $m$, $m^*$ is enhanced compared to $m$ for both attractive ($U < 0$) and repulsive ($U > 0$) interactions. Obviously, the sign of the interaction $U$ does not affect the renormalization of the effective mass $m^*$. This is so, because the existence of an interaction (whatever sign it has) between the particles enforces the motion of many particles whenever one is moved. The behavior of the susceptibility and the compressibility depends on the sign of the interaction. If the interaction is repulsive ($\tilde{u} > 0$), the compressibility decreases ($F_0^s > 0$), implying that it is harder to compress the Fermi liquid. The susceptibility is enhanced ($F_0^a < 0$) in this case, so that it is easier to polarize the spins of the electrons. Conversely, for attractive interactions ($\tilde{u} < 0$), the compressibility is enhanced due to a negative Landau parameter $F_0^s$, whereas the susceptibility is suppressed with a factor $1/(1+F_0^a)$, with $F_0^a > 0$. The attractive case is more subtle because the Fermi liquid becomes unstable at low temperatures, turning into a superfluid or superconductor, by forming so-called Cooper pairs. This represents another non-trivial Fermi surface instability.

### 5.3.2 Distribution function

Finally, we examine the effect of interactions on the ground state properties, using again Rayleigh-Schrödinger perturbation theory. The calculation of the corrections to the ground state $|\Psi_0\rangle$, the filled Fermi sea can be expressed as

$$|\Psi\rangle = |\Psi^{(0)}\rangle + |\Psi^{(1)}\rangle + \cdots \tag{5.102}$$

where

$$|\Psi^{(0)}\rangle = |\Psi_0\rangle \tag{5.103}$$

$$|\Psi^{(1)}\rangle = \frac{U}{\Omega}\sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}}\sum_{s,s'}\frac{\widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q},s}\widehat{c}^\dagger_{\boldsymbol{k}'-\boldsymbol{q},s'}\widehat{c}_{\boldsymbol{k}',s'}\widehat{c}_{\boldsymbol{k},s}}{\epsilon_{\boldsymbol{k}}+\epsilon_{\boldsymbol{k}'}-\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}'-\boldsymbol{q}}}|\Psi_0\rangle. \tag{5.104}$$

The state $|\Psi_0\rangle$ represents the ground state of non-interacting fermions. The lowest order correction involves particle-hole excitations, depleting the Fermi sea by lifting particles virtually above the Fermi energy. How the correction (5.104) affects the distribution function, will be discussed next. The momentum distribution $n_{\boldsymbol{k}s} = \langle\widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s}\rangle$ is obtain as the expectation value,

$$n_{\boldsymbol{k}s} = \frac{\langle\Psi|\widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s}|\Psi\rangle}{\langle\Psi|\Psi\rangle} = n^{(0)}_{\boldsymbol{k}s} + \delta n^{(2)}_{\boldsymbol{k}s} + \cdots \tag{5.105}$$

where $n^{(0)}_{\boldsymbol{k}s}$ is the unperturbed distribution $\Theta(k_F - |\boldsymbol{k}|)$, and

$$\delta n^{(2)}_{\boldsymbol{k}s} = \begin{cases} -\dfrac{U^2}{\Omega^2}\displaystyle\sum_{\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3}\frac{(1-n_{\boldsymbol{k}_1})(1-n_{\boldsymbol{k}_2})n_{\boldsymbol{k}_3}}{(\epsilon_{\boldsymbol{k}}+\epsilon_{\boldsymbol{k}_3}-\epsilon_{\boldsymbol{k}_1}-\epsilon_{\boldsymbol{k}_2})^2}\delta_{\boldsymbol{k}+\boldsymbol{k}_3,\boldsymbol{k}_1+\boldsymbol{k}_2} & |\boldsymbol{k}| < k_F \\[2ex] \dfrac{U^2}{\Omega^2}\displaystyle\sum_{\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3}\frac{n_{\boldsymbol{k}_1}n_{\boldsymbol{k}_2}(1-n_{\boldsymbol{k}_3})}{(\epsilon_{\boldsymbol{k}_1}+\epsilon_{\boldsymbol{k}_2}-\epsilon_{\boldsymbol{k}}-\epsilon_{\boldsymbol{k}_3})^2}\delta_{\boldsymbol{k}+\boldsymbol{k}_3,\boldsymbol{k}_1+\boldsymbol{k}_2} & |\boldsymbol{k}| > k_F \end{cases}. \tag{5.106}$$

This yields the modification of the distribution functions as shown in Figure 5.6. It allows us also to determine the size of the discontinuity of the distribution function at the Fermi surface,

$$n_{\boldsymbol{k}_F^-} - n_{\boldsymbol{k}_F^+} = 1 - \left(\frac{UN(\epsilon_F)}{2}\right)^2 \ln(2), \tag{5.107}$$

where

$$n_{\boldsymbol{k}_F^\pm} = \lim_{|\boldsymbol{k}|-k_F\to 0_\pm} \left(n_{\boldsymbol{k}}^{(0)} + \delta n_{\boldsymbol{k}}^{(2)}\right). \tag{5.108}$$

The jump of $n_{\boldsymbol{k}}$ at the Fermi surface is reduced independently of the sign of the interaction. The reduction is quadratic in the perturbation parameter $UN(\epsilon_F)$. This jump is also a measure for the weight of the quasiparticle state at the Fermi surface.

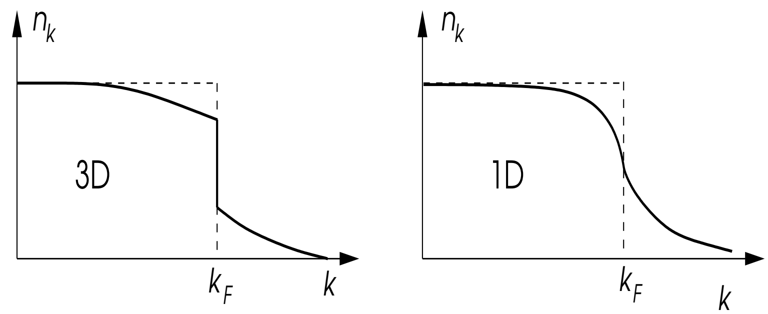



Figure 5.6: Momentum distribution functions of electrons for a three-dimensional (left panel) and one-dimensional (right panel) Fermion system.

### 5.3.3 Fermi liquid in one dimension?

Within a perturbative approach the Fermi liquid theory can be justified for a three-dimensional system and we recognize the one-to-one correspondence between bare electrons and quasiparticles renormalized by (short-ranged) interactions. Now we would like to show that within the same approach problems appear in one-dimensional systems, which are conceptional nature and hint that interaction Fermions in one dimension would not form a Fermi liquid, but a Luttinger liquid, as we will motivate briefly below.

The Landau parameters have been expressed above in terms of the response functions $\chi_0(\boldsymbol{q} = \boldsymbol{k}_F - \boldsymbol{k}'_F)$ and $\tilde{\chi}_0(\boldsymbol{q} = \boldsymbol{k}_F - \boldsymbol{k}'_F)$. For the one-dimensional system, as given in Eqs.(5.94 - 5.96), the relevant contributions come from two configurations, since there are two Fermi points only (instead of a two-dimensional Fermi surface),

$$(k_F, k'_F) \quad \Rightarrow \quad q = k_F - k'_F = 0, \pm 2k_F \, . \tag{5.109}$$

We find that the response functions show singularities for some of these momenta,[8] and we obtain

$$f_{\uparrow\uparrow}(\pm k_F, \pm k_F) = f_{\downarrow\downarrow}(\pm k_F, \mp k_F) \to \infty \tag{5.112}$$

[8] While $\chi_0(q)$ is the Lindhard function given in Eq.(3.96) which diverges logarithmically at $q = \pm 2k_F$, we obtain for the other response function

$$\tilde{\chi}_0(q) = \begin{cases} -\dfrac{2m}{\hbar^2\sqrt{4k_F^2 - q^2}} \ln\left(\dfrac{\sqrt{2k_F+q} - \sqrt{2k_F-q}}{\sqrt{2k_F+q} + \sqrt{2k_F-q}}\right) & |q| < 2k_F \\[2ex] \dfrac{2m}{\hbar^2\sqrt{q^2 - 4k_F^2}} \left(\arctan\sqrt{\dfrac{q+2k_F}{q-2k_F}} + \arctan\sqrt{\dfrac{q-2k_F}{q+2k_F}}\right) & |q| > 2k_F \end{cases} \tag{5.110}$$

as well as

$$f_{\uparrow\downarrow}(k_F, \pm k_F) = f_{\downarrow\uparrow}(k_F, \mp k_F) \to \infty \tag{5.113}$$

giving rise to the divergence of all Landau parameters. Therefore the perturbative approach to a Landau Fermi liquid is not allowed for the one-dimensional Fermi system.
The same message is obtained when looking at the momentum distribution form which had in three dimensions a step giving a measure for the (reduced but finite) quasiparticle weight. The analogous calculation as in Sect.5.3.2 leads here to

$$n^{(2)}_{\boldsymbol{k}s} \approx \begin{cases} \dfrac{1}{8\pi^2}\dfrac{U^2}{\hbar^2 v_F^2}\ \ln \dfrac{k_+}{k - k_F} & k > k_F \\[2ex] -\dfrac{1}{8\pi^2}\dfrac{U^2}{\hbar^2 v_F^2}\ \ln \dfrac{k_-}{k_F - k} & k < k_F \end{cases} . \tag{5.114}$$

Here, $k_\pm$ are cutoff parameters of the order of the Fermi wave vector $k_F$. Apparently the quality of the perturbative calculation deteriorates as $k \to k_{F\pm}$, since we encounter a logarithmic divergence from both sides.

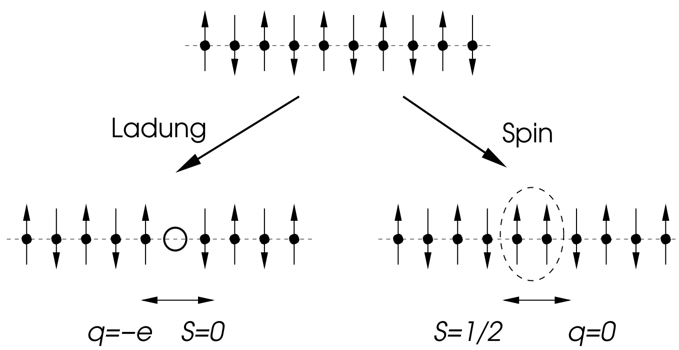



Figure 5.7: Visualization of spin-charge separation. The dominant anti-ferromagnetic spin correlation is staggered. A charge excitation is a vacancy which can move, while spin excitation may be considered as domain wall. Both excitations move independently.

Indeed, a more elaborated approach shows that the distribution function is continuous at $k = k_F$ in one dimension, without any jump. Correspondingly, the quasiparticle weight vanishes and the elementary excitations cannot be described by Fermionic quasiparticles but rather by collective modes. Landau's Theory of Fermi liquids is inappropriate for such systems. This kind of behavior, where the quasiparticle weight vanishes, can be described by the so-called bosonization of fermions in one dimension, a topic that is beyond the scope of these lectures. However, a result worth mentioning, shows that the fermionic excitations in one dimensions decay into independent charge and spin excitations, the so-called *spin-charge separation*. This behavior can be understood with the naive picture of a half-filled lattice with predominantly antiferromagnetic spin correlations. In this case both charge excitations (empty or doubly occupied lattice site) and spin excitations (two parallel neighboring spins) represent different kinds of domain walls, and are free to move at different velocities.

---

which diverges as

$$\lim_{q\to 0_\pm} \tilde{\chi}_0(q) = -\frac{m}{\hbar^2 k_F}\ln\left(\frac{q}{2k_F}\right) , \qquad \lim_{q\to 2k_{F-}} \tilde{\chi}_0(q) = \frac{m}{\hbar^2 k_F} \quad \text{and} \quad \lim_{q\to 2k_{F+}} \tilde{\chi}_0(q) = \frac{\pi m}{2\hbar^2\sqrt{k_F}}\frac{1}{\sqrt{q-2k_F}} . \tag{5.111}$$

# Chapter 6

# Transport properties of metals

The ability to transport electrical current is one of the most remarkable and characteristic properties of metals. At zero temperature, a ideal pure metal is a perfect electrical conductor, i.e., the resistivity is zero. However, disorder due to impurities and lattice defects influence the transport and yield a finite *residual resistivity*, as found in real materials. At finite temperature, electron-electron and electron-phonon scattering lead to a temperature-dependent resistivity. Furthermore, an external magnetic field may influence the resistivity, a phenomenon called magnetoresistance, which leads to the previously studied Hall effect. In this chapter, the effects of a magnetic field will not be considered. Finally, heat transport, which is also mostly mediated by electrons in metals, is going hand in hand with the electric transport. In this context, transport phenomena such as thermoelectricity (Seebeck and Peltier effect) will be analyzed here.

## 6.1 Electrical conductivity

In a normal metal, an electrical current density $\boldsymbol{j}(\boldsymbol{q},\omega)$ (in $\boldsymbol{q},\omega$-space) is induced by an applied electrical field $\boldsymbol{E}(\boldsymbol{q},\omega)$. For a homogeneous isotropic metal, we define the scalar[1] electrical conductivity $\sigma(\boldsymbol{q},\omega)$ within linear response, through

$$\boldsymbol{j}(\boldsymbol{q},\omega) = \sigma(\boldsymbol{q},\omega)\boldsymbol{E}(\boldsymbol{q},\omega). \tag{6.1}$$

The current density $\boldsymbol{j}(\boldsymbol{q},\omega)$ is related to the charge density $\rho(\boldsymbol{r},t) = -en(\boldsymbol{r},t)$, via the continuity equation

$$\frac{\partial}{\partial t}\rho(\boldsymbol{r},t) + \boldsymbol{\nabla}\cdot\boldsymbol{j}(\boldsymbol{r},t) = 0, \tag{6.2}$$

or, in Fourier transformed,

$$\omega\rho(\boldsymbol{q},\omega) - \boldsymbol{q}\cdot\boldsymbol{j}(\boldsymbol{q},\omega) = 0. \tag{6.3}$$

It is interesting to see that a relation between the conductivity $\sigma(\boldsymbol{q},\omega)$ and the dynamical dielectric susceptibility $\chi_0(\boldsymbol{q},\omega)$ defined in equation (3.53) of chapter 3 arises from the equations (6.1) and (6.3). For this, we can calculate

$$\begin{aligned}\chi_0(\boldsymbol{q},\omega) &= -\frac{\rho(\boldsymbol{q},\omega)}{eV(\boldsymbol{q},\omega)} = -\frac{\boldsymbol{q}\cdot\boldsymbol{j}(\boldsymbol{q},\omega)}{e\omega V(\boldsymbol{q},\omega)}\\ &= -\frac{\sigma(\boldsymbol{q},\omega)}{\omega e}\frac{\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},\omega)}{V(\boldsymbol{q},\omega)} = -\frac{\sigma(\boldsymbol{q},\omega)}{\omega}\frac{[i\boldsymbol{q}^2V(\boldsymbol{q},\omega)]}{e^2V(\boldsymbol{q},\omega)}.\end{aligned} \tag{6.4}$$

[1] In an anisotropic material, the conductivity $\hat{\sigma}$ would be a full $3\times 3$ tensor.

In the first line, we used the definition (3.53) of $\chi_0(\boldsymbol{q},\omega)$ and the continuity equation (6.3). To the second line, we made use of the definition (6.1) of $\sigma(\boldsymbol{q},\omega)$ and then replaced $\boldsymbol{E}(\boldsymbol{q},\omega)$ by $-i\boldsymbol{q}V(\boldsymbol{q},\omega)$, which is nothing else than the Fourier transform of the equation

$$-e\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{\nabla}_{\boldsymbol{r}} V(\boldsymbol{r},t). \tag{6.5}$$

From this calculations we conclude that

$$\chi_0(\boldsymbol{q},\omega) = \frac{-iq^2}{e^2\omega}\sigma(\boldsymbol{q},\omega), \tag{6.6}$$

and thus

$$\varepsilon(\boldsymbol{q},\omega) = 1 - \frac{4\pi e^2}{q^2}\chi_0(\boldsymbol{q},\omega) = 1 + \frac{4\pi i}{\omega}\sigma(\boldsymbol{q},\omega). \tag{6.7}$$

In the limit of large wavelengths $q \ll k_F$, we know from previous discussions[2] that $\varepsilon(0,\omega) = 1 - \omega_p^2/\omega^2$. Then the conductivity simplifies to

$$\sigma(\omega) = \frac{i\omega_p^2}{4\pi\omega}. \tag{6.8}$$

One might conclude from this result that the conductivity is purely imaginary in the small-$q$ limit. However, this conclusion is wrong, since the real part of $\sigma(\omega)$ is related to its imaginary part via the Kramers-Kronig relation. Defining $\sigma_1$ ($\sigma_2$) as the real (imaginary) part of $\sigma$, this relation states that

$$\sigma_1(\omega) = -\frac{1}{\pi}\,\mathcal{P}\left[\int_{-\infty}^{+\infty} d\omega'\,\frac{1}{\omega-\omega'}\sigma_2(\omega')\right] \tag{6.9}$$

and

$$\sigma_2(\omega) = \frac{1}{\pi}\,\mathcal{P}\left[\int_{-\infty}^{+\infty} d\omega'\,\frac{1}{\omega-\omega'}\sigma_1(\omega')\right]. \tag{6.10}$$

A simple calculation with $\sigma_2$ from equation (6.8), yields

$$\sigma_1(\omega) = \frac{\omega_p^2}{4}\delta(\omega), \tag{6.11}$$

$$\sigma_2(\omega) = \frac{\omega_p^2}{4\pi\omega}. \tag{6.12}$$

Obviously this metal is perfectly conducting ($\sigma \to \infty$ for $\omega \to 0$), which comes from the fact that we considered systems without dissipation so far.
An additional important property coming from complex analysis, is the existence of the so-called *$f$-sum rule*,

$$\int_0^{\infty} d\omega'\,\sigma_1(\omega') = \frac{1}{2}\int_{-\infty}^{+\infty} d\omega'\,\sigma_1(\omega') = \frac{\omega_p^2}{8} = \frac{\pi e^2 n}{2m}. \tag{6.13}$$

This relation is valid for all electronic systems (including semiconductors).

[2] In the small-$\boldsymbol{q}$ limit we approximate $\chi_0(\boldsymbol{q},\omega) \approx nq^2/m\omega^2$ from equation (3.67).

## 6.2 Transport equations and relaxation time

We introduce here Boltzmann's transport theory as as rather simple and efficient way to deal with dissipation and relaxation of non-stationary electronic states in metals.

### 6.2.1 The Boltzmann equation

In order to tackle the problem of a finite conductivity, it is suitable to use a formalism similar to Landau's Fermi liquid theory, based on a distribution function of quasiparticles. In transport theory, the distribution function can be used to describe the deviation of the system from an equilibrium. If the system is isolated from external influence, equilibrium is reached through relaxation after some time, a process which is accompanied with an increase of entropy as discussed in statistical physics. Analogously to the theory of transport phenomena, let us introduce the distribution function[3] $f(\boldsymbol{k}, \boldsymbol{r}, t)$, where

$$f(\boldsymbol{k}, \boldsymbol{r}, t) \frac{d^3 k}{(2\pi)^3} d^3 r \tag{6.14}$$

is the number of particles in the infinitesimal phase space volume $d^3 r d^3 k/(2\pi)^3$ centered at $(\boldsymbol{k}, \boldsymbol{r})$, at time $t$. Such a description is only applicable if the temporal and spacial variations occur at long wavelengths and small frequencies, respectively, i.e., if typically $q \ll k_F$ and $\hbar\omega \ll \epsilon_F$. The total number of particles $N$ is given by

$$N = \int \frac{d^3 k}{(2\pi)^3} d^3 r f(\boldsymbol{k}, \boldsymbol{r}, t). \tag{6.15}$$

The equilibrium distribution $f_0$ for the fermionic quasiparticles is given by the Fermi-Dirac distribution,

$$f_0(\boldsymbol{k}, \boldsymbol{r}, t) = \frac{1}{e^{(\epsilon_{\boldsymbol{k}} - \mu)/k_B T} + 1}, \tag{6.16}$$

and is independent of space $\boldsymbol{r}$ and time $t$. The general distribution function $f(\boldsymbol{k}, \boldsymbol{r}, t)$ obeys the *Boltzmann equation*

$$\frac{D}{Dt} f(\boldsymbol{k}, \boldsymbol{r}, t) = \left( \frac{\partial}{\partial t} + \dot{\boldsymbol{r}} \cdot \boldsymbol{\nabla}_{\boldsymbol{r}} + \dot{\boldsymbol{k}} \cdot \boldsymbol{\nabla}_{\boldsymbol{k}} \right) f(\boldsymbol{k}, \boldsymbol{r}, t) = \left( \frac{\partial f}{\partial t} \right)_{\text{coll}}, \tag{6.17}$$

where the substantial derivative in phase space $D/Dt$ is defined as the total temporal derivative in a frame moving with the phase-space volume. The right-hand side is called *collision integral* and describes the rate of change in $f$ due to collision processes. Without scattering, the equation (6.17) would represent a continuity equation for $f$. Now, consider the temporal derivatives of $\boldsymbol{r}$ and $\boldsymbol{k}$ from a quasi-classical viewpoint. In absence of a magnetic field, we find

$$\dot{\boldsymbol{r}} = \frac{\hbar \boldsymbol{k}}{m}, \tag{6.18}$$

$$\hbar \dot{\boldsymbol{k}} = -e\boldsymbol{E}, \tag{6.19}$$

i.e., the force $\hbar\dot{\boldsymbol{k}}$, which is our central interest, originates from the electric field. The collision integral may be expressed via the probability $W(\boldsymbol{k}, \boldsymbol{k}')$ to scatter a quasiparticle with wave vector $\boldsymbol{k}$ to $\boldsymbol{k}'$. For simple scattering on static potentials, the collision integral is given by

$$\left( \frac{\partial f}{\partial t} \right)_{\text{coll}} = -\int \frac{d^3 k'}{(2\pi)^3} \big[ W(\boldsymbol{k}, \boldsymbol{k}') f(\boldsymbol{k}, \boldsymbol{r}, t)(1 - f(\boldsymbol{k}', \boldsymbol{r}, t)) \tag{6.20}$$

$$- W(\boldsymbol{k}', \boldsymbol{k}) f(\boldsymbol{k}', \boldsymbol{r}, t)(1 - f(\boldsymbol{k}, \boldsymbol{r}, t)) \big]. \tag{6.21}$$

[3] For simplicity we neglect spin the electron spin. In general there would be a distribution function $f_\sigma(\boldsymbol{k}, \boldsymbol{r}, t)$ for each spin species $\sigma$.

The first term, describing the scattering[4] from $\boldsymbol{k}$ to $\boldsymbol{k}'$, requires a quasiparticle at $\boldsymbol{k}$, hence the factor $f(\boldsymbol{k}, \boldsymbol{r}, t)$, and the absence of a particle at $\boldsymbol{k}'$, therefore the factor $1 - f(\boldsymbol{k}', \boldsymbol{r}, t)$. This process describes the scattering out of the phase space volume $d^3k/(2\pi)^3$, i.e., reduces the number of particles in it. Therefore, it enters the collision integral with negative sign. The second term describes the opposite process and, according to its positive sign, increases the number of particles in the phase space volume $d^3k/(2\pi)^3$. For a system with time inversion symmetry, we have $W(\boldsymbol{k}, \boldsymbol{k}') = W(\boldsymbol{k}', \boldsymbol{k})$. Assuming this, we can combine both terms and end up with

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int \frac{d^3k'}{(2\pi)^3} W(\boldsymbol{k}, \boldsymbol{k}') \big[f(\boldsymbol{k}', \boldsymbol{r}, t) - f(\boldsymbol{k}, \boldsymbol{r}, t)\big]. \tag{6.22}$$

The Boltzmann equation is a complicated integro-differential equation and suitable approximations are required. Usually, we study processes close to equilibrium, where the deviation $f(\boldsymbol{k}, \boldsymbol{r}, t) - f_0(\boldsymbol{k}, \boldsymbol{r}, t)$ is small compared to $f(\boldsymbol{k}, \boldsymbol{r}, t)$. Here, to generalize we assume $f_0(\boldsymbol{k}, \boldsymbol{r}, t)$ to be a *local* equilibrium distribution for which the temperature $T = T(\boldsymbol{r}, t)$ and the chemical potential $\mu = \mu(\boldsymbol{r}, t)$ vary slowly in $\boldsymbol{r}$ and $t$, such that $f_0(\boldsymbol{k}, \boldsymbol{r}, t)$ can still be expressed via the local Fermi-Dirac distribution (6.16). At small deviations from equilibrium (or local equilibrium), we can approximate the collision integral by the so-called *relaxation-time approximation*. For simplicity, we assume that the system is isotropic, such that the quasiparticle dispersion $\epsilon_{\boldsymbol{k}}$ only depends on $|\boldsymbol{k}|$ and, furthermore, that the scattering probabilities are elastic and depend on the angle between $\boldsymbol{k}$ and $\boldsymbol{k}'$. Then, we make the Ansatz

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f(\boldsymbol{k}, \boldsymbol{r}, t) - f_0(\boldsymbol{k}, \boldsymbol{r}, t)}{\tau(\epsilon_{\boldsymbol{k}})}. \tag{6.23}$$

The time scale $\tau(\epsilon_{\boldsymbol{k}})$ is called *relaxation time* and gives the characteristic time within which the system relaxes to equilibrium.

Consider the simplest case of a system at constant temperature subject to a small uniform electric field $\boldsymbol{E}(t)$. With $f(\boldsymbol{k}, \boldsymbol{r}, t) = f_0(\boldsymbol{k}, \boldsymbol{r}, t) + \delta f(\boldsymbol{k}, \boldsymbol{r}, t)$, we can calculate the Fourier-transform of Boltzmann equation (6.17) in relaxation-time approximation and find, after linearizing in $\delta f$,

$$-i\omega \delta f(\boldsymbol{k}, \omega) - \frac{e\boldsymbol{E}(\omega)}{\hbar} \frac{\partial f_0(\boldsymbol{k})}{\partial \boldsymbol{k}} = -\frac{\delta f(\boldsymbol{k}, \omega)}{\tau(\epsilon_{\boldsymbol{k}})}. \tag{6.24}$$

In order to come to this expression, we used that $f(\boldsymbol{k}, \boldsymbol{r}, t) = f(\boldsymbol{k}, t)$ for $E = E(t)$, and assumed for linearizing equation (6.24) that $\delta f \propto |\boldsymbol{E}|$. Thus, the equation (6.24) is consistent to linear order in $|\boldsymbol{E}|$ and can be easily solved as

$$\delta f(\boldsymbol{k}, \omega) = \frac{e\tau \boldsymbol{E}(\omega)}{\hbar(1 - i\omega\tau)} \frac{\partial f_0(\boldsymbol{k})}{\partial \boldsymbol{k}} = \frac{e\tau \boldsymbol{E}(\omega)}{\hbar(1 - i\omega\tau)} \frac{\partial f_0(\epsilon)}{\partial \epsilon} \frac{\partial \epsilon_{\boldsymbol{k}}}{\partial \boldsymbol{k}}. \tag{6.25}$$

This result leads straightforwardly to the quasiparticle current $\boldsymbol{j}(\omega)$,

$$\boldsymbol{j}(\omega) = -2e \int \frac{d^3k}{(2\pi)^3} \boldsymbol{v}_{\boldsymbol{k}} f(\boldsymbol{k}, \omega) = -\frac{e^2}{4\pi^3} \int d^3k \frac{\tau(\epsilon_{\boldsymbol{k}})[\boldsymbol{E}(\omega) \cdot \boldsymbol{v}]\boldsymbol{v}}{1 - i\omega\tau(\epsilon_{\boldsymbol{k}})} \frac{\partial f_0(\epsilon_{\boldsymbol{k}})}{\partial \epsilon_{\boldsymbol{k}}}, \tag{6.26}$$

which in turn can be simplified to

$$j_\alpha(\omega) = \sum_\beta \sigma_{\alpha\beta}(\omega) E_\beta(\omega), \tag{6.27}$$

[4]Note that if spin was considered here, this collision term would account for scattering processes where spin is conserved. However, there are in principle also scattering process where the electron spin can be transferred to the lattice (spin-orbit coupling) or an impurity (Kondo effect) and would not be conserved independently.

where the conductivity tensor $\sigma_{\alpha\beta}$ reads

$$\sigma_{\alpha\beta} = -\frac{e^2}{4\pi^3}\int d\epsilon \frac{\partial f_0(\epsilon)}{\partial \epsilon}\frac{\tau(\epsilon)}{1-i\omega\tau(\epsilon)}\int d\Omega_{\boldsymbol{k}} k^2 \frac{v_{\alpha\boldsymbol{k}} v_{\beta\boldsymbol{k}}}{\hbar|\boldsymbol{v}_{\boldsymbol{k}}|}. \tag{6.28}$$

This corresponds to the Ohmic law. Note that $\sigma_{\alpha\beta} = \sigma\delta_{\alpha\beta}$ in isotropic systems. We recover in this case the expression (6.1) for the conductivity, which we introduced at the beginning of this chapter. In the following, we consider the result (6.28) for an isotropic system in different limiting cases.

### 6.2.2 The Drude form

For $\omega\tau \gg 1$ equation (6.28) becomes independent on the relaxation time. In an isotropic system $\sigma_{\alpha\beta} = \sigma\delta_{\alpha\beta}$ at low temperatures $T \ll T_F$, this leads to

$$\sigma(\omega) \approx i\frac{e^2 m^2 v_F}{4\pi^3\hbar^3\omega}\int d\Omega_{\boldsymbol{k}} v_{Fz}^2 = i\frac{e^2 n}{m\omega} = i\frac{\omega_p^2}{4\pi\omega}, \tag{6.29}$$

which reproduces the result from equation (6.8). However now, this does not mean that our system is a perfect conductor. We are actually interested in the static limit, where the "dc conductivity" ($\omega = 0$) reduces to

$$\sigma = -\frac{e^2 n}{m}\int d\epsilon \frac{\partial f_0}{\partial \epsilon}\tau(\epsilon) = \frac{e^2 n\bar{\tau}}{m} = \frac{\omega_p^2\bar{\tau}}{4\pi}. \tag{6.30}$$

Since the function $\partial f_0/\partial\epsilon$ is strongly peaked around the Fermi energy $\epsilon_F$, we introduced a mean relaxation time $\bar{\tau}$. In the form (6.30), the result recovers the well-known Drude[5] form of the conductivity.

If the relaxation time $\tau$ depends only weakly on energy, we can simply calculate the optical conductivity at finite frequency,

$$\sigma(\omega) = \frac{\omega_p^2}{4\pi}\frac{\bar{\tau}}{1-i\omega\bar{\tau}} = \frac{\omega_p^2}{4\pi}\left(\frac{\bar{\tau}}{1+\omega^2\bar{\tau}^2} + \frac{i\bar{\tau}^2\omega}{1+\omega^2\bar{\tau}^2}\right) = \sigma_1 + i\sigma_2. \tag{6.31}$$

Note that the real part satisfies the $f$-sum rule,

$$\int_0^\infty d\omega\, \sigma_1(\omega) = \int_0^\infty d\omega\, \frac{\omega_p^2}{4\pi}\frac{\bar{\tau}}{1+\omega^2\bar{\tau}^2} = \frac{\omega_p^2}{8} \tag{6.32}$$

and that $\sigma(\omega)$ recovers the behavior of equation (6.12) in the limit $\bar{\tau} \to 0$. This form of the conductivity yields the dielectric function

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2\bar{\tau}}{\omega(i+\omega\bar{\tau})} = 1 - \frac{\omega_p^2\bar{\tau}^2}{1+\omega^2\bar{\tau}^2} + \frac{i}{\omega}\frac{\omega_p^2\bar{\tau}}{1+\omega^2\bar{\tau}^2}, \tag{6.33}$$

which can be used to discuss the optical properties of metals. The complex index of refraction $n + ik$ is given through $(n+ik)^2 = \varepsilon$. Next, we discuss three important regimes of frequency.

[5]Note, that the phenomenological Drude theory of electron transport can be deduced from purely classical considerations.

**Relaxation-free regime** $(\omega\bar{\tau} \ll 1 \ll \omega_p\bar{\tau})$

In this limit, the real ($\varepsilon_1$) and imaginary ($\varepsilon_2$) part of the dielectric function (6.33) read

$$\varepsilon_1(\omega) \approx -\omega_p^2\bar{\tau}^2, \tag{6.34}$$
$$\varepsilon_2(\omega) \approx \frac{\omega_p^2\bar{\tau}}{\omega}. \tag{6.35}$$

The real part $\varepsilon_1$ is constant and negative, whereas the imaginary part $\varepsilon_2$ becomes singular in the limit $\omega \to 0$. Thus, the refractive index turns out to be dominated by $\varepsilon_2$

$$n(\omega) \approx k(\omega) \approx \sqrt{\frac{\varepsilon_2(\omega)}{2}} \approx \sqrt{\frac{\omega_p^2\bar{\tau}}{2\omega}}, \tag{6.36}$$

As a result, the reflectivity $R$ is practically 100%, since

$$R = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}, \tag{6.37}$$

and $n(\omega) \gg 1$. The absorption index $k(\omega)$ determines the penetration depth $\delta$ through

$$\delta(\omega) = \frac{c}{\omega k(\omega)} \approx \frac{c}{\omega_p}\sqrt{\frac{2}{\omega\bar{\tau}}}. \tag{6.38}$$

With this, the skin depth of a metal with the famous relation $\delta(\omega) \propto \omega^{-1/2}$ is reproduced within the relaxation time approximation of the Boltzmann equation. While length $\delta(\omega)$ is in the centimeter range for frequencies of the order of $10 - 100$Hz, the Debye length $c/\omega_p$, is only of the order of 100Å for $\hbar\omega_p = 10$ eV. (cf. Figure 6.1).

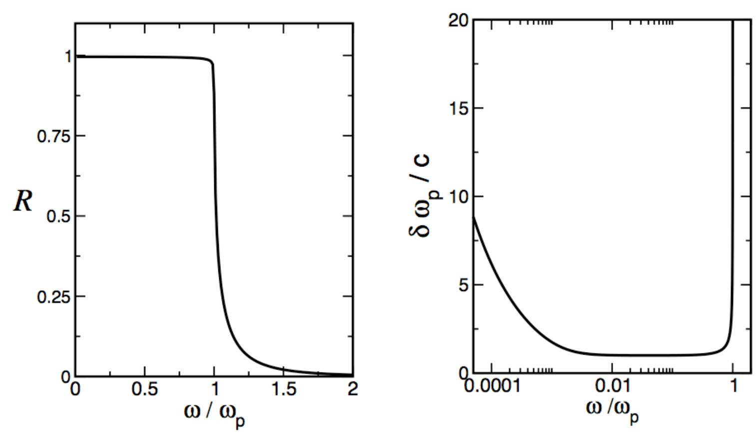



Figure 6.1: The frequency dependent reflectivity and penetration depth for $\omega_p\bar{\tau} = 500$.

**Relaxation regime** $(1 \ll \omega\bar{\tau} \ll \omega_p\bar{\tau})$

Here, we can expand the dielectric function (6.33) in $(\omega\bar{\tau})^{-1}$, yielding

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2} + i\frac{\omega_p^2}{\omega^3\bar{\tau}}. \tag{6.39}$$

The real part $\varepsilon_1 \approx -\omega_p^2/\omega^2$ is large and negative and dominates in magnitude over $\varepsilon_2$. For the optical properties, we obtain

$$k(\omega) \approx \frac{\omega_p}{\omega} \tag{6.40}$$

$$n(\omega) \approx \frac{\omega_p}{2\omega^2 \bar{\tau}}. \tag{6.41}$$

We find $k(\omega) \gg n(\omega) \gg 1$, which implies a large reflectivity of metals in this frequency range as well. Note that visible frequencies are part of this regime (see Figures 6.2 and 6.3). The frequency dependence of the penetration depth becomes weak, and its magnitude is approximately given by the Debye length, $\delta \sim c/\omega_p$.

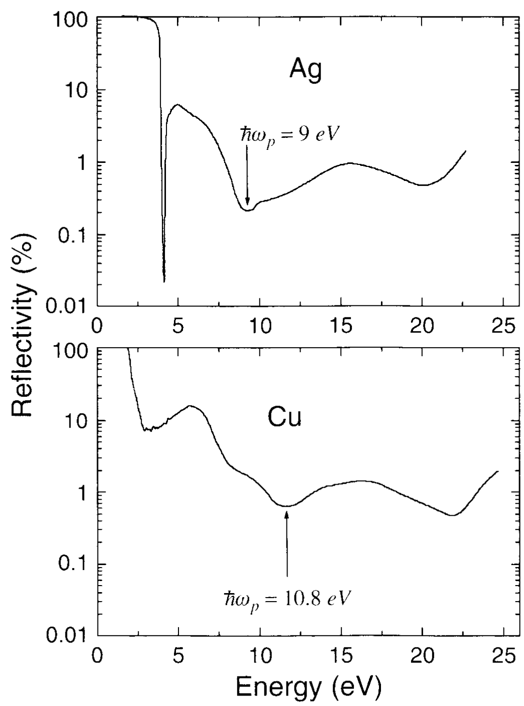



Figure 6.2: Reflectance spectra for silver and copper. In both cases the drop of reflectance is due to optical transitional between the completely filled $d$-band and the partially filled $s$-band. Note the logarithmic scale for the reflectivity. (Source: *An introduction to the optical spectroscopy of inorganic solids*, J. García Solé, L.E. Bausá and D. Jaque, Wiley (2005))

**Ultraviolet regime ($\omega \approx \omega_p$ and $\omega > \omega_p$)**

In this regime, the imaginary part of $\varepsilon$ is approximately zero and the real part has the well known form

$$\varepsilon_1(\omega) = 1 - \frac{\omega_p^2}{\omega^2}, \tag{6.42}$$

such that the reflectivity drops drastically, from close to unity towards zero (cf. Figure 6.1). Metals become nearly transparent in the range $\omega > \omega_p$. In Figure 6.1, one also notices the rapid increase in the penetration depth $\delta$, showing the transparency of the metal.

In all these considerations, we have neglected the contributions to the dielectric function due to the ion cores (core electrons and nuclei). They do, however, influence the reflecting properties of metals; particularly, the value of $\omega_p$ is reduced to $\omega_p' = \omega_p/\sqrt{\varepsilon_\infty}$, where $\varepsilon_\infty$ is the frequency-independent part of the dielectric function due to the ions. With this, the reflectivity for frequencies above $\omega_p'$ approaches $R_\infty = (\varepsilon_\infty - 1)^2/(\varepsilon_\infty + 1)^2$, and $0 < R_\infty < 1$ (see Figure 6.2 and 6.3).

**Color of metals**

The practically full reflectance for frequencies below $\omega_p$ is a typical feature of metals. Since for most metals, the plasma frequency lies well above the range of visible light ($\hbar\omega = 1.5 - 3.5eV$), they appear shiny to our eye. While most polished metal surfaces appear shiny white, like silver, there are some metals with a color, like gold which is yellow and copper which is reddish. White shininess results from reflectance on the whole visible frequency range, while for colored metals there is a certain threshold above which the reflectance drops and frequencies towards blue are not or much weaker reflected. In most cases this drop is not connected with the plasma frequency, but with light absorption due to interband transitions. Note that the single band metal which was used for the Drude theory does not allow for optical absorption apart from the plasma excitation. Interband transition play a particularly important role for the noble metals, Cu, Ag and Au. For these metals, the reflectance drop is caused by the transition from the completely occupied $d$-band to the partially filled $s$-band, $3d \to 4s$ in case of Cu. For copper, this drop appear below 2.5eV so that predominantly red light is reflected (see Figure 6.2). For gold, this threshold frequency is slightly higher, but still in the visible, while for silver, it lies beyond the visible range (see Figure 6.2). For all these cases, the plasma frequency is not so easily recognizable in the reflectance. On the other hand, aluminum shows a reflectance rather close to the expected behavior (see. Figure 6.3). Also here, there is a small reduction of the reflection due to interband absorption. However, this effect is weak and the strong drop occurs at the plasma frequency of $\hbar\omega_p = 15.8eV$. Like silver also polished aluminum is white shiny.

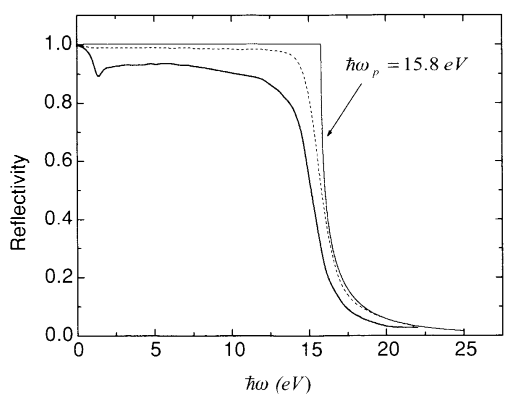



Figure 6.3: Reflectance spectrum of aluminum. The slight reduction of reflectivity below $\omega_p$ is due to interband transitions. The thin solid line is the theoretical behavior for $\tau = 0$ and the dashed line for finite $\tau$. (Source: *An introduction to the optical spectroscopy of inorganic solids*, J. García Solé, L.E. Bausá and D. Jaque, Wiley (2005))

### 6.2.3 The relaxation time

By replacing the collision integral the a relaxation time approximation, we implicitly introduced a connection between the scattering rate $W(\boldsymbol{k}, \boldsymbol{k}')$ and the relaxation time $\tau$. This relation,

$$\frac{f(\boldsymbol{k}) - f_0(\boldsymbol{k})}{\tau(\epsilon_{\boldsymbol{k}})} = \int \frac{d^3k'}{(2\pi)^3} W(\boldsymbol{k}, \boldsymbol{k}')\{f(\boldsymbol{k}) - f(\boldsymbol{k}')\}, \tag{6.43}$$

will be studied for an isotropic system, with elastic scattering, and a small external field $\boldsymbol{E}$. The solution of equation (6.24) is of the form

$$f(\boldsymbol{k}) = f_0(\boldsymbol{k}) + A(k)\boldsymbol{k}\cdot\boldsymbol{E}, \tag{6.44}$$

such that

$$f(\boldsymbol{k}) - f(\boldsymbol{k}') = A(k)(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{E}. \tag{6.45}$$

Without loss of generality, we define $\hat{z} \parallel \boldsymbol{k}$, and introduce the parametrization of the angles $\theta$, polar angle of $\boldsymbol{E}$ and $\theta'$ ($\phi'$) polar (azimuth) angle of $\boldsymbol{k}'$), leading to

$$\boldsymbol{k}\cdot\boldsymbol{E} = kE\cos\theta, \tag{6.46}$$

$$\boldsymbol{k}\cdot\boldsymbol{k}' = kk'\cos\theta', \tag{6.47}$$

$$\boldsymbol{k}'\cdot\boldsymbol{E} = k'E(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\phi'). \tag{6.48}$$

For elastic scattering, $k = k'$, we obtain

$$f(\boldsymbol{k}) - f(\boldsymbol{k}') = A(k)kE\big[\cos\theta(1-\cos\theta') - \sin\theta\sin\theta'\cos\phi'\big]. \tag{6.49}$$

Inserting this into the right-hand side of equation (6.43), the $\phi'$-dependent part of the integration vanishes for an isotropic system, and we are left with

$$\frac{f(\boldsymbol{k}) - f_0(\boldsymbol{k})}{\tau(\epsilon_{\boldsymbol{k}})} = \int d\Omega_{\boldsymbol{k}'}[f(\boldsymbol{k}) - f(\boldsymbol{k}')]W(\boldsymbol{k},\boldsymbol{k}') \tag{6.50}$$

$$= A(k)kE\cos\theta\int d\Omega_{\boldsymbol{k}'}(1-\cos\theta')W(\boldsymbol{k},\boldsymbol{k}') \tag{6.51}$$

$$= [f(\boldsymbol{k}) - f_0(\boldsymbol{k})]\int d\Omega_{\boldsymbol{k}'}(1-\cos\theta')W(\boldsymbol{k},\boldsymbol{k}'). \tag{6.52}$$

The factor $[f(\boldsymbol{k}) - f_0(\boldsymbol{k})]$ can be dropped on both sides of the equation, resulting in

$$\frac{1}{\tau(\epsilon_{\boldsymbol{k}})} = \int \frac{d^3k'}{(2\pi)^3}W(\boldsymbol{k},\boldsymbol{k}')(1-\cos\theta'), \tag{6.53}$$

where one should remember that, for elastic scattering, the quasiparticle energy $\epsilon_{\boldsymbol{k}} = \epsilon_{\boldsymbol{k}'}$ is conserved in the collision process. The scattering probability $W(\boldsymbol{k},\boldsymbol{k}')$ accounts for this restriction. In the next few sections we discuss different scattering processes, looking at collision probabilities, relaxation times and the resulting conductivity and resistivity contributions.

## 6.3 Impurity scattering

### 6.3.1 Potential scattering

Every deviation from the perfect periodicity of the ionic lattice is a source of quasiparticle scattering, leading to the loss of their original momentum. Without translational invariance, the conservation of momentum is lost, the energy, however, is still conserved. Possible static scatterers are among others vacancies, dislocations, and impurity atoms. The scattering rate $W(\boldsymbol{k},\boldsymbol{k}')$ for a potential $\widehat{V}$ can be determined applying Fermi's golden rule,[6]

$$W(\boldsymbol{k},\boldsymbol{k}') = \frac{2\pi}{\hbar}n_{\text{imp}}|\langle\boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2\delta(\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}'}). \tag{6.54}$$

[6] This corresponds to the first Born approximation in scattering theory. Note, that this approximation is insufficient to describe resonant scattering.

By $n_{\text{imp}}$ we denote the density of impurities, assuming only one species of them. For small densities $n_{\text{imp}}$, it is reasonable to neglect interference effects between different impurities. According to equation (6.53), the relaxation time $\tau$ of a quasiparticle with momentum $\hbar\boldsymbol{k}$ is given by

$$\frac{1}{\tau(\epsilon_{\boldsymbol{k}})} = \frac{2\pi}{\hbar} n_{\text{imp}} \int \frac{d^3k'}{(2\pi)^3} |\langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2 (1 - \hat{k}\cdot\hat{k}')\delta(\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}'}) \tag{6.55}$$

$$= n_{\text{imp}}(\hat{k}\cdot\boldsymbol{v}_{\boldsymbol{k}}) \int \frac{d\sigma}{d\Omega}(\boldsymbol{k},\boldsymbol{k}')(1 - \hat{k}\cdot\hat{k}')\frac{d\Omega_{\boldsymbol{k}'}}{4\pi}, \tag{6.56}$$

with the differential scattering cross section $d\sigma/d\Omega$ and $\hat{k} = \boldsymbol{k}/|\boldsymbol{k}|$. Here, we used the connection[7] between Fermi's golden rule and the Born approximation. Note the difference in the expressions for the relaxation time $\tau$ in equation (6.53) and for the lifetime $\tilde{\tau}$,

$$\frac{1}{\tilde{\tau}} = \int \frac{d^3k}{(2\pi)^3} W(\boldsymbol{k},\boldsymbol{k}'), \tag{6.60}$$

given by Fermi's golden rule. The factor $(1 - \cos\theta')$ in equation (6.53) gives more weight to backscattering ($\theta' \approx \pi$) compared to forward scattering ($\theta' \approx 0$), since the former has more influence in impeding transport. This explains why $\tau$ is also termed *transport lifetime.*

Assuming defects in the form of point charges $Ze$, whose screened potential is

$$\langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle = \frac{4\pi Z e^2}{|\boldsymbol{k} - \boldsymbol{k}'|^2 + k_{\text{TF}}^2}. \tag{6.61}$$

In the limit of very strong screening, $k_{\text{TF}} \gg k_F$, the differential cross section becomes independent of the deviation[8] $(\boldsymbol{k} - \boldsymbol{k}')$, the transport and the usual lifetime become equal, $\tau = \tilde{\tau}$, and

$$\frac{1}{\tau} \approx \frac{\pi}{\hbar} N(\epsilon_F) n_{\text{imp}} \left(\frac{4\pi Z e^2}{k_{\text{TF}}^2}\right)^2. \tag{6.62}$$

With this, we are now able to determine the conductivity for scattering on Coulomb defects, assuming $s$-wave scattering only. Then, since $\tau(\epsilon)$ depends weakly on energy, equation (6.30) yields

$$\sigma = \frac{e^2 n \tau(\epsilon_F)}{m}, \tag{6.63}$$

or, for the specific resistivity $\rho = 1/\sigma$,

$$\rho = \frac{m}{e^2 n \tau(\epsilon_F)}. \tag{6.64}$$

---

[7] The scattering of particles with momentum $\hbar\boldsymbol{k}$ into the solid angle $d\Omega_{\boldsymbol{k}'}$ around $\boldsymbol{k}'$ yields

$$W(\boldsymbol{k},\boldsymbol{k}')d\Omega_{\boldsymbol{k}'} = \frac{2\pi}{\hbar\Omega} \sum_{\boldsymbol{k}'\in d\Omega_{\boldsymbol{k}'}} |\langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2 \delta(\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}'}) \tag{6.57}$$

$$= \frac{2\pi}{\hbar} d\Omega_{\boldsymbol{k}'} \int\limits_{\boldsymbol{k}'\in d\Omega_{\boldsymbol{k}'}} \frac{d^3k'}{(2\pi)^3} \langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2 \delta(\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}'}) = \frac{2\pi}{\hbar} d\Omega_{\boldsymbol{k}'} N(\epsilon) |\langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2. \tag{6.58}$$

The scattering per incoming particle current $j_{\text{in}} d\sigma(\boldsymbol{k},\boldsymbol{k}') = W(\boldsymbol{k},\boldsymbol{k}')d\Omega_{\boldsymbol{k}'}$ determines the differential cross section

$$\hat{k}\cdot\boldsymbol{v}_{\boldsymbol{k}} \frac{d\sigma}{d\Omega}(\boldsymbol{k},\boldsymbol{k}') = \frac{2\pi}{\hbar}\frac{N(\epsilon)}{4\pi} |\langle \boldsymbol{k}'|\widehat{V}|\boldsymbol{k}\rangle|^2. \tag{6.59}$$

leading to equation (6.56).

[8] In the context of partial wave expansion, one speaks of *s-wave scattering*, i.e., $\delta_{l>0} \to 0$.

Both $\sigma$ and $\rho$ are independent of temperature. This contribution is called the *residual resistivity* of a metal, which approaches zero for a perfect material. The temperature dependence of the resistivity is induced in other scattering processes like electron-phonon scattering and electron-electron scattering, which will be considered below. The so-called *residual resistance ratio* $RRR = R(T = 300K)/R(T = 0)$ is an often used quantity to benchmark the quality of a material. It is defined as the ratio between the resistance $R$ at room temperature and the resistance at zero temperature. The bigger the $RRR$, the better the quality of the material. The typical value of $RRR$ for common copper is 40-50, while the $RRR$ for very clean aluminum reaches values up to 20000.

### 6.3.2 Kondo effect

There are impurity atoms inducing so-called *resonant scattering*. If the resonance occurs close to the Fermi energy, the scattering rate is strongly energy dependent, inducing a more pronounced temperature dependence of the resistivity. An important example is the scattering off magnetic impurities with a spin degree of freedom, yielding a dramatic energy dependence of the scattering rate. This problem was first studied by Kondo in 1964 in order to explain the peculiar minima in resistivity in some materials. The coupling between the local spin impurities $\boldsymbol{S}_i$ at $\boldsymbol{R}_i$ and the quasiparticle spin $\boldsymbol{s}$ has the exchange form

$$\begin{aligned}
\widehat{V}_K &= \sum_i \left(\widehat{V}_K\right)_i = J\sum_i \widehat{\boldsymbol{S}}_i \cdot \widehat{\boldsymbol{s}}(\boldsymbol{r})\delta(\boldsymbol{r}-\boldsymbol{R}_i) &\quad (6.65)\\
&= J\sum_i \left(\widehat{S}_i^z\widehat{s}^z(\boldsymbol{r}) + \frac{1}{2}\widehat{S}_i^+\widehat{s}^-(\boldsymbol{r}) + \frac{1}{2}\widehat{S}_i^-\widehat{s}^+(\boldsymbol{r})\right)\delta(\boldsymbol{r}-\boldsymbol{R}_i) &\quad (6.66)\\
&= \frac{J\hbar}{2\Omega}\sum_{\boldsymbol{k},\boldsymbol{k}',i}\left[\widehat{S}_i^z(\widehat{c}^\dagger_{\boldsymbol{k}\uparrow}\widehat{c}_{\boldsymbol{k}'\uparrow} - \widehat{c}^\dagger_{\boldsymbol{k}\downarrow}\widehat{c}_{\boldsymbol{k}'\downarrow}) + S_i^+\widehat{c}^\dagger_{\boldsymbol{k}\downarrow}\widehat{c}_{\boldsymbol{k}'\uparrow} + S_i^-\widehat{c}_{\boldsymbol{k}\uparrow}\widehat{c}_{\boldsymbol{k}'\downarrow}\right]e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}_i}. &\quad (6.67)
\end{aligned}$$

Here, it becomes important that spin flip processes, which change the spin state of the impurity and that of the scattered electron, are enabled. The results for the scattering rate are presented here without derivation,

$$W(\boldsymbol{k},\boldsymbol{k}') \approx J^2S(S+1)\left[1 + 2JN(\epsilon_F)\ln\left(\frac{D}{|\epsilon_{\boldsymbol{k}} - \epsilon_F|}\right)\right], \quad (6.68)$$

where $D$ is the bandwidth and we have assumed that $JN(\epsilon_F) \ll 1$. The relaxation time is found to be

$$\frac{1}{\tau(\epsilon_{\boldsymbol{k}})} \approx \frac{J^2S(S+1)}{\hbar}N(\epsilon)\left[1 + 2JN(\epsilon_F)\ln\left(\frac{D}{|\epsilon_{\boldsymbol{k}} - \epsilon_F|}\right)\right]. \quad (6.69)$$

Note that $W(\boldsymbol{k},\boldsymbol{k}')$ does not depend on angle, meaning that the process is described by $s$-wave scattering. The energy dependence is singular at the Fermi energy, indicating that we are not dealing with simple resonant potential scattering, but with a much more subtle many-particle effect involving the electrons very near the Fermi surface. The fact that the local spins $\boldsymbol{S}_i$ can flip, makes the scattering center dynamical, because the scatterer is constantly changing. The scattering process of an electron is influenced by previous scattering events, leading to the singularity at $\epsilon_F$. This cannot be described within the first Born approximation, but requires at least the second approximation or the full solution.[9] As mentioned before, the resonant behavior

[9] We refer to J.M. Ziman, *Principles of the Theory of Solids*, and A.C. Hewson, *The Kondo Problem to Heavy Fermions* for more details.

induces a strong temperature dependence of the conductivity. Indeed,

$$\sigma(T) = \frac{e^2 k_F^3}{6\pi^2 m} \int d\epsilon \frac{1}{4k_B T \cosh^2(\epsilon - \epsilon_F)/2k_B T)} \tau(\epsilon) \tag{6.70}$$

$$\approx \frac{e^2 n}{8mk_B T} \int d\tilde{\epsilon} \frac{J^2 S(S+1)[1 - 2JN(\epsilon_F)\ln(D/\tilde{\epsilon})]}{\cosh^2(\tilde{\epsilon}/2k_B T)}. \tag{6.71}$$

A simple substitution in the integral leads to

$$\sigma(T) \approx \frac{e^2 n}{2m} J^2 S(S+1) \left[1 - 2JN(\epsilon_F) \ln\left(\frac{D}{k_B T}\right)\right]. \tag{6.72}$$

Usual contributions to the resistance, like electron-phonon scattering discussed below, typically decrease with temperature. The contribution (6.72) to the conductivity is strongly increasing, inducing a minimum in the resistance, when we crossover from the decreasing behavior at high temperatures to the low-temperature increase of $\rho(T)$. At even lower temperatures, the conductivity would decrease and eventually even turn negative which is an artifact of our approximation. In reality, the conductivity saturates at a finite value when the temperature is lowered below a characteristic Kondo temperature $T_K$,

$$k_B T_K = D e^{-1/JN(\epsilon_F)}, \tag{6.73}$$

a characteristic energy scale of this system. The real behavior of the conductivity at temperatures $T \ll T_K$ is not accessible by simple perturbation theory. This regime, known as the Kondo problem, represents one of the most interesting correlation phenomena of many-particle physics.

## 6.4 Electron-phonon interaction

Even in perfect metals, the conductivity becomes non-zero at finite temperature. The thermally induced distortions of the lattice, phonons, act as fluctuating scattering centers. In the language of electron-phonon interaction, electrons are scattered via absorption and emission of phonons, which induce local fluctuations in volume (cf. Chapter 3). The corresponding coupling term was given in equation (3.127) and simplifies with the definition (3.114) to

$$\mathcal{H}_{\text{int}} = 2i \sum_{\boldsymbol{k},\boldsymbol{q},s} \tilde{V}_{\boldsymbol{q}} \sqrt{\frac{\hbar}{2\rho_0 \omega_{\boldsymbol{q}}}} |\boldsymbol{q}| (\widehat{b}_{\boldsymbol{q}} - \widehat{b}^\dagger_{-\boldsymbol{q}}) \widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q},s} \widehat{c}_{\boldsymbol{k}s}. \tag{6.74}$$

The interaction is similar to the interaction between electrons and photons. The dominant processes consist of single-phonon processes, i.e., the absorption or emission of one phonon. Energy and momentum are conserved, such that, for the scattering of an electron from momentum $\boldsymbol{k}$ to $\boldsymbol{k}'$ due to the emission of a phonon with momentum $\boldsymbol{q}$, we have

$$\boldsymbol{k} = \boldsymbol{k}' + \boldsymbol{q} + \boldsymbol{G}, \tag{6.75}$$

$$\epsilon_{\boldsymbol{k}} = \hbar\omega_{\boldsymbol{q}} + \epsilon_{\boldsymbol{k}'}, \tag{6.76}$$

Here, $\omega_{\boldsymbol{q}} = c_s q$ is the phonon spectrum, while the reciprocal lattice vector $\boldsymbol{G}$ allows for scattering[10] in nearby Brillouin zones. By this, the phase space available for scattering is strongly reduced, especially near the Fermi energy. Note that $\hbar\omega_{\boldsymbol{q}} \leq \hbar\omega_D \ll \epsilon_F$. In order to calculate

[10]This so-called *Umklapp phenomenon* will be discussed in some more detail later in this chapter.

the scattering rates, the matrix elements[11] of the available processes,

$$\langle \boldsymbol{k}+\boldsymbol{q}; N_{\boldsymbol{q}'}|(\widehat{b}_{\boldsymbol{q}} - \widehat{b}^\dagger_{-\boldsymbol{q}})\widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q},s}\widehat{c}_{\boldsymbol{k}s}|\boldsymbol{k}; N'_{\boldsymbol{q}'}\rangle = \langle \boldsymbol{k}+\boldsymbol{q}|\widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q},s}\widehat{c}_{\boldsymbol{k}s}|\boldsymbol{k}\rangle\Big(\sqrt{N'_{\boldsymbol{q}'}}\;\delta_{N_{\boldsymbol{q}'},N'_{\boldsymbol{q}'}-1}\;\delta_{\boldsymbol{q},\boldsymbol{q}'} - \sqrt{N'_{\boldsymbol{q}'}+1}\;\delta_{N_{\boldsymbol{q}'},N'_{\boldsymbol{q}'}+1}\;\delta_{\boldsymbol{q},-\boldsymbol{q}'}\Big), \tag{6.77}$$

need to be calculated. From Fermi's golden rule we obtain

$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{2\pi}{\hbar}\sum_{\boldsymbol{q}}|g(\boldsymbol{q})|^2\Big[&\big[f(\boldsymbol{k})(1-f(\boldsymbol{k}+\boldsymbol{q}))(N_{-\boldsymbol{q}}+1)\\ &- f(\boldsymbol{k}+\boldsymbol{q})(1-f(\boldsymbol{k}))N_{-\boldsymbol{q}}\big]\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}+\hbar\omega_{-\boldsymbol{q}})\\ &-\big[f(\boldsymbol{k}+\boldsymbol{q})(1-f(\boldsymbol{k}))(N_{\boldsymbol{q}}+1)\\ &- f(\boldsymbol{k})(1-f(\boldsymbol{k}+\boldsymbol{q}))N_{\boldsymbol{q}}\big]\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}-\hbar\omega_{\boldsymbol{q}})\Big],\end{aligned} \tag{6.78}$$

where $g(\boldsymbol{q}) = \tilde{V}_{\boldsymbol{q}}|\boldsymbol{q}|\sqrt{2\hbar/\rho_0\omega_{\boldsymbol{q}}}$. Each of these four terms describes one of the single phonon scattering processes depicted in Figure 6.4.

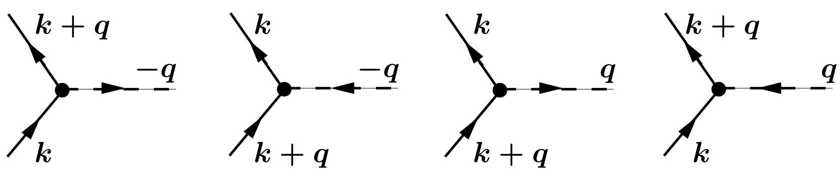



Figure 6.4: The four single-phonon electron-phonon scattering processes.

The collision integral leads to a complicated integro-differential equation, whose solution is very tedious. Instead of a full rigorous calculation including the non-equilibrium redistribution of phonons, we will consider the behavior in various temperature regimes by an approximate treatment of the phonons. The characteristic temperature of phonons, the Debye temperature $\Theta_D \ll T_F$, is much smaller than the Fermi temperature. Hence, the phonon energy is virtually unimportant for the energy conservation, $\epsilon_{\boldsymbol{k}'=\boldsymbol{k}+\boldsymbol{q}} \approx \epsilon_{\boldsymbol{k}}$. Therefore we are allowed to impose momentum conservation $\epsilon_{\boldsymbol{k}+\boldsymbol{q}} = \epsilon_{\boldsymbol{k}}$ and consider the lattice distortion as being essentially static, in the sense of an adiabatic Born-Oppenheimer approximation. The approximate collision integral then reads

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{2\pi}{\hbar}\sum_{\boldsymbol{q}}|g(\boldsymbol{q})|^2 2N(\omega_{\boldsymbol{q}})[f(\boldsymbol{k}+\boldsymbol{q})-f(\boldsymbol{k})]\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}), \tag{6.79}$$

where we assume the occupation of phonon states according to the equilibrium distribution for bosons,

$$N(\omega_{\boldsymbol{q}}) = \frac{1}{e^{\hbar\omega_{\boldsymbol{q}}/k_BT}-1}. \tag{6.80}$$

This approximation includes all important aspects of the electron-phonon scattering we need to derive the temperature dependence of $\rho(T)$.

[11] In analogy to the discussion on electromagnetic radiation, the phenomenon of *spontaneous phonon emission* due to zero-point fluctuations appears. It is formally visible in the additional "+1" in the factors $(N_{\pm\boldsymbol{q}})+1$.

In analogy to previous approaches, we obtain for with relaxation-time Ansatz

$$\frac{1}{\tau(\epsilon_{\boldsymbol{k}})} = \frac{2\pi}{\hbar}\frac{\lambda}{N(\epsilon_F)}\int \frac{d^3q}{(2\pi)^3}\hbar\omega_{\boldsymbol{q}}N(\omega_{\boldsymbol{q}})(1-\cos\theta)\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}), \tag{6.81}$$

where $|\boldsymbol{k}| = |\boldsymbol{k}+\boldsymbol{q}| = k_F$, meaning that only the electrons in a thin shell close to the Fermi surface are relevant. Furthermore, we parametrized $g(\boldsymbol{q})$ according to

$$|g(\boldsymbol{q})|^2 = \frac{\lambda}{2N(\epsilon_F)\Omega}\hbar\omega_{\boldsymbol{q}}, \tag{6.82}$$

where $\lambda$ is a dimensionless electron-phonon coupling constant. In usual metals $\lambda < 1$. As in the case of defect scattering, the relaxation time depends only weakly on the electron energy. But, unlike previously, the direct temperature dependence enters via the dependence on temperature of the phonon occupation $N(\omega_{\boldsymbol{q}})$.

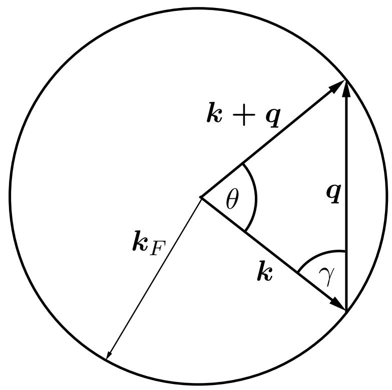



Figure 6.5: The geometry of electron-phonon scattering.

In order to perform the integration in equation (6.81), we have to re-express $\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}})$ by writing

$$\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}) = \delta\left(\frac{\hbar^2}{2m}(q^2-2k_Fq\cos\gamma)\right) = \frac{m}{\hbar^2 k_F q}\delta\left(\frac{q}{2k_F}-\cos\gamma\right), \tag{6.83}$$

where $\gamma$ is defined in Figure 6.5. From there, we also see that $2\gamma+\theta=\pi$, and thus, find the relation

$$1-\cos\theta = 1+\cos(2\gamma) = 2\cos^2(\gamma). \tag{6.84}$$

Obviously, we have to integrate $q$ over the range $[0, 2k_F]$ on the right-hand side of equation (6.81), which can be reformulated to

$$\frac{1}{\tau(\epsilon_F, T)} = \frac{-\lambda}{N(\epsilon_F)}\frac{m}{\hbar^2\pi k_F}\int_0^{2k_F} dq\, q\omega_{\boldsymbol{q}}N(\omega_{\boldsymbol{q}})\int_0^{\pi/2} d\gamma\ \sin\gamma\cos^2(\gamma)\,\delta\big(\frac{q}{2k_F}-\cos\gamma\big) \tag{6.85}$$

$$= \frac{\lambda}{4N(\epsilon_F)}\frac{mc_s}{\hbar^2\pi k_F^3}\int_0^{2k_F}\frac{q^4dq}{e^{\hbar c_s q/k_BT}-1} \tag{6.86}$$

$$= \frac{\lambda}{4N(\epsilon_F)}\frac{mc_sk_F^2}{\hbar^2\pi}\left(\frac{T}{\Theta_D}\right)^5\int_0^{2\Theta_D/T}\frac{y^4dy}{e^y-1}, \tag{6.87}$$

where we have approximated the Debye temperature by $k_B\Theta_D \approx 2\hbar c_s k_F$. We notice the two distinct characteristic temperature regimes,

$$\frac{1}{\tau} = \begin{cases} 6\zeta(5)\lambda\pi\frac{k_B\Theta_D}{\hbar}\left(\frac{T}{\Theta_D}\right)^5, & T \ll \Theta_D, \\ \lambda\pi\frac{k_B\Theta_D}{\hbar}\left(\frac{T}{\Theta_D}\right), & T \gg \Theta_D. \end{cases} \tag{6.88}$$

The prefactors depend on the details of the approximation, whereas the qualitative temperature dependence does not. We finally obtain the conductivity and resistivity from equation (6.28),

$$\sigma = \frac{e^2 n}{m}\tau(T), \tag{6.89}$$

$$\rho = \frac{m}{e^2 n}\frac{1}{\tau(T)}, \tag{6.90}$$

where we used the weak energy dependence of $\tau(\epsilon \approx \epsilon_F, T)$. With this, we obtain the well-known Bloch-Grüneisen form

$$\rho(T) \propto \begin{cases} T^5, & T \ll \Theta_D, \\ T, & T \gg \Theta_D. \end{cases} \tag{6.91}$$

At high temperatures, $\rho$ is determined by the occupation of phonon states

$$N(\omega_q) \approx \frac{k_B T}{\hbar\omega_q} \tag{6.92}$$

which change the scattering strength (amplitude) of the lattice modulation linear in $T$. At low temperature only the lowest phonon states are occupied $\hbar\omega_q < k_B T$ yield $q < k_B T/\hbar c_s$. Thus, at low temperatures only long-wave length modulations of the lattice generate a scattering potentials which deflects electrons only slightly from their trajectories (forward scattering dominates). This represents a restriction of the scattering phase space becoming ever smaller with decreasing temperature.

## 6.5 Electron-electron scattering

In Chapter 5 we have learned, that, taking a short-ranged electron-electron interaction into account, the lifetime of a quasiparticle strongly increases close to the Fermi surface. The basic reason was the constraint of the scattering phase space due to the Pauli principle. The lifetime, which we can identify with the relaxation time here, has the form

$$\frac{1}{\tau(\epsilon)} \propto (\epsilon - \epsilon_F)^2. \tag{6.93}$$

Note that this phase space reduction does not restrict the momentum transfer, so that in principle $0 < q < 2k_F$ is valid for all energies $\epsilon$. This allows determining the conductivity from equation (6.28), and we find

$$\sigma(T) \propto \frac{1}{T^2} \tag{6.94}$$

The resistivity $\rho \propto T^2$ increases quadratically in $T$. This is a key property of a Fermi liquid and is often considered an identifying criterion.

**Umklapp process**

An important point, kept quiet so far, requires some explanation. One could argue, that the momentum of the Fermi liquid is conserved upon the collision of two electrons. With this argument, it is not quite clear what causes a finite resistance. This argument, however, is based on translational invariance and ignores the existence of the underlying lattice. In the sense that the kinematics (momentum conservation) is also satisfied for electrons being scattered from the Fermi surface of one Brillouin zone to the one of another Brillouin zone, while incorporating a reciprocal lattice vector. The equation of momentum conservation,

$$\boldsymbol{k} = \boldsymbol{k}' + \boldsymbol{G}, \tag{6.95}$$

where $\boldsymbol{G}$ is a reciprocal vector of the lattice, allows for scattering to other Brillouin zones ($\boldsymbol{G} \neq 0$). By this, the momentum is transferred to a static deformation of the lattice. Such processes are termed *Umklapp* processes and play an important role in electron-phonon scattering as well (see equation (6.75)).

## 6.6 Matthiessen's rule and the Ioffe-Regel limit

Matthiessen's rule states, that the scattering rates of different scattering processes can simply be added, leading to

$$W(\boldsymbol{k}, \boldsymbol{k}') = W_1(\boldsymbol{k}, \boldsymbol{k}') + W_2(\boldsymbol{k}, \boldsymbol{k}'), \tag{6.96}$$

or, expressed in the relaxation time approximation,

$$\frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_2}, \tag{6.97}$$

and

$$\rho = \frac{m}{ne^2\tau} = \frac{m}{ne^2}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right) = \rho_1 + \rho_2. \tag{6.98}$$

This rule is not a theorem and corresponds effectively to a serial coupling of resistors in a classical circuit. It is applicable, if the different scattering processes are independent. Actually, the assumption that the impurity scattering rate depends linearly on the impurity density $n_{\mathrm{imp}}$ is already an application of Matthiessen's rule. Mutual influences of impurities, e.g., through interference effects due to the coherent scattering of an electron at different impurities, would invalidate this simplification. An example where Matthiessen's rule is violated is a one-dimensional system, where a single scatterer $i$ induces a finite resistance $R_i$. Two serial scatterers then lead to a total resistance

$$R = R_1 + R_2 + \frac{2e^2}{h} R_1 R_2 \geq R_1 + R_2. \tag{6.99}$$

The reason is, that in one-dimensional systems, the interference of backscattered waves is unavoidable and no impurity can be treated as isolated. Furthermore, every particle traversing the whole system has to pass all scatterers. The more general Matthiessen's rule,

$$\rho \geq \rho_1 + \rho_2, \tag{6.100}$$

is still valid. Another source of deviation from Matthiessen's rule arises, if the relaxation time depends on $\boldsymbol{k}$, since then the averaging is not the same for all scattering processes. The electron-phonon coupling can be modified by the scattering on impurities, most importantly in the presence of anisotropic Fermi surfaces. For the analysis of resistance data of simple metals, we often assume the validity of Matthiessen's rule. A typical example is the resistance minimum explained by Kondo, where

$$\rho(T) = \rho_0 + \rho_{e-p}(T) + \rho_K(T) + \rho_{e-e}(T) \tag{6.101}$$

$$= \rho_0 + \alpha T^5 + \beta(1 + 2JN(\epsilon_F)\ln(D/k_B T)) + \gamma T^2, \tag{6.102}$$

where $\alpha$, $\beta$, and $\gamma$ are numerical constants. Upon decreasing temperature, the Kondo term is increasing, whereas the electron-phonon and electron-electron contributions decrease. Consequently, there is a minimum.

We now turn to the discussion of resistivity in the high-temperature limit. Believing the previous considerations entirely, the electrical resistivity would grow indefinitely with temperature. In most cases, however, the resistivity will saturate at a finite limiting value. We can understand this from simple considerations writing the mean free path $\ell = v_F \tau(\epsilon_F)$ as the mean distance an electron travels freely between two collisions. The lattice constant $a$ is a natural lower boundary to $\ell$ in the crystal lattice. Furthermore, we assumed so far that scattering occurs between two states with sharp momenta $\boldsymbol{k}$ and $\boldsymbol{k}'$. If the de Broglie wavelength becomes comparable to the mean free path, the framework becomes unfounded and $k_F^{-1}$ would become a boundary for $\ell$. In most systems $a$ and $k_F^{-1}$ are comparable lengths. Empirically, the resistivity is described via the formula

$$\frac{1}{\rho(T)} = \frac{1}{\rho_{BT}(T)} + \frac{1}{\rho_{\max}}, \tag{6.103}$$

corresponding to the parallel addition of two resistivities; on one hand, $\rho_{BT}(T)$, which we have investigated using the Boltzmann transport theory, and on the other hand the limiting value $\rho_{\max}$. This is in clear disagreement to Matthiessen's rule, which is to be expected, since for $k_F\ell \sim 1$, complex interference effects will arise. The saturated resistivity $\rho_{\max}$ can be estimated from the Jellium model,

$$\rho_{\max} = \frac{m}{e^2 n \tau(\epsilon_F)} = \frac{3\pi^2 m}{e^2 k_F^3 \tau(\epsilon_F)} = \frac{h}{e^2}\frac{3\pi}{2k_F^2 \ell} \tag{6.104}$$

$$\sim \frac{h}{e^2}\frac{3\pi}{2k_F}, \tag{6.105}$$

where we used $\ell^{-1} \sim k_F$. For a typical value $k_F \sim 10^8\text{cm}^{-1}$ of the Fermi wave vector, we find $\rho_{\max} \sim 1\text{m}\Omega\text{cm}$, which is called the Ioffe-Regel[12] limit. Establishing a quantitative estimate of $\rho_{\max}$ for a given material is often difficult. There are even materials whose resistivity surpasses the Ioffe-Regel limit.

## 6.7 General transport coefficients

Simultaneously with charge, electrons will also transport energy, i.e., heat and entropy. This is why charge and heat transport are naturally interconnected. In the following, we generalize the transport theory set up above to include this interplay.

[12] The saturated resistivity $\rho_{\max} \sim 1\text{m}\Omega\text{cm} = 1000\mu\Omega\text{cm}$ should be compared to the room-temperature resistivity of good conductors,

| metal | Cu | Au | Ag | Pt | Al | Sn | Na | Fe | Ni | Pb |
|---|---|---|---|---|---|---|---|---|---|---|
| $\rho$ [$\mu\Omega$cm] | 1.7 | 2.2 | 1.6 | 10.5 | 2.7 | 11 | 4.6 | 9.8 | 7 | 21 |

which are all well below $\rho_{\max}$.

### 6.7.1 Generalized Boltzmann equation

We consider a metal with weakly space-dependent temperature $T(\boldsymbol{r})$ and chemical potential $\mu(\boldsymbol{r})$. The distribution function then reads

$$\delta f(\boldsymbol{k};\boldsymbol{r}) = f(\boldsymbol{k};\boldsymbol{r}) - f_0(\boldsymbol{k}, T(\boldsymbol{r}), \mu(\boldsymbol{r})), \tag{6.106}$$

where

$$f_0(\boldsymbol{k}, T(\boldsymbol{r}), \mu(\boldsymbol{r})) = \frac{1}{e^{(\epsilon_{\boldsymbol{k}}-\mu(\boldsymbol{r}))/k_B T(\boldsymbol{r})}+1}. \tag{6.107}$$

In addition, we require the charge density to remain constant in space, i.e.,

$$\int d^3k \, \delta f(\boldsymbol{k};\boldsymbol{r}) = 0 \tag{6.108}$$

for all $\boldsymbol{r}$. In this section, we introduce the *electrochemical potential* $\eta(\boldsymbol{r}) = e\phi(\boldsymbol{r}) + \mu(\boldsymbol{r})$ generating the general force field $\boldsymbol{\mathcal{E}} = -\boldsymbol{\nabla}(e\phi + \mu)$, where $\phi(\boldsymbol{r})$ denotes the electrostatic potential which produces the electric field $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$. With this, the Boltzmann equation for the stationary situation reads

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \boldsymbol{v}_{\boldsymbol{k}} \cdot \boldsymbol{\nabla}_{\boldsymbol{r}} f + \dot{\boldsymbol{k}} \cdot \boldsymbol{\nabla}_{\boldsymbol{k}} f \tag{6.109}$$

$$= -\frac{\partial f}{\partial \epsilon_{\boldsymbol{k}}} \boldsymbol{v}_{\boldsymbol{k}} \cdot \left(\frac{\epsilon_{\boldsymbol{k}} - \mu}{T} \boldsymbol{\nabla}_{\boldsymbol{r}} T - \boldsymbol{\mathcal{E}}\right). \tag{6.110}$$

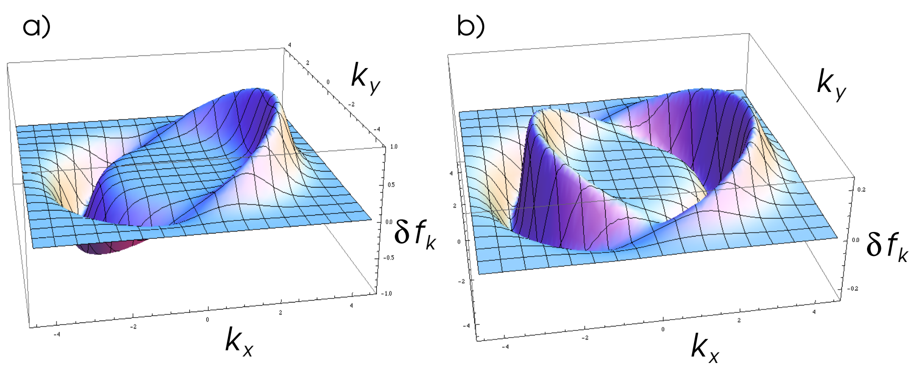



Figure 6.6: Schematic view of the distribution functions $\delta f(\boldsymbol{k})$ on a slice cut through the $k$-space ($k_z = 0$) with a circular Fermi surface for two situations. On the left panel (a), for an applied electric field along the negative $x$-direction, on the right panel (b) for a temperature gradient in $x$-direction.

In the relaxation time approximation for the collision integral, we obtain the solution

$$\delta f(\boldsymbol{k}) = -\frac{\partial f_0}{\partial \epsilon_{\boldsymbol{k}}} \tau(\epsilon_{\boldsymbol{k}}) \boldsymbol{v}_{\boldsymbol{k}} \cdot \left(\boldsymbol{\mathcal{E}} - \frac{\epsilon_{\boldsymbol{k}} - \mu}{T} \boldsymbol{\nabla}_{\boldsymbol{r}} T\right), \tag{6.111}$$

which allows us to calculate the charge and heat currents,

$$\boldsymbol{J}_e = 2\int \frac{d^3k}{(2\pi)^3} e \boldsymbol{v}_{\boldsymbol{k}} \delta f_{\boldsymbol{k}}, \tag{6.112}$$

$$\boldsymbol{J}_q = 2\int \frac{d^3k}{(2\pi)^3} (\epsilon_{\boldsymbol{k}} - \mu) \boldsymbol{v}_{\boldsymbol{k}} \delta f_{\boldsymbol{k}}, \tag{6.113}$$

respectively. Inserting the solution (6.111) into the two definitions above yields

$$\boldsymbol{J}_e = e^2\hat{K}^{(0)}\boldsymbol{\mathcal{E}} + \frac{e}{T}\hat{K}^{(1)}\left(-\boldsymbol{\nabla}T\right), \tag{6.114}$$

$$\boldsymbol{J}_q = e\hat{K}^{(1)}\boldsymbol{\mathcal{E}} + \frac{1}{T}\hat{K}^{(2)}\left(-\boldsymbol{\nabla}T\right), \tag{6.115}$$

where $\boldsymbol{\nabla}$ should be understood as $\boldsymbol{\nabla}_{\boldsymbol{r}}$ and the tensors $\hat{K}^{(n)}$ ($n \in \mathbb{N}_0$) are defined as

$$K^{(n)}_{\alpha\beta} = -\frac{1}{4\pi^3}\int d\epsilon\, \frac{\partial f_0}{\partial \epsilon}\tau(\epsilon)(\epsilon-\mu)^n \int d\Omega_{\boldsymbol{k}} \frac{v_{F\alpha}v_{F\beta}}{\hbar|\boldsymbol{v}_F|}. \tag{6.116}$$

In the case $T \ll T_F$ we can calculate the coefficients,[13]

$$K^{(0)}_{\alpha\beta} = \frac{1}{4\pi^3\hbar}\tau(\epsilon_F)\int d\Omega_{\boldsymbol{k}}\frac{v_{F\alpha}v_{F\beta}}{|\boldsymbol{v}_F|}, \tag{6.119}$$

$$\hat{K}^{(1)} = \frac{\pi^2}{3}(k_BT)^2\left.\frac{\partial}{\partial\epsilon}\hat{K}^{(0)}(\epsilon)\right|_{\epsilon=\epsilon_F}, \tag{6.120}$$

$$\hat{K}^{(2)} = \frac{\pi^2}{3}(k_BT)^2\hat{K}^{(0)}(\epsilon_F). \tag{6.121}$$

We measure the electrical resistivity assuming thermal equilibrium, $\boldsymbol{\nabla}T = 0$ for all $\boldsymbol{r}$. With this, we find the expression

$$\hat{\sigma} = e^2\hat{K}^{(0)}. \tag{6.122}$$

In order to determine the thermal conductivity $\hat{\kappa}$ relating the heat current $\boldsymbol{J}_q$ to $\boldsymbol{\nabla}T$ when no external electric field $\boldsymbol{E}$ is applied, we set $\boldsymbol{J}_e = 0$ as for an open circuit. Then, the equations (6.114) and (6.115) reveal the appearance of an electrochemical field

$$\boldsymbol{\mathcal{E}} = \frac{1}{T}\big(\hat{K}^{(0)}\big)^{-1}\hat{K}^{(1)}\boldsymbol{\nabla}T. \tag{6.123}$$

Thus, the heat current is given by

$$\mathbf{J}_q = -\frac{1}{T}\left(\hat{K}^{(2)} - \hat{K}^{(1)}\big(\hat{K}^{(0)}\big)^{-1}\hat{K}^{(1)}\right)\boldsymbol{\nabla}T = -\hat{\kappa}\boldsymbol{\nabla}T. \tag{6.124}$$

In simple metals, the second term in (6.124) is often negligible as compared to the first one and we obtain in this case

$$\hat{\kappa} = \frac{1}{T}\hat{K}^{(2)} = \frac{\pi^2k_B^2}{3}T\hat{K}^{(0)} = \frac{\pi^2}{3}\frac{k_B^2}{e^2}T\hat{\sigma}, \tag{6.125}$$

which is the well-known *Wiedemann-Franz law.* Note, that we can write the thermal conductivity in the form

$$\hat{\kappa} = \frac{C}{e^2N(\epsilon_F)}\hat{\sigma}, \tag{6.126}$$

where $C = \pi^2k_B^2T/3$ denotes the electronic specific heat.

[13] If a function $g(\epsilon)$ depends only weakly on $\epsilon$ in the vicinity of $\epsilon_F$, we can use the Taylor expansion to derive a general approximation for following integrals

$$-\int d\epsilon g(\epsilon)\frac{\partial f_0}{\partial\epsilon} = g(\epsilon_F) + \frac{\pi^2}{6}(k_BT)^2\left.\frac{\partial^2 g(\epsilon)}{\partial\epsilon^2}\right|_{\epsilon=\epsilon_F} + \ldots \tag{6.117}$$

and

$$-\int d\epsilon g(\epsilon)(\epsilon-\epsilon_F)\frac{\partial f_0}{\partial\epsilon} = \frac{\pi^2}{3}(k_BT)^2\left.\frac{\partial g(\epsilon)}{\partial\epsilon}\right|_{\epsilon=\epsilon_F} + \ldots, \tag{6.118}$$

in the limit $T \to 0$. We used that $\mu \to \epsilon_F$ in that asymptotic case.

### 6.7.2 Thermoelectric effect

Equation (6.123) shows, that a temperature gradient induces an electric field. For a simple, isotropic system, this relation reduces to

$$\boldsymbol{\mathcal{E}} = Q\boldsymbol{\nabla}T, \tag{6.127}$$

with the so-called *Seebeck coefficient*

$$Q = -\frac{\pi^2}{3}\frac{k_B^2 T}{e}\frac{\sigma'(\epsilon)}{\sigma(\epsilon)}\bigg|_{\epsilon=\epsilon_F}. \tag{6.128}$$

Using $\sigma(\epsilon) = n(\epsilon)e^2\tau(\epsilon)/m$, we investigate $\sigma'(\epsilon)$,

$$\sigma'(\epsilon) = \frac{\tau'(\epsilon)}{\tau(\epsilon)}\sigma(\epsilon) + \frac{n'(\epsilon)}{n(\epsilon)}\sigma(\epsilon) = \frac{\tau'(\epsilon)}{\tau(\epsilon)}\sigma(\epsilon) + \frac{N(\epsilon)}{n(\epsilon)}\sigma(\epsilon), \tag{6.129}$$

and obtain an additional contribution to $Q$, if the relaxation time depends strongly on energy. This is most prominent in collision processes in which resonant scattering is involved (e.g., the Kondo effect). In the opposite situation, namely, when the first term is irrelevant, the Seebeck coefficient

$$Q = -\frac{\pi^2}{3}\frac{k_B^2 T}{e}\frac{N(\epsilon_F)}{n(\epsilon_F)} = -\frac{S}{ne} \tag{6.130}$$

is simply reduced to the entropy per electron. For simple metals such as the alkali metals we may estimate the low-temperature values using equation (6.130)

$$Q = -\frac{\pi^2}{2}\frac{k_B^2 T}{e\epsilon_F} = -\frac{\pi^2}{2}\frac{k_B T}{eT_F} \tag{6.131}$$

which for $T_F(Na, K) \approx 3 \times 10^4 K$ leads to $Q = -14mVK^{-1} \times T[K]$. A comparison with experiments in Fig.6.7 shows that the order of magnitude works reasonably well for Na and K. However, for Li and Cs even the sign is different. Differences occur through phonon effects, such as the so-called phonon drag which we have neglected here.
In the following, we consider two different types of thermoelectric effects.

#### Seebeck effect

The first is the *Seebeck effect*, where a thermoelectric voltage appears in a bi-metallic system (cf. Figure 6.8). With equation (6.127), a temperature gradient across metal $B$ induces an electromotive force[14]

$$U_{\text{EMF}} = \int d\boldsymbol{l}\cdot\boldsymbol{\mathcal{E}} \tag{6.132}$$

$$= Q_A\int_{T_0}^{T_1} d\boldsymbol{l}\cdot\boldsymbol{\nabla}T + Q_B\int_{T_1}^{T_2} d\boldsymbol{l}\cdot\boldsymbol{\nabla}T + Q_A\int_{T_2}^{T_0} d\boldsymbol{l}\cdot\boldsymbol{\nabla}T \tag{6.133}$$

$$= (Q_B - Q_A)(T_2 - T_1). \tag{6.134}$$

The resulting voltage $V_{\text{therm}} = U_{\text{EMF}}$ appears between the two ends of a second metal $A$, whose contacts are kept at the same temperature $T_0$. Here, a bi-metallic configuration was chosen to reveal voltage differences across the contacts which are absent in a single metal.

[14] The term *electromotive force*, first introduced by Alessandro Volta, is misleading in the sense, that it measures a voltage instead of a force. For further explanations on this ambiguity consult [Neal Graneaum, *In the grip of the distant universe*, World Scientific (2006), page 191, ISBN 9812567542]. This is why we term it $U_{\text{EMF}}$.

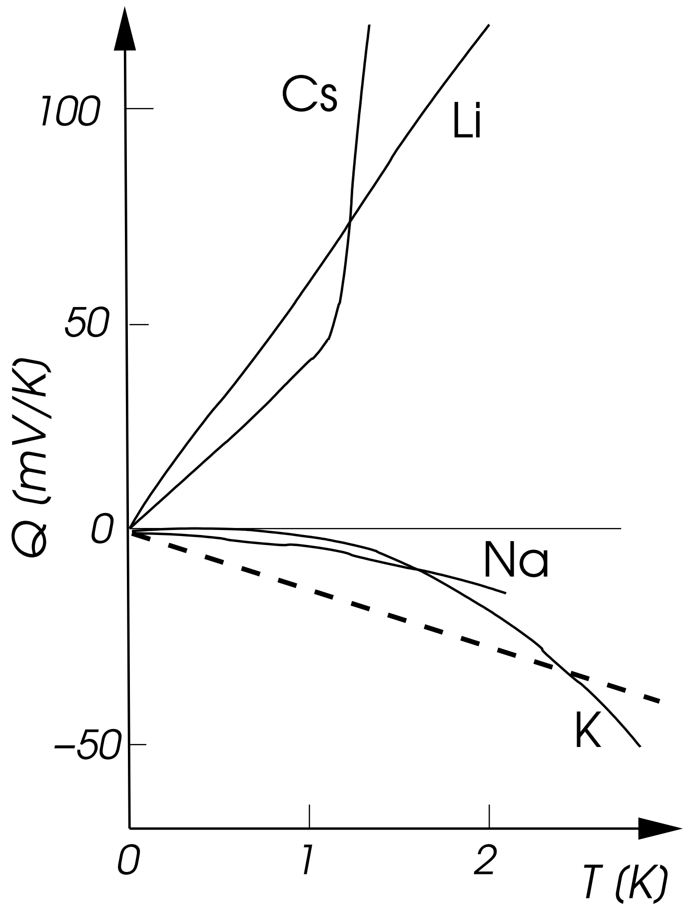



Figure 6.7: Seebeck coefficient for the Alkali metals Li, Na, K, and Cs at low temperatures. The dashed line represents the estimate for Na and K following Eq.(6.131). (adapted from D.K.C. MacDonald, *Thermoelectricity: an introduction to the principles*, Dover (2006).)

**Peltier effect**

The second phenomenon, termed *Peltier effect*, emerges in a system kept at the same temperature everywhere. Here, an electric current $J_e$ between the two contacts of the metal $A$ (see Figure 6.8) induces a heat current in the bi-metallic system, such that heat is transferred from one reservoir (top) to another (bottom). This follows from the equations (6.114) and (6.114) by assuming $\nabla T = 0$, where

$$\begin{cases} J_e = e^2 K^{(0)} E \\ \\ J_q = eK^{(1)} E \end{cases} \tag{6.135}$$

implies

$$J_q = \frac{K^{(1)}}{eK^{(0)}} J_e = \Pi J_e. \tag{6.136}$$

The coupling $\Pi = TQ$ between $J_q$ and $J_e$ is called *Peltier coefficient*. According to Figure 6.8, a contribution to the heat current is to be expected from both metals $A$ and $B$,

$$J_q = (\Pi_A - \Pi_B) J_e = T_0 (Q_A - Q_B) J_e. \tag{6.137}$$

This means, that the heat transfer between reservoirs can be controlled by electrical current.

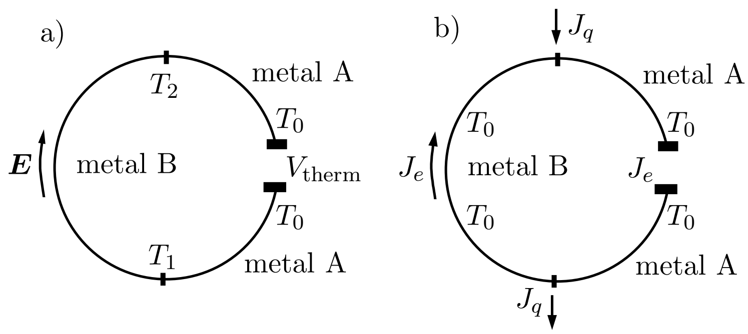



Figure 6.8: Schematics of thermoelectric effects. On the left panel (a) a representation of Seebeck effect is given, where the symbol $E$ is used instead of $\mathcal{E}$. On the right panel (b) the Peltier effect is represented. In our analysis, both systems were effectively one-dimensional.

It has to be emphasized here that the bi-metal design of the devices in Fig. 6.8 serves the observation of the two effects which both represent bulk effects of the two metals A and B. By no means, it should be mistaken as an effect originating from the inter-metal contacts.

## 6.8 Anderson localization in one-dimensional systems

Transport in one spatial dimension is very special, since there are only two different directions to go: forward and backward. We introduce the transfer matrix formalism and use it to express the conductivity through the Landauer formula. We will then investigate the effects of multiple scattering at different obstacles, leading to the so-called *Anderson localization*, which turns a metal into an insulator.

### 6.8.1 Landauer Formula for a single impurity

The transmission and reflection at an arbitrary potential with finite support[15] in one dimension can be described by a transfer matrix $T$.

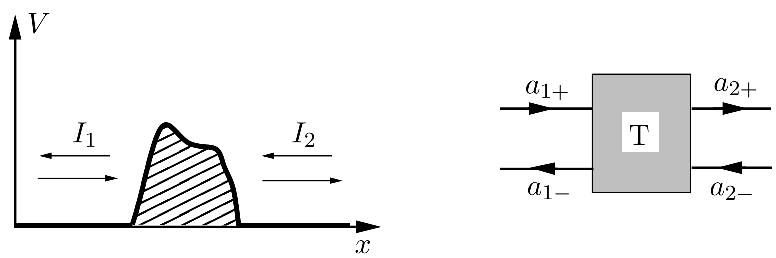



Figure 6.9: Transfer matrices are sufficient to describe potential scattering in one dimension.

In this situation, a suitable choice for a basis of the electron states is the set of plane waves $\{e^{\pm ikx}\}$ (cf. Figure 6.9) moving in the positive (negative) $x$-direction with wave vector $+k$ $(-k)$. Only plane waves with the same $|k|$ on the left $(I_1)$ and right $(I_2)$ side of the scatterer are interconnected. Therefore, we write

$$\psi_1(x) = a_{1+}e^{ikx} + a_{1-}e^{-ikx}, \tag{6.138}$$

$$\psi_2(x) = a_{2+}e^{ikx} + a_{2-}e^{-ikx}, \tag{6.139}$$

[15]The support (dt. Träger) of a function is the set of all points, where the function takes non zero values.

where $\psi_1$ ($\psi_2$) is defined in the area $I_1$ ($I_2$). The vectors $a_i = (a_{i+}, a_{i-})$ $i \in \{1, 2\}$ are connected via the linear relation,

$$a_2 = \hat{T} a_1 = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} a_1, \tag{6.140}$$

with the $2 \times 2$ transfer matrix $\hat{T}$. The conservation of current ($J_1 = J_2$) requires that $\hat{T}$ is unimodular, i.e., $\det T = 1$. Here,

$$J = \frac{i\hbar}{2m} \left( \frac{d\psi^*(x)}{dx} \psi(x) - \psi^*(x) \frac{d\psi(x)}{dx} \right), \tag{6.141}$$

such that, for a plane wave $\psi(x) = (1/\sqrt{L}) e^{ikx}$ in a system of length $L$, the current results in

$$J = v/L \tag{6.142}$$

with the velocity $v$ defined as $v = \hbar k/m$. Time reversal symmetry implies that, simultaneously with $\psi(x)$, the complex conjugate $\psi^*(x)$ is a solution of the stationary Schrödinger equation. From this, we find $T_{11} = T_{22}^*$ and $T_{12} = T_{21}^*$, such that

$$\hat{T} = \begin{pmatrix} T_{11} & T_{12} \\ T_{12}^* & T_{11}^* \end{pmatrix}. \tag{6.143}$$

It is easily shown that a shift of the scattering potential by a distance $x_0$ changes the coefficient $T_{12}$ of $\hat{T}$ by a phase factor $e^{i2kx_0}$. Meanwhile, the coefficient $T_{11}$ remains unchanged.

With the Ansatz for a right moving incoming wave ($\propto e^{ikx}$), producing a reflected ($\propto re^{-ikx}$) and transmitted ($\propto te^{ikx}$) part, the wave functions on both sides of the scatterer read

$$\psi_1(x) = \frac{1}{\sqrt{L}} (e^{ikx} + re^{-ikx}), \tag{6.144}$$

$$\psi_2(x) = \frac{1}{\sqrt{L}} (te^{ikx}). \tag{6.145}$$

The coefficients of $\hat{T}$ can be determined explicitly in this situation, resulting in

$$\hat{T} = \begin{pmatrix} \frac{1}{t^*} & -\frac{r^*}{t^*} \\ -\frac{r}{t} & \frac{1}{t} \end{pmatrix}. \tag{6.146}$$

Here, the conservation of currents imposes the condition $1 = |r|^2 + |t|^2$. Furthermore, we can find a relation between the parameters ($r$, $t$) of the potential barrier and the electric resistivity. For this, we notice that the incoming current density $J_0$ is split into a reflected $J_r$ and transmitted $J_t$ part, all given by

$$J_0 = -\frac{1}{L} ve = -n_0 ve, \tag{6.147}$$

$$J_r = -\frac{|r|^2}{L} ve = -n_r ve, \tag{6.148}$$

$$J_t = -\frac{|t|^2}{L} ve = -n_t ve, \tag{6.149}$$

with the velocity $v = \hbar k/m$, the electron charge $-e$, and the particle densities $n_0$, $n_r$, and $n_t$ corresponding to the incoming, reflected and transmitted particles respectively. The electron density on the two sides of the barrier is given by

$$n_1 = n_0 + n_r = \frac{1 + |r|^2}{L}, \tag{6.150}$$

$$n_2 = n_t = \frac{|t|^2}{L}. \tag{6.151}$$

From this consideration, a density difference $\delta n = n_1 - n_2 = (1+|r|^2 - |t|^2)/L = 2|r|^2/L$ results between the left and the right side of the scatterer. The resistance $R$ of the barrier is defined by the ratio between the voltage drop over the resistor $\delta V$ and the transmitted current $J_t$, i.e.

$$R = \frac{\delta V}{J_t} \tag{6.152}$$

Consequently, a relation between $\delta V$ and the electron density $\delta n$ remains to be established to determine $R$. The connection is easily found via the existing energy difference $\delta E = -e\delta V$ between the two sides of the resistor, such that the expression

$$\delta n = \frac{dn}{dE}\delta E \tag{6.153}$$

$$= \frac{dn}{dE}(-e\delta V) \tag{6.154}$$

produces the wished relation. Here, $\frac{dn}{dE}dE$ is the number of states per unit length in the energy interval $[E, E+dE]$ and we find

$$\frac{dn}{dE} = \frac{1}{L}\sum_{\boldsymbol{k},s}\delta\left(E - \frac{\hbar^2k^2}{2m}\right) = 2\int\frac{dk}{2\pi}\delta\left(E - \frac{\hbar^2k^2}{2m}\right) = \frac{1}{\pi\hbar v(E)}. \tag{6.155}$$

The resistance $R$ is finally obtained from the equations (6.152), (6.153), and (6.155), leading to

$$R = \frac{h}{e^2}\frac{|r|^2}{|t|^2}, \tag{6.156}$$

The *Klitzing constant* $R_K = h/e^2 \approx 25.8k\Omega$ is a resistance quantum named after the discoverer of the Quantum Hall Effect. The result (6.156) is the famous *Landauer formula*, which is valid for all one-dimensional systems and whose application often extends to the description of mesoscopic systems and quantum wires.

### 6.8.2 Scattering at two impurities

We consider now two spatially separated scattering potentials, represented by $\hat{T}_1$ and $\hat{T}_2$ each determined by $r_1, t_1$ and $r_2, t_2$ respectively.

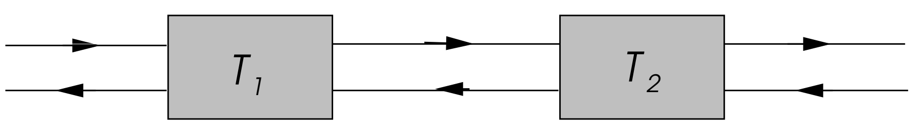



Figure 6.10: Two spatially separated scattering potentials with transmission matrices $\hat{T}_1$ and $\hat{T}_2$ respectively.

The particles are multiply scattered at these potentials in a unknown manner, but the global result can again be expressed via a simple transfer matrix $\hat{T} = \hat{T}_1\hat{T}_2$, given by the matrix multiplication of each transfer matrix. All previously found properties remain valid for the new matrix $\hat{T}$, given by

$$\hat{T} = \begin{pmatrix} \frac{1}{t_1^*t_2^*} + \frac{r_1^*r_2}{t_1^*t_2} & -\frac{r_2^*}{t_1^*t_2^*} - \frac{r_1^*}{t_1^*t_2} \\ -\frac{r_1}{t_1t_2^*} - \frac{r_2}{t_1t_2} & \frac{1}{t_1t_2} + \frac{r_1r_2^*}{t_1t_2^*} \end{pmatrix} = \begin{pmatrix} \frac{1}{t^*} & -\frac{r^*}{t^*} \\ -\frac{r}{t} & \frac{1}{t} \end{pmatrix}. \tag{6.157}$$

For the ratio between reflection and transmission probability we find

$$\frac{|r|^2}{|t|^2} = \frac{1}{|t|^2} - 1 \tag{6.158}$$

$$= \frac{1}{|t_1|^2|t_2|^2}\left|1 + \frac{r_1 r_2^* t_2}{t_2^*}\right|^2 - 1 \tag{6.159}$$

$$= \frac{1}{|t_1|^2|t_2|^2}\left(1 + |r_1|^2|r_2|^2 + \frac{r_1 r_2^* t_2}{t_2^*} + \frac{r_1^* r_2 t_2^*}{t_2}\right) - 1. \tag{6.160}$$

Assuming a (random) distance $d = x_2 - x_1$ between the two potential barriers, we may average over this distance. Note, that for $x_1 = 0$, we find $r_2 \propto e^{-2ikd}$, while $r_1$, $t_1$, and $t_2$ are independent on $d$. Consequently, all terms containing an odd power of $r_2$ or $r_2^*$ vanish after averaging over $d$. The remainders of equation (6.160) can be collected to

$$\left.\frac{|r|^2}{|t|^2}\right|_{\text{avg}} = \frac{1}{|t_1|^2|t_2|^2}\left(1 + |r_1|^2|r_2|^2\right) - 1 \tag{6.161}$$

$$= \frac{|r_1|^2}{|t_1|^2} + \frac{|r_2|^2}{|t_2|^2} + 2\frac{|r_1|^2}{|t_1|^2}\frac{|r_2|^2}{|t_2|^2}. \tag{6.162}$$

Even though two scattering potentials are added in series, an additional non-linear combination emerges beside the sum of the two ratios $|r_i|^2/|t_i|^2$. It results from the Landauer formula applied to two scatterers, that resistances do not add linearly to the total resistance. Adding $R_1$ and $R_2$ serially, the total resistance is not given by $R = R_1 + R_2$, but by

$$R = R_1 + R_2 + 2\frac{R_1 R_2}{R_K} > R_1 + R_2, \tag{6.163}$$

with $R_K = h/e^2$ This result is a consequence of the unavoidable multiple scattering in one dimensions. This effect is particularly prominent if $R_i \gg h/e^2$ for $i \in \{1,2\}$, where resistances are then multiplied instead of summed.

### 6.8.3 Anderson localization

Let us consider a system with many arbitrarily distributed scatterers, and let $\rho$ be a mean resistance per unit length. $R(\ell)$ shall be the resistance between points 0 and $\ell$. The change in resistance by advancing an infinitesimal $\delta\ell$ is found from equation (6.163), resulting in

$$dR = \rho d\ell + 2\frac{R(\ell)}{R_K}\rho d\ell, \tag{6.164}$$

which yields

$$\int_0^{\ell} \rho d\ell = \int_0^{R(\ell)} \frac{dR}{1 + 2R/R_K}, \tag{6.165}$$

and thus,

$$\rho\ell = \frac{h}{2e^2}\ln\left(1 + 2R(\ell)/R_K\right). \tag{6.166}$$

Finally, solving this equation for $R(\ell)$, we find

$$R(\ell) = \frac{R_K}{2}\left(e^{2\rho\ell/R_K} - 1\right). \tag{6.167}$$

Obviously, $R$ grows almost exponentially fast for increasing $\ell$. This means, that for large $\ell$, the system is an insulator for arbitrarily small but finite $\rho > 0$. The reason for this is that, in one dimension, all states are bound states in the presence of disorder. This phenomenon is called *Anderson localization*. Even though all states are localized, the energy spectrum is continuous, as infinitely many bound states with different energy exist. The mean localization length $\xi$ of individual states, related to the mean extension of a wave function, is found from equation (6.167) to be $\xi = \rho/R_K$. The transmission amplitude is reduced[16] on this length scale, since $|t| \approx 2e^{-\ell/\xi}$ for $\ell \gg \xi$. In one dimension, there is no linearly increasing electric resistance, $R(\ell) \approx \rho\ell$. For non-interacting particles, only two extreme situations are possible. Either, the potential is perfectly periodic and the states correspond to Bloch waves. Then, coherent constructive interference produces extended states[17] that propagate freely throughout the system, resulting in a perfect conductor without resistance. On the other hand, if the scattering potential is disordered, all states are strictly localized. In this case, there is no propagation and the system is an insulator. In three-dimensional systems, the effects of multiple scattering are far less drastic and the Ohmic law is applicable. Localization effects in two dimensions is a very subtle topic and part of today's research in solid state physics.

[16] For an expanded discussion of this topic, the article [P.W. Anderson, D.J. Thouless, E. Abrahams, and D.S. Fisher, *New method for a scaling theory of localization*, Physical Review B **22**, 3519 (1980)] is recommended.

[17] We have also seen in the context of chiral edge states in the Quantum Hall state, that perfect conductance in a one-dimensional channel is obtained if there is no backscattering due to the lack of states moving in the opposite direction. In chiral states, particle move only in one direction.

# Chapter 7

# Magnetism in metals

Magnetic ordering in metals can be viewed as an instability of the Fermi liquid state. We enter this new behavior of metals through a detailed description of the Stoner ferromagnetism. The discussion of antiferromagnetism and spin density wave phases will be only brief here. In Stoner ferromagnets the magnetic moment is provided by the spin of itinerant electrons. Magnetism due to localized magnetic moments will be considered in the context of Mott insulators which are subject of the next chapter.

Well-known examples of elemental ferromagnetic metals are iron (Fe), cobalt (Co) and nickel (Ni) belonging to the $3d$ transition metals, where the $3d$-orbital character is dominant for the conduction electrons at the Fermi energy. These orbitals are rather tightly bound to the atomic cores such that the electron mobility is reduced, enhancing the effect of interaction which is essential for the formation of a magnetic state.

Other forms of magnetism, such as antiferromagnetism and the spin density wave state are found in the $3d$ transition metals Cr and Mn. On the other hand, $4d$ and $5d$ transition metals within the same columns of the periodic system are not magnetic. Their $d$-orbitals are more extended, leading to a higher mobility of the electrons, such that the mutual interaction is insufficient to trigger magnetism. It is, however, possible to find ferromagnetism in $ZrZn_2$ where zink (Zn) may act as a spacer reducing the mobility of the $4d$-electrons of zirconium (Zr). The $4d$-elements Pd and Rh and the $5d$-element Pt are, however, nearly ferromagnetic. Going further in the periodic table, the $4f$-orbitals appearing in the lanthanides are nearly localized and can lead to ferromagnetism, as illustrated by the elements going from Gd through Tm in the periodic system.

Magnetism appears through a phase transition, meaning that the metal is non-magnetic at temperatures above a critical temperature $T_c$, the Curie-temperature (cf. Table 7.1). In many cases, magnetism appears at $T_c$ as a continuous, second order phase transition involving the spontaneous violation of symmetry. This transition is lacking latent heat (no discontinuity in entropy and volume) but instead features a discontinuity in the specific heat.

| element | $T_c$ (K) | type | element | $T_c$ (K) | type |
|---|---|---|---|---|---|
| Fe | 1043 | ferromagnet (3d) | Gd | 293 | ferromagnet (4f) |
| Co | 1388 | ferromagnet (3d) | Dy | 85 | ferromagnet (4f) |
| Ni | 627 | ferromagnet (3d) | Cr | 312 | spin density wave (3d) |
| $ZrZn_2$ | 22 | ferromagnet | $\alpha$-Mn | 100 | antiferromagnet |
| Pd | – | paramagnet | Pt | – | paramagnet |
| $HfZn_2$ | – | paramagnet | | | |

Table 7.1: Selection of (ferro)magnetic materials with their respective form of magnetism and the critical temperature $T_c$.

## 7.1 Stoner instability

In the following section, we study the emergence of the metallic ferromagnetism originating from the Stoner mechanism. In close analogy to the first Hund's rule, the exchange interaction among the electrons plays a crucial role here. The alignment of the electronic spins in a favored direction allows the system to reduce the energy contribution due to Coulomb repulsion. According to Landau's theory of Fermi liquids, the interaction between electrons renormalizes the spin susceptibility $\chi_0$ to

$$\chi = \frac{m^*}{m}\frac{\chi_0}{1+F_0^a}, \tag{7.1}$$

which obviously diverges for $F_0^a \to -1$ and leads to a correlation driven instability of the Fermi liquid, as we will discuss here.

### 7.1.1 Stoner model within the mean field approximation

Consider a model of conduction electrons with a repulsive contact interaction,

$$\mathcal{H} = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s} + U\int d^3r\, d^3r'\, \widehat{\rho}_\uparrow(\boldsymbol{r})\delta(\boldsymbol{r}-\boldsymbol{r}')\widehat{\rho}_\downarrow(\boldsymbol{r}'), \tag{7.2}$$

where we use the electron density $\widehat{\rho}_s(\boldsymbol{r}) = \widehat{\Psi}^\dagger_s(\boldsymbol{r})\widehat{\Psi}_s(\boldsymbol{r})$ and the field operator $\widehat{\Psi}^\dagger_s$ ($\widehat{\Psi}_s$) follows from the definition (3.12) [(3.13)]. Due to the Pauli exclusion principle, the contact interaction is only active between electrons with opposite spins. This is a consequence of the exchange hole in the two-particle correlation between electrons of identical spin (cf. Figure 3.1). The derivation of the full solution of this model goes beyond the scope of this lecture. However, a mean field approximation will already provide very useful insights [1]. We rewrite,

$$\widehat{\rho}_s(\boldsymbol{r}) = n_s + [\widehat{\rho}_s(\boldsymbol{r}) - n_s], \tag{7.3}$$

where

$$n_s = \langle \widehat{\rho}_s(\boldsymbol{r})\rangle, \tag{7.4}$$

and $\langle\dot{}\rangle$ represents the expectation value of the argument. We stipulate that the deviation from the mean value $n_s$ shall be small. In other words

$$\langle [\widehat{\rho}_s(\boldsymbol{r}) - n_s]^2\rangle \ll n_s^2. \tag{7.5}$$

Inserting equation (7.3) into the Hamiltonian (7.2), we obtain

$$\mathcal{H}_{\rm mf} = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s} + U\int d^3r\, [\widehat{\rho}_\uparrow(\boldsymbol{r})n_\downarrow + \widehat{\rho}_\downarrow(\boldsymbol{r})n_\uparrow - n_\uparrow n_\downarrow] \tag{7.6}$$

$$= \sum_{\boldsymbol{k},s} \left(\epsilon_{\boldsymbol{k}} + Un_{-s}\right)\widehat{c}^\dagger_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s} - U\Omega n_\uparrow n_\downarrow, \tag{7.7}$$

the so-called *mean field Hamiltonian*, describing electrons which move in the uniform background of electrons of opposite spin coupling via the spin dependent exchange interaction. Fluctuations are ignored here. The advantage of this approximation is, that the many-body problem is now reduced to an effective one-particle problem, where only the mean electron interaction is taken

[1] Note that the following mean field calculation is equivalent to a variational approach using simple many-body wavefunction (Slater determinant) with different concentrations of up and down spins.

into account. This is equivalent to a generalized Hartree-Fock approximation and enables us to calculate certain expectation values, such as the density of one spin species

$$n_\uparrow = \frac{1}{\Omega}\sum_{\boldsymbol{k}}\langle \hat{c}^\dagger_{\boldsymbol{k}\uparrow}\hat{c}_{\boldsymbol{k}\uparrow}\rangle = \frac{1}{\Omega}\sum_{\boldsymbol{k}} f(\epsilon_{\boldsymbol{k}} + Un_\downarrow) \tag{7.8}$$

$$= \int d\epsilon\, \frac{1}{\Omega}\sum_{\boldsymbol{k}} \delta(\epsilon - \epsilon_{\boldsymbol{k}} - Un_\downarrow) f(\epsilon) \tag{7.9}$$

$$= \int d\epsilon\, \frac{1}{2} N(\epsilon - Un_\downarrow) f(\epsilon). \tag{7.10}$$

An analogous result is found for the opposite spin direction. These mean densities are determined self-consistently, namely such that the insertion of $n_s$ in into the mean field Hamiltonian (7.7) provides the correct output according to the expectation values given in equation (7.8). Furthermore, the constraint that the total number of electrons is conserved, must be implemented. The real magnetization $M = \mu_B m$ is proportional to $m$ defined via

$$n_s = \frac{1}{2}\left[(n_\uparrow + n_\downarrow) + s(n_\uparrow - n_\downarrow)\right] = \frac{n_0 + sm}{2}, \tag{7.11}$$

where $n_0$ is the total particle density. This leads to the two coupled equations

$$n_0 = \frac{1}{2}\int d\epsilon\, \Big(N(\epsilon - Un_\downarrow) + N(\epsilon - Un_\uparrow)\Big) f(\epsilon), \tag{7.12}$$

$$m = \frac{1}{2}\int d\epsilon\, \Big(N(\epsilon - Un_\downarrow) - N(\epsilon - Un_\uparrow)\Big) f(\epsilon), \tag{7.13}$$

or equivalently

$$n_0 = \frac{1}{2}\sum_s \int d\epsilon N\left(\epsilon - \frac{Un_0}{2} - s\frac{Um}{2}\right) f(\epsilon), \tag{7.14}$$

$$m = -\frac{1}{2}\sum_s s \int d\epsilon N\left(\epsilon - \frac{Un_0}{2} - s\frac{Um}{2}\right) f(\epsilon), \tag{7.15}$$

which usually can not be solved analytically and must be treated numerically.

### 7.1.2 Stoner criterion

An approximate solution can be found if $m \ll n_0$. In this limit, the equations (7.14) and (7.15) are solved by adapting the chemical potential $\mu$. For low temperatures and small magnetization we can expand $\mu$ as

$$\mu(m, T) = \epsilon_F + \Delta\mu(m, T). \tag{7.16}$$

The constant energy shift $-Un_0/2$ appearing in the equations (7.14) and (7.15) has been absorbed into $\epsilon_F$. The Fermi-Dirac distribution takes the form

$$f(\epsilon) = \frac{1}{e^{\beta[\epsilon - \mu(m,T)]} + 1}, \tag{7.17}$$

where $\beta = (k_B T)^{-1}$. After expanding equation (7.14) for small $m$, one obtains

$$n_0 \approx \int d\epsilon f(\epsilon)\left[N(\epsilon) + \frac{1}{2}\left(\frac{Um}{2}\right)^2 N''(\epsilon)\right] \tag{7.18}$$

$$\approx \int_0^{\epsilon_F} d\epsilon\big[N(\epsilon)\big] + N(\epsilon_F)\Delta\mu + \frac{\pi^2}{6}(k_B T)^2 N'(\epsilon_F) + \frac{1}{2}\left(\frac{Um}{2}\right)^2 N'(\epsilon_F), \tag{7.19}$$

where we introduced the abbreviations $N'(\epsilon) = dN(\epsilon)/d\epsilon$ and $N''(\epsilon) = d^2N(\epsilon)/d\epsilon^2$. Since the first term on the right side of equation (7.19) is identical to $n_0$, $\Delta\mu(m,T)$ is immediately found to be given by

$$\Delta\mu(m,T) \approx -\frac{N'(\epsilon_F)}{N(\epsilon_F)}\left[\frac{\pi^2}{6}(k_BT)^2 + \frac{1}{2}\left(\frac{Um}{2}\right)^2\right]. \tag{7.20}$$

Analogously, the expansion of equation (7.15) in $m$ and $T$, results in

$$m \approx \int d\epsilon f(\epsilon)\left[N'(\epsilon)\frac{Um}{2} + \frac{1}{3!}N'''(\epsilon)\left(\frac{Um}{2}\right)^3\right] \tag{7.21}$$

$$\approx \left[N(\epsilon_F) + \frac{\pi^2}{6}(k_BT)^2N''(\epsilon_F) + \frac{1}{3!}\left(\frac{Um}{2}\right)^2 N''(\epsilon_F) + \Delta\mu N'(\epsilon_F)\right]\left(\frac{Um}{2}\right), \tag{7.22}$$

and finally, inserting the result for $\Delta\mu$ from (7.20) into (7.22), we find

$$m = N(\epsilon_F)\left[1 - \frac{\pi^2}{6}(k_BT)^2\Lambda_1^2(\epsilon_F)\right]\left(\frac{Um}{2}\right) - N(\epsilon_F)\Lambda_2^2(\epsilon_F)\left(\frac{Um}{2}\right)^3, \tag{7.23}$$

where

$$\Lambda_1^2(\epsilon_F) = \left(\frac{N'(\epsilon_F)}{N(\epsilon_F)}\right)^2 - \frac{N''(\epsilon_F)}{N(\epsilon_F)}, \quad \text{and} \quad \Lambda_2^2(\epsilon_F) = \frac{1}{2}\left(\frac{N'(\epsilon_F)}{N(\epsilon_F)}\right)^2 - \frac{N''(\epsilon_F)}{3!N(\epsilon_F)}. \tag{7.24}$$

The structure of equation (7.23) is $m = am + bm^3$, where $b$ is assumed to be negative. Thus, two types of solutions emerge

$$m^2 = \begin{cases} 0, & a < 1, \\ \dfrac{1-a}{b}, & a \geq 1. \end{cases} \tag{7.25}$$

With this, $a = 1$ corresponds to a critical value.

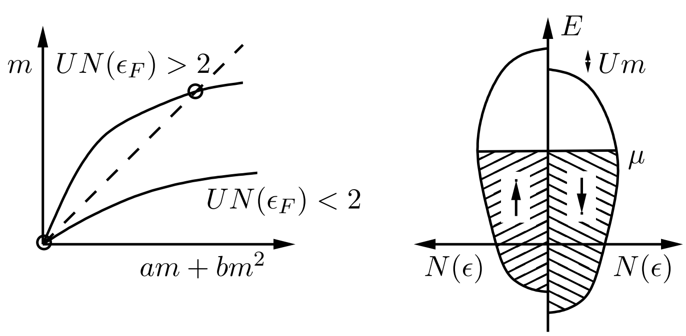



Figure 7.1: Graphical solution of equation (7.23) and the resulting magnetization. The Fermi sea of each spin configuration is shifted by $\pm Um/2$, resulting in a finite total magnetization.

In our situation, this condition corresponds to

$$1 = \frac{1}{2}UN(\epsilon_F)\left[1 - \frac{\pi^2}{6}(k_BT_C)^2\Lambda_1^2(\epsilon_F)\right], \tag{7.26}$$

yielding

$$k_BT_C = \frac{\sqrt{6}}{\pi\Lambda_1(\epsilon_F)}\sqrt{1 - \frac{2}{UN(\epsilon_F)}} \propto \sqrt{1 - \frac{U_c}{U}} \tag{7.27}$$

for $U > U_c = 2/N(\epsilon_F)$. This is an instability condition for the nonmagnetic Fermi liquid state with $m = 0$, and $T_C$ is the *Curie temperature*, below which the ferromagnetic state appears (see Figure 7.1). The temperature dependence of the magnetization $M$ of the ferromagnetic state ($T < T_C$) is given by

$$M(T) = \mu_B m(T) \propto \sqrt{T_C - T}, \tag{7.28}$$

close to the phase transition ($T_C - T \ll T_C$). Note that the Curie temperature $T_C$ is nonzero for $UN(\epsilon_F) > 2$, and $T_C \to 0$ in the limit $UN(\epsilon_F) \to 2_+$. For $UN(\epsilon_F) < 2$ no phase transition occurs. This condition for a finite transition temperature $T_C$ is known as the *Stoner criterion*. This simple model also describes a so-called *quantum phase transition*, i.e., a phase transition that appears at $T = 0$ as a function of system parameters, which in our case are the density of states $N(\epsilon_F)$ and the Coulomb repulsion $U$. While thermal fluctuations destroy the ordered state at some finite temperature via entropy increase, entropy is irrelevant at $T = 0$. The order is now suppressed by quantum fluctuations (Heisenberg's uncertainty principle).

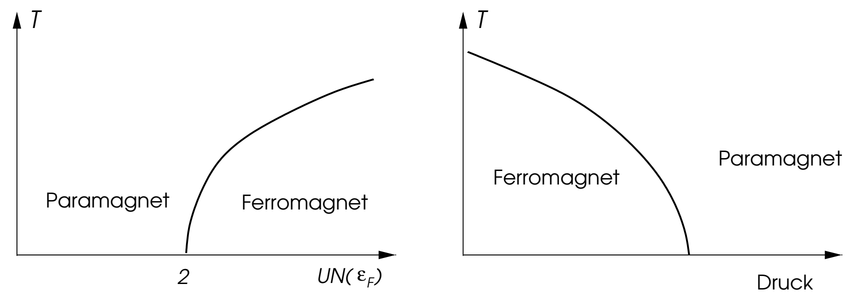



Figure 7.2: Phase diagram of a Stoner ferromagnet in the $T$-$UN(\epsilon_F)$ and $T$-$p$ plane, respectively.

The density of states as an internal parameter can, for example be changed by applying an external pressure. By reducing the lattice constant, pressure may facilitate the motion of the conduction electrons and increase the Fermi velocity. Consequently, the density of states is reduced (cf. Figure 7.2). Indeed, pressure is able to destroy ferromagnetism in weakly ferromagnetic materials as $ZrZn_2$, MnSi, and $UGe_2$. In other materials, the Curie temperature is high enough, such that the technologically applicable pressure is insufficient to suppress magnetism. It is, however, possible, that pressure leads to other transitions, such as structural phase transitions, that eventually destroy magnetism. This is seen in iron (Fe), where a pressure of about 12 GPa induces a transition from magnetic iron with body-centered crystal (bcc) structure to a nonmagnetic, hexagonal close packed (hcp) structure (cf. Figure 7.3).
While this structural transition is a quantum phase transition as well, it appear as a discontin-

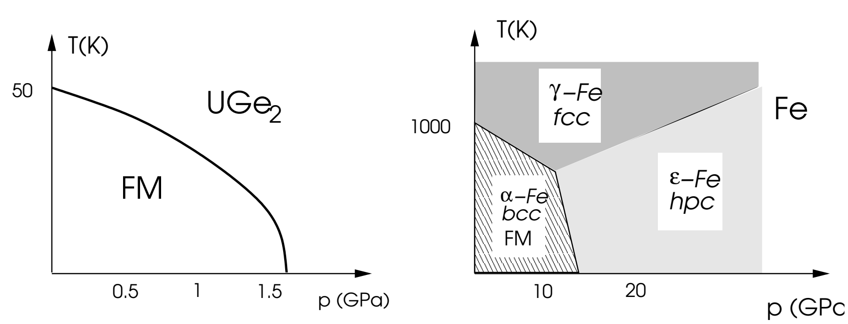



Figure 7.3: Phase diagrams of $UGe_2$ and Fe.

uous, first order[2] transition. In some cases, pressure can also induce an increase in $N(\epsilon_F)$, for example in metals with multiple bands, where compression leads to a redistribution of charge. One example is the ruthenate $Sr_3Ru_2O_7$ for which uniaxial pressure along the $z$-axis leads to magnetism.

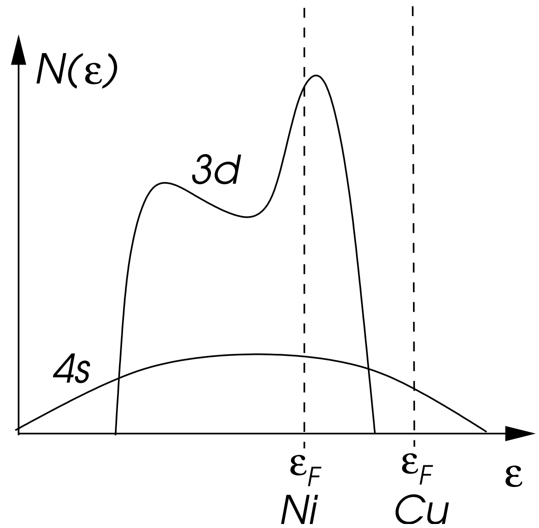



Figure 7.4: The position of the Fermi energy of Cu and Ni, respectively.

Let us eventually turn to the question, why Cu, being a direct neighbor of Ni in the 3d-row of the periodic table, is not ferromagnetic, even though both elemental metals share the same fcc crystal structure. The answer is given by the Stoner instability criterion $UN(\epsilon_F) = 2$. While the conduction electrons at the Fermi level of Ni have 3d-character and belong to a narrow band with a large density of states, the Fermi energy of Cu is situated in the broad 4s-band and constitutes a much smaller density of states (cf. Figure 7.4). With this, the Cu conduction electrons are much less localized and feature a weaker tendency towards ferromagnetic order. Copper is known to be a better conductor than Ni for the same reason.

### 7.1.3 Spin susceptibility for $T > T_C$

Next, we study the response of a metallic system to an external perturbation. For this purpose, we apply an infinitesimal magnetic field $H$ along the $z$-axis, which induces a spin polarization due to the Zeeman coupling. From the self-consistency equations (7.14) and (7.15) we obtain

$$m = -\frac{1}{2}\int d\epsilon f(\epsilon)\sum_s s\, N\left(\epsilon - \mu_B sH - s\frac{Um}{2}\right) \tag{7.29}$$

$$\approx \int d\epsilon f(\epsilon) N'(\epsilon)\left(\frac{Um}{2} + \mu_B H\right) \tag{7.30}$$

$$= N(\epsilon_F)\left[1 - \frac{\pi^2}{6}(k_B T)^2 \Lambda_1(\epsilon_F)^2\right]\left(\frac{Um}{2} + \mu_B H\right) \tag{7.31}$$

to lowest order in $m$ and $H$. Solving this equation for $m$ yields

$$M = \mu_B m = \frac{\chi_0(T)}{1 - U\chi_0(T)/2\mu_B^2} H, \tag{7.32}$$

[2] The Stoner instability is a simplification of the quantum phase transition. In most cases, a discontinuous phase transition originates in the band structure or in fluctuation effects, which were ignored here, For more details consult [D. Belitz and T.R. Kirkpatrick, Phys. Rev. Lett. **89**, 247202 (2002)].

and consequently, the magnetic susceptibility $\chi$ reads

$$\chi = \frac{M}{H} = \frac{\chi_0(T)}{1 - U\chi_0(T)/2\mu_B^2}, \tag{7.33}$$

where the bare susceptibility $\chi_0$ is given by

$$\chi_0(T) = \mu_B^2 N(\epsilon_F)\left[1 - \frac{\pi^2}{6}(k_B T)^2 \Lambda_1(\epsilon_F)^2\right]. \tag{7.34}$$

We see, that the denominator of the susceptibility $\chi(T)$ vanishes exactly when the Stoner instability criterion is fulfilled. The diverging susceptibility

$$\chi(T) \approx \frac{\chi_0(T_C)}{\frac{T_C^2}{T^2} - 1}, \tag{7.35}$$

indicates the instability. Note that for $T \to T_C$ from above, the susceptibility diverges like $\chi(T) \propto |T_C - T|^{-1}$ corresponding to the mean field behavior, since the mean field critical exponent $\gamma$ for the susceptibility takes the value $\gamma = 1$.

## 7.2 General spin susceptibility and magnetic instabilities

The ferromagnetic state is characterized by a uniform magnetization. There are, however, magnetically ordered states which do not feature a nonzero net magnetization. Examples are spin density wave (SDW) states, antiferromagnets and spin spiral states. In this section, we analyze general instability conditions for metallic systems.

### 7.2.1 General dynamic spin susceptibility

We consider a magnetic field, oscillating in time and with spacial modulation

$$\boldsymbol{H}(\boldsymbol{r}, t) = \boldsymbol{H}_0 e^{i\boldsymbol{q}\cdot\boldsymbol{r} - i\omega t} e^{\eta t}, \tag{7.36}$$

and calculate the resulting magnetization, for the corresponding Fourier component. For that, we proceed analogously as in Chapter 3 and define the spin density operator $\widehat{\boldsymbol{S}}(\boldsymbol{r})$ in real space,

$$\widehat{\boldsymbol{S}}(\boldsymbol{r}) = \frac{\hbar}{2}\sum_{s,s'} \widehat{\Psi}_s^\dagger(\boldsymbol{r})\boldsymbol{\sigma}_{ss'}\widehat{\Psi}_{s'}(\boldsymbol{r}) = \frac{\hbar}{2}\begin{pmatrix} \widehat{\Psi}_\uparrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\downarrow(\boldsymbol{r}) + \widehat{\Psi}_\downarrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\uparrow(\boldsymbol{r}) \\ -i\widehat{\Psi}_\uparrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\downarrow(\boldsymbol{r}) + i\widehat{\Psi}_\downarrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\uparrow(\boldsymbol{r}) \\ \widehat{\Psi}_\uparrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\uparrow(\boldsymbol{r}) - \widehat{\Psi}_\downarrow^\dagger(\boldsymbol{r})\widehat{\Psi}_\downarrow(\boldsymbol{r}) \end{pmatrix} \tag{7.37}$$

with momentum space representation

$$\widehat{\boldsymbol{S}}_{\boldsymbol{q}} = \int d^3r \widehat{\boldsymbol{S}}(\boldsymbol{r}) e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} = \frac{\hbar}{2\Omega}\sum_{\boldsymbol{k},s,s'} c_{\boldsymbol{k},s}^\dagger \boldsymbol{\sigma}_{ss'} c_{\boldsymbol{k}+\boldsymbol{q},s'} = \frac{1}{\Omega}\sum_{\boldsymbol{k}} \widehat{\boldsymbol{S}}_{\boldsymbol{k},\boldsymbol{q}}, \tag{7.38}$$

where $\widehat{\boldsymbol{S}}_{\boldsymbol{k},\boldsymbol{q}} = (\hbar/2) c_{\boldsymbol{k}+\boldsymbol{q},s}^\dagger \boldsymbol{\sigma}_{ss'} c_{\boldsymbol{k},s'}$. The Hamiltonian of the electronic system with contact interaction is given by

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_Z + \mathcal{H}_{\text{int}}, \tag{7.39}$$

where

$$\mathcal{H}_0 = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}_{\boldsymbol{k}s}^\dagger \widehat{c}_{\boldsymbol{k}s}, \tag{7.40}$$

$$\mathcal{H}_Z = -\frac{g\mu_B}{\hbar}\int d^3r \boldsymbol{H}(\boldsymbol{r}, t)\cdot\widehat{\boldsymbol{S}}(\boldsymbol{r}), \tag{7.41}$$

$$\mathcal{H}_{\text{int}} = U\int d^3r \widehat{\rho}_\uparrow(\boldsymbol{r})\widehat{\rho}_\downarrow(\boldsymbol{r}). \tag{7.42}$$

The operator $\mathcal{H}_Z$ describes the *Zeeman coupling* between the electrons of the metal and the perturbing field. We investigate a magnetic field

$$\boldsymbol{H} = \frac{1}{2} H^+(\boldsymbol{q},\omega) e^{-i\omega t} e^{\eta t} \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix} \tag{7.43}$$

in the $xy$-plane. The Zeeman term then simplifies to

$$\mathcal{H}_Z = -\frac{g\mu_B}{\hbar\Omega} \sum_{\boldsymbol{k}} H^+(\boldsymbol{q},\omega) \widehat{S}^-_{\boldsymbol{k},-\boldsymbol{q}} e^{i\boldsymbol{q}\cdot\boldsymbol{r}-i\omega t+\eta t} + \text{h.c.}, \tag{7.44}$$

where the Hermitian conjugate (h.c.) is ignored in the following. We see immediately that only $\widehat{S}^-_{\boldsymbol{k},-\boldsymbol{q}}$ couples to the field $H^+(\boldsymbol{q},\omega)$. In the framework of linear response theory, this coupling will induce a magnetization $M^+_{\text{ind}} = (\mu_B/\hbar)\langle S^+_{\text{ind}}(\boldsymbol{q},\omega)\rangle$. Using the same equations of motion as in Section 3.2,

$$i\hbar \frac{\partial}{\partial t} \widehat{S}^+_{\boldsymbol{k},\boldsymbol{q}} = [\widehat{S}^+_{\boldsymbol{k},\boldsymbol{q}}, \mathcal{H}], \tag{7.45}$$

we can determine this induced magnetization, first without the interaction term (U=0). We obtain for the given Fourier component,

$$i\hbar \frac{\partial}{\partial t} \widehat{S}^+_{\boldsymbol{k},\boldsymbol{q}}(t)_{\boldsymbol{k},\boldsymbol{q}} = (\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}}) \widehat{S}^+_{\boldsymbol{k},\boldsymbol{q}}(t) + g\hbar\mu_B (\widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q}\uparrow} \widehat{c}_{\boldsymbol{k}+\boldsymbol{q}\uparrow} - \widehat{c}^\dagger_{\boldsymbol{k}\downarrow} \widehat{c}_{\boldsymbol{k}\downarrow}) H^+(\boldsymbol{q},\omega) e^{-i\omega t}. \tag{7.46}$$

Performing the Fourier transform from frequency space to real time, and applying the thermal average we obtain,

$$(\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega + i\hbar\eta) \langle S^+_{\boldsymbol{k},\boldsymbol{q}} \rangle = -g\hbar\mu_B (n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} - n_{\boldsymbol{k}\downarrow}) H^+(\boldsymbol{q},\omega), \tag{7.47}$$

which then leads to the induced spin density ($\propto$ magnetization),

$$\langle S^+_{\text{ind}}(\boldsymbol{q},\omega)\rangle = \frac{1}{\Omega} \sum_{\boldsymbol{k}} \langle S^+_{\boldsymbol{k},\boldsymbol{q}} \rangle = \frac{\hbar}{\mu_B} \chi_0(\boldsymbol{q},\omega) H^+(\boldsymbol{q},\omega), \tag{7.48}$$

with

$$\chi_0(\boldsymbol{q},\omega) = -\frac{g\mu_B^2}{\Omega} \sum_{\boldsymbol{k}} \frac{n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} - n_{\boldsymbol{k}\downarrow}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega + i\hbar\eta}. \tag{7.49}$$

Note that the form of the bare susceptibility $\chi_0(\boldsymbol{q},\omega)$ is similar to the Lindhard function (3.48). The result (7.48) describes the induced spin density within linear response approximation.
In the next step, we want included the effects of the interaction. Analogously to the charge density wave state found in Section 3.2, the induced spin density generates in turn an effective field. The induced spin polarization appears in the exchange interaction as an effective magnetic field. We rewrite the contact interaction term in equation (7.39) in the form

$$\begin{aligned} \mathcal{H}_{\text{int}} &= \frac{U}{\Omega} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}'} \widehat{c}^\dagger_{\boldsymbol{k}+\boldsymbol{q}'\uparrow} \widehat{c}_{\boldsymbol{k}\uparrow} \widehat{c}^\dagger_{\boldsymbol{k}'-\boldsymbol{q}'\downarrow} \widehat{c}_{\boldsymbol{k}'\downarrow} \\ &= -\frac{U}{\Omega} \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{q}'} \widehat{c}^\dagger_{\boldsymbol{k}\uparrow} \widehat{c}_{\boldsymbol{k}+\boldsymbol{q}'\downarrow} \widehat{c}^\dagger_{\boldsymbol{k}'\downarrow} \widehat{c}_{\boldsymbol{k}'-\boldsymbol{q}'\uparrow} + const. \\ &= -\frac{U}{\Omega\hbar^2} \sum_{\boldsymbol{q}'} \widehat{S}^+_{\boldsymbol{q}'} \widehat{S}^-_{-\boldsymbol{q}'}. \end{aligned} \tag{7.50}$$

The induced spin polarization $\langle S^+_{\text{ind}}(\boldsymbol{q},\omega)\rangle$ acts through the exchange interaction as an effective (local) field, as can be seen by replacing $\widehat{S}^+_{\boldsymbol{q}'} \to \langle S^+_{\text{ind}}(\boldsymbol{q},\omega)\rangle\delta_{\boldsymbol{q},\boldsymbol{q}'}$ in Eq.(7.51) ,

$$-\frac{U}{\Omega\hbar^2}\sum_{\boldsymbol{q}'}\widehat{S}^+_{\boldsymbol{q}'}\widehat{S}^-_{-\boldsymbol{q}'} \to -\frac{U}{\hbar^2}\langle S^+(\boldsymbol{q},\omega)\rangle\widehat{S}^-_{-\boldsymbol{q}} = -\frac{g\mu_B}{\hbar}H^+_{\text{ind}}(\boldsymbol{q},\omega)\widehat{S}^-_{-\boldsymbol{q}} \tag{7.51}$$

where the effective magnetic field $H^+_{\text{ind}}(\boldsymbol{q},\omega)$ finally reads

$$H^+_{\text{ind}}(\boldsymbol{q},\omega) = \frac{U}{g\mu_B\hbar}\langle S^+(\boldsymbol{q},\omega)\rangle. \tag{7.52}$$

This induced field acts on the spins as well, such that the total response of the spin density on the external field becomes

$$\begin{aligned} M^+(\boldsymbol{q},\omega) &= \frac{\mu_B}{\hbar}\langle S^+(\boldsymbol{q},\omega)\rangle \\ &= \chi_0(\boldsymbol{q},\omega)[H^+(\boldsymbol{q},\omega) + H^+_{\text{ind}}(\boldsymbol{q},\omega)] \\ &= \chi_0(\boldsymbol{q},\omega)H^+(\boldsymbol{q},\omega) + \chi_0(\boldsymbol{q},\omega)\frac{U}{g\mu_B\hbar}\langle S^+(\boldsymbol{q},\omega)\rangle \\ &= \chi_0(\boldsymbol{q},\omega)H^+(\boldsymbol{q},\omega) + \chi_0(\boldsymbol{q},\omega)\frac{U}{g\mu_B^2}M^+(\boldsymbol{q},\omega). \end{aligned} \tag{7.53}$$

With the definition

$$M^+(\boldsymbol{q},\omega) = \chi(\boldsymbol{q},\omega)H^+(\boldsymbol{q},\omega) \tag{7.54}$$

of the susceptibility in the random phase approximation (RPA) (see Section 3.2), we find

$$\chi(\boldsymbol{q},\omega) = \frac{\chi_0(\boldsymbol{q},\omega)}{1 - \frac{U}{2\mu_B^2}\chi_0(\boldsymbol{q},\omega)}. \tag{7.55}$$

This form of the susceptibility is found to be valid for all field directions, as long as spin-orbit coupling is neglected and the spin is isotropic (Heisenberg model). Within the random phase approximation, the generalization of the Stoner criterion for the appearance of an instability of the system at finite temperature reads

$$1 = \frac{U}{2\mu_B^2}\chi_0(\boldsymbol{q},\omega). \tag{7.56}$$

For the limiting case $(\boldsymbol{q},\omega) \to (\boldsymbol{0},0)$ corresponding to a uniform, static external field, we obtain for the bare susceptibility

$$\begin{aligned} \chi_0(\boldsymbol{q},0) &= -\frac{2\mu_B^2}{\Omega}\sum_{\boldsymbol{k}}\frac{n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} - n_{\boldsymbol{k}\downarrow}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}}} \\ &\xrightarrow{\boldsymbol{q}\to 0} -\frac{2\mu_B^2}{\Omega}\sum_{\boldsymbol{k}}\frac{\partial f(\epsilon_{\boldsymbol{k}})}{\partial\epsilon_{\boldsymbol{k}}} = \chi_0(T), \end{aligned} \tag{7.57}$$

which corresponds to the Pauli susceptibility ($g = 2$). Then, $\chi(T)$ from equation (7.58) is again cast into the form (7.33) and describes the instability of the metal with respect to ferromagnetic spin polarization, when the denominator vanishes. Similar to the charge density wave, the isotropic deformation for $\boldsymbol{q} = \boldsymbol{0}$ is not the leading instability, when $\chi_0(\boldsymbol{Q},0) > \chi_0(\boldsymbol{0},0)$ for a finite $\boldsymbol{Q}$. Then, another form of magnetic order will occur.

### 7.2.2 Instability with finite wave vector $\boldsymbol{Q}$

In order to show that, indeed, the Stoner instability does not always prevail among all possible magnetic instabilities, we first go through a simple argument based on the local susceptibility. For that, we define the local magnetic moment along the $z$-axis, $M(\boldsymbol{r}) = \mu_B \langle \widehat{\rho}_\uparrow(\boldsymbol{r}) - \widehat{\rho}_\downarrow(\boldsymbol{r})\rangle$, and consider the nonlocal relation

$$M(\boldsymbol{r}) = \int d^3r' \, \tilde{\chi}_0(\boldsymbol{r}-\boldsymbol{r}')H_z(\boldsymbol{r}'), \tag{7.58}$$

within the linear response approximation. In Fourier space, the same relation reads

$$M_{\boldsymbol{q}} = \chi_0(\boldsymbol{q})H_{\boldsymbol{q}}, \tag{7.59}$$

with

$$\chi_0(\boldsymbol{q}) = \int d^3r \, \tilde{\chi}_0(\boldsymbol{r})e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}. \tag{7.60}$$

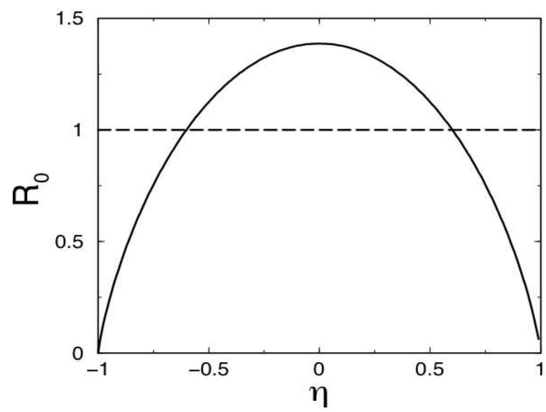



Figure 7.5: $R_0$, the ratio between the local susceptibility and the static susceptibility, plotted for a box-shaped band with width $2D$. Depending where the Fermi energy lies $\eta = \epsilon_F/D$, the susceptibility is dominated by the contribution $\chi_0(\boldsymbol{q}=0)$ or by the susceptibility at finite $\boldsymbol{q}$.

Now, compare $\chi_0(\boldsymbol{q}=0)$ with $\overline{\chi(\boldsymbol{q})} = \tilde{\chi}_0(\boldsymbol{r}=0)$, i.e., the uniform and the local susceptibility at $T=0$. The local susceptibility appears to be the average of $\chi_0(\boldsymbol{q})$ over all $\boldsymbol{q}$,

$$\overline{\chi_0(\boldsymbol{q})} = \frac{2\mu_B^2}{\Omega^2}\sum_{\boldsymbol{k},\boldsymbol{q}} \frac{n_{\boldsymbol{k}+\boldsymbol{q}} - n_{\boldsymbol{k}}}{\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}}} = \frac{\mu_B^2}{2}\int d\epsilon N(\epsilon)\int d\epsilon' N(\epsilon')\frac{f(\epsilon)-f(\epsilon')}{\epsilon'-\epsilon}, \tag{7.61}$$

and must be compared to $\chi_0(\boldsymbol{q}=0) = \mu_B^2 N(\epsilon_F)$ $(f(\epsilon) = \Theta(\epsilon_F - \epsilon))$. The local susceptibility depends on the density of states and the Fermi energy of the system. A very good qualitative understanding can be obtained by a very simple form

$$N(\epsilon) = \begin{cases} \dfrac{1}{D}, & -D \le \epsilon \le D, \\ \\ 0, & |\epsilon| > D, \end{cases} \tag{7.62}$$

for the density of states. $N(\epsilon)$ forms a band with the shape of a box with band width $2D$. With this rough approximation, the integral in equation (7.65) is easily evaluated. The ratio between $\overline{\chi(\boldsymbol{q})}$ and $\chi_0(\boldsymbol{q}=0)$ is then found to be

$$R_0 = \frac{\overline{\chi_0(\boldsymbol{q})}}{\chi_0(\boldsymbol{q}=0)} = \ln\left(\frac{4}{1-\eta^2}\right) + \eta\ln\left(\frac{1-\eta}{1+\eta}\right), \tag{7.63}$$

with $\eta = \epsilon_F/D$. For both small and large band fillings ($\epsilon_F$ close to the band edges), the tendency towards ferromagnetism dominates (cf. Figure 7.5), whereas when $\epsilon_F$ tends towards the middle of the band, the susceptibility $\chi_0(\boldsymbol{q})$ will cease to be maximal at $\boldsymbol{q} = 0$, and magnetic ordering with a well-defined finite $\boldsymbol{q} = \boldsymbol{Q}$ becomes more probable.

### 7.2.3 Influence of the band structure

Whether magnetic order arises at finite $\boldsymbol{q}$ or not depends strongly on the details of the band structure. The argument given above, comparing the local ($\boldsymbol{r} = 0$) to the uniform ($\boldsymbol{q} = 0$) susceptibility is nothing more than a vague indicator for a possible instability at nonzero $\boldsymbol{q}$. A crucial ingredient for the appearance of magnetic order at a given $\boldsymbol{q} = \boldsymbol{Q}$ is the so-called *nesting* of the Fermi surface, i.e. within extended areas in close proximity to the Fermi surface the energy dispersion satisfies the nesting condition,

$$\xi_{\boldsymbol{k}+\boldsymbol{Q}} = -\xi_{\boldsymbol{k}} \tag{7.64}$$

where $\xi_{\boldsymbol{k}} = \epsilon_{\boldsymbol{k}} - \epsilon_F$ and $\boldsymbol{Q}$ is some fixed vector. If the Fermi surface of a material features such a nesting trait, the susceptibility will be dominated by the contribution from this vector $\boldsymbol{Q}$. In order to see this, let us investigate the static susceptibility $\chi_0(\boldsymbol{q})$ for $\boldsymbol{q} = \boldsymbol{Q}$ under the assumption, that equation (7.68) holds for all $\boldsymbol{k}$. Thus,

$$\chi_0(\boldsymbol{Q};T) = \frac{2\mu_B^2}{\Omega}\sum_{\boldsymbol{k}}\frac{n_{\boldsymbol{k}+\boldsymbol{Q}} - n_{\boldsymbol{k}}}{\xi_{\boldsymbol{k}} - \xi_{\boldsymbol{k}+\boldsymbol{Q}}} = \mu_B^2\int\frac{d^3k}{(2\pi)^3}\frac{\tilde{f}(-\xi_{\boldsymbol{k}}) - \tilde{f}(\xi_{\boldsymbol{k}})}{\xi_{\boldsymbol{k}}}, \tag{7.65}$$

where $\tilde{f}(\xi) = f(\xi + \epsilon_F) = f(\epsilon)$ and $f$ is the Fermi-Dirac distribution. Under the further assumption that $\xi_{\boldsymbol{k}}$ is weakly angle dependent, we find

$$\chi_0(\boldsymbol{Q};T) = \mu_B^2\int\frac{d^3k}{(2\pi)^3}\frac{\tanh(\xi_{\boldsymbol{k}}/2k_BT)}{\xi_{\boldsymbol{k}}} = \frac{\mu_B^2}{2}\int d\xi N(\xi + \epsilon_F)\frac{\tanh(\xi/2k_BT)}{\xi}. \tag{7.66}$$

In order to approximate this integral properly, we notice that the integral only converges if a cutoff $\epsilon_0$ is introduced which we take at half the bandwidth $D$, i.e. $\epsilon_0 \sim D$. Furthermore, the leading contribution comes from the immediate vicinity of the Fermi energy ($\xi \approx 0$), so that, within the integration range, a constant density of states $N(\epsilon_F)$ can be assumed. Finally, the susceptibility reads

$$\chi_0(\boldsymbol{Q};T) \approx -\mu_B^2 N(\epsilon_F)\int_0^{\epsilon_0} d\xi\frac{\tanh(\xi/2k_BT)}{\xi} \tag{7.67}$$

$$= \mu_B^2 N(\epsilon_F)\left(\ln\left(\frac{\epsilon_0}{2k_BT}\right) + \ln\left(\frac{4e^\gamma}{\pi}\right)\right) \approx \mu_B^2 N(\epsilon_F)\ln\left(\frac{1.14\epsilon_0}{2k_BT}\right), \tag{7.68}$$

where we assumed $\epsilon_0 \gg k_BT$, and where $\gamma \approx 0.57721$ is the Euler constant. The bare susceptibility $\chi_0$ diverges logarithmically at low temperatures. Inserting the result (7.72) into the generalized Stoner relation (7.59), results in

$$0 = 1 - \frac{UN(\epsilon_F)}{2}\ln\left(\frac{1.14\epsilon_0}{2k_BT_c}\right), \tag{7.69}$$

with the critical temperature

$$k_BT_c = 1.14\epsilon_0 e^{-2/UN(\epsilon_F)}. \tag{7.70}$$

A finite critical temperature persists for arbitrarily small positive values of $UN(\epsilon_F)$. The nesting condition for a given $\boldsymbol{Q}$ leads to a maximum of $\chi_0(\boldsymbol{q}, 0; T)$ at $\boldsymbol{q} = \boldsymbol{Q}$ and triggers the relevant instability in the system. The latter finally stabilizes in a magnetic ordered phase with wave vector $\boldsymbol{Q}$, the so-called *spin density wave*. The spin density distribution takes, for example, the form

$$\boldsymbol{S}(\boldsymbol{r}) = \hat{z} S \cos(\boldsymbol{Q} \cdot \boldsymbol{r}), \tag{7.71}$$

without a uniform component. In comparison, the charge density wave was a modulation of the charge density with a much smaller amplitude than the height of the uniform density, i.e.,

$$\rho(\boldsymbol{r}) = \rho_0 + \delta\rho \cos(\boldsymbol{Q} \cdot \boldsymbol{r}), \tag{7.72}$$

with $\delta\rho \ll \rho_0$. The spin density state frequently appear in low-dimensional systems like organic conductors, or in transition metals such as chrome (Cr) for example. In all cases, nesting plays an important role (cf. Figure 7.6).

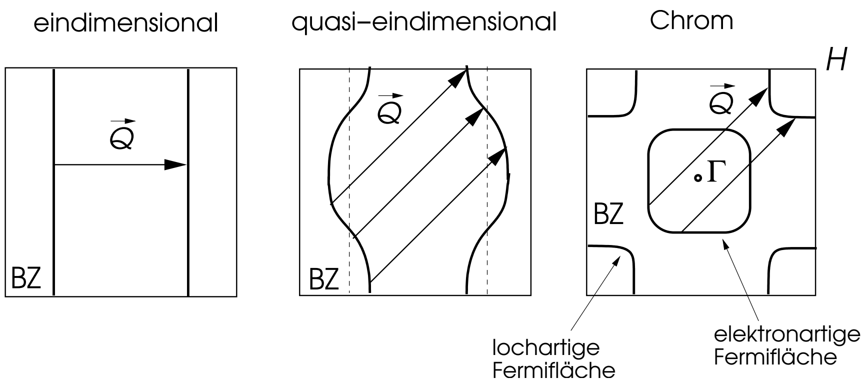



Figure 7.6: Sketch of Fermi surfaces favorable for nesting. In purely one-dimensional systems (left panel) there is a well-defined nesting vector pointing from one end of the Fermi surface to the other one. In quasi-one-dimensional systems nesting is almost perfect (central panel). In special cases (e.g. Cr) the Fermi surface(s) of three-dimensional systems show nesting properties (right panel) promoting an instability of the susceptibility at finite $\boldsymbol{q}$.

In quasi-one-dimensional electron systems, a main direction of motion dominates over two other directions with weak dispersion. In this case, the nesting condition is very probable to be fulfilled, as it is schematically shown in the center panel of Figure 7.6. Chromium is a three-dimensional metal, where nesting occurs between a electron-like Fermi surface around the Γ-point and a hole-like Fermi surface at the Brillouin zone boundary ($H$-point). These Fermi surfaces originate in different bands (right panel in Figure 7.6). Chromium has a cubic body centered crystal structure, where the $H$-point at $(\pi/a, 0, 0)$ leads to the nesting vector $\boldsymbol{Q}_x \parallel (1,0,0)$ and equivalent vectors in $y$- and $z$-direction, which are incommensurable with the lattice.

The textbook example of nesting is found in a tight-binding model of a simple cubic lattice with nearest-neighbor hopping at half filling. The band structure is given by

$$\epsilon_{\boldsymbol{k}} = -2t[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)], \tag{7.73}$$

where $a$ is the lattice constant and $t$ the hopping term. Because of half filling, the chemical potential $\mu$ lies at $\mu = 0$. Obviously, $\epsilon_{\boldsymbol{k}+\boldsymbol{Q}} = -\epsilon_{\boldsymbol{k}}$ holds for all $\boldsymbol{k}$, for the nesting vector $\boldsymbol{Q} = (\pi/a)(1,1,1)$. This full nesting trait is a signature of the total particle-hole symmetry. Analogously to the Peierls instability, the spin density wave induces the opening of a gap at the Fermi surface. This is another example of a Fermi surface instability. In this situation, the gap is confined to the areas of the Fermi surface obeying the nesting condition. Contrarily to the ferromagnetic order, the material can become insulating when forming the spin density wave state.

## 7.3 Stoner excitations

In this last section, we discuss the elementary excitations of the ferromagnetic ground state, including both particle-hole excitations and collective modes. We focus on spin excitations, for which we make the Ansatz

$$|\psi_{\boldsymbol{q}}\rangle = \sum_{\boldsymbol{k}} f_{\boldsymbol{k}} \widehat{c}^{\dagger}_{\boldsymbol{k}+\boldsymbol{q},\downarrow} \widehat{c}_{\boldsymbol{k}\uparrow} |\psi_g\rangle. \tag{7.74}$$

In this excitation, an electron is extracted from the ground state $|\psi_g\rangle$ and is replaced by an electron with opposite spin. We limit ourselves to excitations with a fixed momentum transfer $\boldsymbol{q}$. A selection factor $n_{\boldsymbol{k}\uparrow}(1 - n_{\boldsymbol{k}+\boldsymbol{q},\downarrow})$ ensures that an electron with $(\boldsymbol{k}, \downarrow)$ is available to be removed, and that the state $(\boldsymbol{k}+\boldsymbol{q}, \uparrow)$ is unoccupied. We solve the Schrödinger equation

$$\mathcal{H}|\psi_{\boldsymbol{q}}\rangle = (E_g + \hbar\omega_{\boldsymbol{q}})|\psi_{\boldsymbol{q}}\rangle. \tag{7.75}$$

After some calculations, the eigenvalue condition takes the form

$$\frac{1}{U} = \frac{1}{\Omega} \sum_{\boldsymbol{k}} \frac{n_{\boldsymbol{k}\downarrow} - n_{\boldsymbol{k}+\boldsymbol{q}\uparrow}}{\hbar\omega_{\boldsymbol{q}} - \epsilon_{\boldsymbol{k}+\boldsymbol{q}\downarrow} + \epsilon_{\boldsymbol{k}\uparrow}}, \tag{7.76}$$

corresponding to a root of the generalized RPA Stoner criterion (7.59). During the derivation of equation (7.80), one notices that a subset of the eigenvalues corresponds to the continuum of electron-hole excitations with the spectrum

$$\hbar\omega_{\boldsymbol{q}} = \epsilon_{\boldsymbol{k}+\boldsymbol{q},\downarrow} - \epsilon_{\boldsymbol{k}\uparrow} = \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} + U(n_\uparrow - n_\downarrow), \tag{7.77}$$

where we use the definition $\epsilon_{\boldsymbol{k}s} = \epsilon_{\boldsymbol{k}} + Un_{-s}$. In addition to these single particle excitations, collective modes exist. Similar to the excitons, these modes can be interpreted as a bound state of an electron and a hole. In the limit $\boldsymbol{q} \to 0$, the equation (7.80) simplifies to

$$\frac{1}{U} = \frac{n_\downarrow - n_\uparrow}{\hbar\omega_0 - U(n_\uparrow - n_\downarrow)}, \tag{7.78}$$

meaning that $\hbar\omega_0 = 0$ is a particular solution which we will interpret later. Before, we expand the right-hand side of equation (7.80) for $\hbar\omega - \epsilon_{\boldsymbol{k}+\boldsymbol{q}} + \epsilon_{\boldsymbol{k}} \ll \Delta = U(n_\uparrow - n_\downarrow)$,

$$\frac{1}{U} = \frac{1}{\Delta\Omega} \sum_{\boldsymbol{k}} \frac{n_{\boldsymbol{k}+\boldsymbol{q}\uparrow} - n_{\boldsymbol{k}\downarrow}}{1 - \frac{\hbar\omega}{\Delta} + \frac{1}{\Delta}(\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}})}, \tag{7.79}$$

using $\epsilon_{\boldsymbol{k}+\boldsymbol{q},\downarrow} - \epsilon_{\boldsymbol{k}\uparrow} = \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} + \Delta$. Together with equation (7.82), we find

$$0 = \frac{U}{\Delta} \sum_{\boldsymbol{k}} (n_{\boldsymbol{k}+\boldsymbol{q}\downarrow} - n_{\boldsymbol{k}\uparrow}) \left[ \left(\hbar\omega_{\boldsymbol{q}} + \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}}\right) - \frac{1}{\Delta}\left(\hbar\omega_{\boldsymbol{q}} + \epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}}\right)^2 \right] + \dots \tag{7.80}$$

$$\approx \hbar\omega_{\boldsymbol{q}} - \frac{U}{\Delta} \sum_{\boldsymbol{k}} \frac{n_{\boldsymbol{k}\uparrow} + n_{\boldsymbol{k}\downarrow}}{2} (\boldsymbol{q}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}})^2 \epsilon_{\boldsymbol{k}} - \frac{U}{\Delta^2} \sum_{\boldsymbol{k}} (n_{\boldsymbol{k}\downarrow} - n_{\boldsymbol{k}\uparrow})(\boldsymbol{q}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\,\epsilon_{\boldsymbol{k}})^2 + O(\boldsymbol{q}^4) \tag{7.81}$$

up to second order in $\boldsymbol{q}$. The assumption of a simple parabolic form for the band energies $\epsilon_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m^*$, and a weak magnetization $n_\uparrow - n_\downarrow \ll n_0$ allows the explicit evaluation of the condition (7.85). Few calculation steps lead to

$$\hbar\omega_{\boldsymbol{q}} = \frac{\hbar^2 q^2}{2m^*} \frac{1}{\Delta} \left( Un_0 - \frac{4\epsilon_F}{3} \right) = vq^2. \tag{7.82}$$

Note, that $v$ is positive, since the instability criterion for this case reads $U_c = 4\epsilon_F/3n_0 = 2/N(\epsilon_F)$ and

$$\frac{\hbar\omega_{\boldsymbol{q}}}{q^2} = \frac{\hbar^2}{2m^*}\frac{n_0}{3\sqrt{3}}\sqrt{\frac{1}{-}U_c U} \geq 0. \tag{7.83}$$

This collective excitation features a $q^2$-dependent dispersion, vanishing in the limit $\boldsymbol{q} \to 0$. This effect is a consequence of the ferromagnetic state breaking a continuous symmetry. The rotation symmetry is broken by the choice of a given direction of magnetization. A uniform $\boldsymbol{q} = 0$ rotation of the magnetization does not cost any energy $\omega_0 = 0$. This result was already found in equation (7.82) and is predicted by the so-called *Goldstone theorem*.[3] Such an infinitesimal rotation is induced by a global spin rotation,

$$\sum_{\boldsymbol{k}} \hat{c}^\dagger_{\boldsymbol{k}\downarrow}\hat{c}_{\boldsymbol{k}\uparrow} = \hat{S}^-_{tot}. \tag{7.84}$$

Since the elementary excitations have an energy gap of the order of $\Delta$ at small $\boldsymbol{q}$, the collective excitations, which are termed *magnons*, are well-defined quasiparticles describing propagating spin waves. When these modes enter the particle-hole continuum, they are damped in the same way as plasmons (cf. Figure 7.7). Being a bound state composed of an electron and a hole, magnons are bosonic quasiparticles.

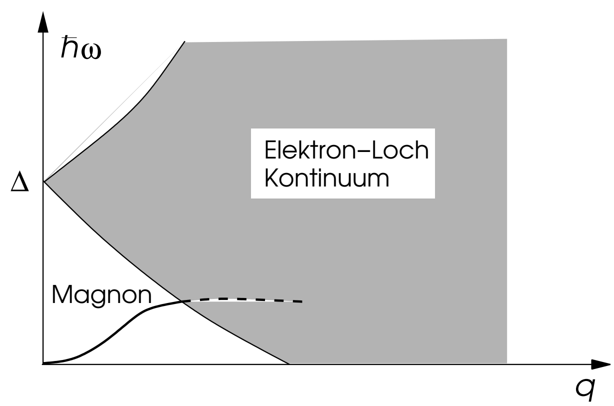



Figure 7.7: Elementary particle-hole excitation spectrum (light gray) and collective modes (magnons, solid line) of the Stoner ferromagnet.

[3]The Goldstone theorem states that, in a system with a short-ranged interaction, a phase which is reached by the breaking of a continuous symmetry features a collective excitation with arbitrarily small energy, so-called *Goldstone modes*. These modes have bosonic character. In the case of the Stoner ferromagnet, these modes are the magnons or spin waves.

# Chapter 8

# Magnetism of localized moments

Up to now, we have mostly assumed that the interaction between electrons leads to secondary effects. This was, essentially, the message of the Fermi liquid theory, the standard model of condensed matter physics. There, the interactions of course renormalize the properties of a metal, but their description is still possible by using a language of nearly independent fermionic quasiparticles with a few modifications. Even in connection with the magnetism of itinerant electrons, where interactions proved to be crucial, the description in terms of extended Bloch states. Many properties were determined by the band structure of the electrons in the lattice, i.e., the electrons were preferably described in $k$-space.
However, in this chapter, we will consider situations, were it is less clear wether we should describe the electrons in momentum or in real space. The problem becomes obvious with the following Gedanken experiment: We look at a regular lattice of H-atoms. The lattice constant should be large enough such that the atoms can be considered to be independent for now. In the ground state, each H-atom contains exactly one electron in the 1$s$-state, which is the only atomic orbital we consider at the moment. The transfer of one electron to another atom would cost the relatively high energy of $E(H^+)+E(H^-)-2E(H) \sim 15$eV, since it corresponds to an ionization. Therefore, the electrons remain localized on the individual H-atoms and the description of the electron states is obviously best done in real space. The reduction of the lattice constant will gradually increase the overlap of the electron wave functions of neighboring atoms. In analogy to the $H_2$ molecule, the electrons can now extend on neighboring atoms, but the cost in energy remains that of an “ionization”. Thus, transfer processes are only possible virtually, there are not yet itinerant electrons in the sense of a metal.

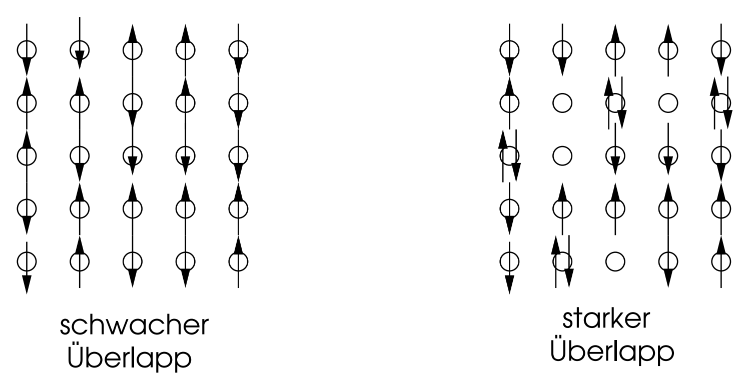



Figure 8.1: Possible states of the electrons in a lattice with weak or strong overlap of the electron wave functions, respectively.

On the other hand, we know the example of the alkali metals, which release their outermost ns-electron into an extended Bloch state and build a metallic (half-filled) band. This would actually

work well for the H-atoms for sufficiently small lattice constant too.[1] Obviously, a transition between the two limiting behaviors should exist. This metal-insulator transition, which occurs, if the gain of kinetic energy surpasses the energy costs for the charge transfer. The insulating side is known as a Mott insulator.
While the obviously metallic state is reliably described by the band picture and can be sufficiently well approximated by the previously discussed methods, this point of view becomes obsolete when approaching the metal-insulator transition. According to band theory, a half-filled band must produce a metal, which definitely turns wrong when entering the insulating side of the transition. Unfortunately, no well controlled approximation for the description of this metal-insulator transition exists, since there are no small parameters for a perturbation theory.
Another important aspect is the fact, that in a standard Mott insulator each atom features an electron in the outermost occupied orbital and, hence, a degree of freedom in the form of a localized spin $s = 1/2$, in the simplest case. While charge degrees of freedom (motion of electrons) are frozen at small temperatures, the same does not apply to these spin degrees of freedom. Many interesting magnetic phenomena are produced by the coupling of these spins. Other, more general forms of Mott insulators exist as well, which include more complex forms of localized degrees of freedom, e.g., partially occupied degenerate orbital states.

## 8.1 Mott transition

First, we investigate the metal-insulator transition. Its description is difficult, since it does not constitute a transition between an ordered and a disordered state in the usual sense. We will, however, use some simple considerations which will allow us to gain some insight into the behavior of such systems.

### 8.1.1 Hubbard model

We introduce a model, which is based on the tight-binding approximation we have introduced in chapter 1. It is inevitable to go back to a description based on a lattice and give up continuity. The model describes the motion of electrons, if their wave functions on neighboring lattice sites only weakly overlap. Furthermore, the Coulomb repulsion, leading to an increase in energy, if a site is doubly occupied, is taken into account. We include this with the lattice analogue of the contact interaction. The model, called *Hubbard model*, has the form

$$\mathcal{H} = -t \sum_{\langle i,j\rangle,s} (\widehat{c}^\dagger_{is}\widehat{c}_{js} + \text{h.c.}) + U \sum_i \widehat{n}_{i\uparrow}\widehat{n}_{i\downarrow}, \tag{8.1}$$

where we consider hopping between nearest neighbors only, via the matrix element $-t$. Note, that $\widehat{c}^{(\dagger)}_{is}$ are real-space field operators on the lattice (site index $i$) and $\widehat{n}_{is} = \widehat{c}^\dagger_{is}\widehat{c}_{is}$ is the density operator. We focus on half filling, $n = 1$, one electron per site on average. There are two obvious limiting cases:

- *Insulating atomic limit:* We put $t = 0$. The ground state has exactly one electron on each lattice site. This state is, however, highly degenerate. In fact, the degeneracy is $2^N$ (number of sites $N$), since each electron has spin 1/2, i.e.,

$$|\Phi_{A0}\{s_i\}\rangle = \prod_i \widehat{c}^\dagger_{i,s_i}|0\rangle, \tag{8.2}$$

  where the spin configuration $\{s_i\}$ can be chosen arbitrarily. We will deal with the lifting of this degeneracy later. The first excited states feature one lattice site without electron

[1]In nature, this can only be induced by enormous pressures metallic hydrogen probably exists in the centers of the large gas planets Jupiter and Saturn due to the gravitational pressure.

and one doubly occupied site. This state has energy $U$ and its degeneracy is even higher, i.e., $2^{N-2}N(N-1)$. Even higher excited states correspond to more empty and doubly occupied sites. The system is an insulator and the density of states is shown in Figure 8.2.

- *Metallic band limit:* We set $U = 0$. The electrons are independent and move freely via hopping processes. The band energy is found through a Fourier transform of the Hamiltonian. With

$$\widehat{c}_{is} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \widehat{c}_{\boldsymbol{k}s} e^{i\boldsymbol{k}\cdot\boldsymbol{r}_i}, \tag{8.3}$$

we can rewrite

$$-t \sum_{\langle i,j\rangle,s} (\widehat{c}^{\dagger}_{is}\widehat{c}_{js} + h.c.) = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \widehat{c}^{\dagger}_{\boldsymbol{k}s}\widehat{c}_{\boldsymbol{k}s}, \tag{8.4}$$

where

$$\epsilon_{\boldsymbol{k}} = -t \sum_{\boldsymbol{a}} e^{i\boldsymbol{k}\cdot\boldsymbol{a}} = -2t\left(\cos k_x a + \cos k_y a + \cos k_z a\right), \tag{8.5}$$

and the sum runs over all vectors $\boldsymbol{a}$ connecting nearest neighbors. The density of states is also shown in Figure 8.2. Obviously, this system is metallic, with a unique ground state

$$|\Phi_{B0}\rangle = \prod_{\boldsymbol{k}} \Theta(-\epsilon_{\boldsymbol{k}}) \widehat{c}^{\dagger}_{\boldsymbol{k}\uparrow}\widehat{c}^{\dagger}_{\boldsymbol{k}\downarrow}|0\rangle. \tag{8.6}$$

Note, that $\epsilon_F = 0$ at half filling, whereas the bandwidth $2D = 12t$.

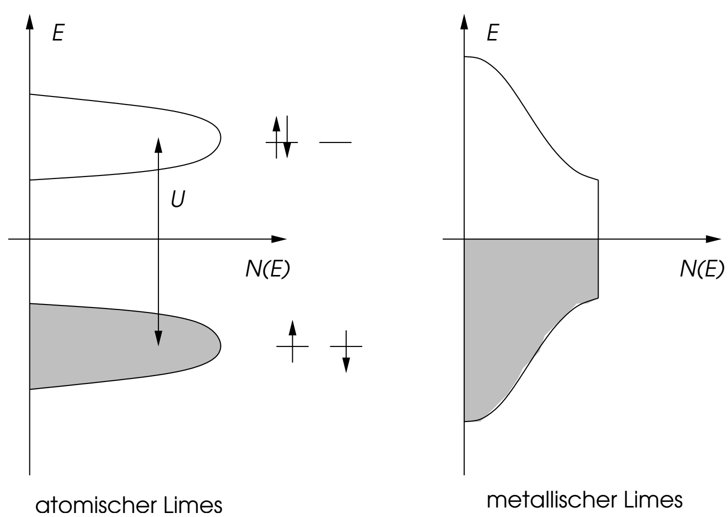



Figure 8.2: Density of states of the Hubbard model in the atomic limit (left) and in the free limit (right).

### 8.1.2 Insulating state

We consider the two lowest energy sectors for the case $t \ll U$. The ground state sector $\alpha$ has already been defined: one electrons sits on each lattice site. The lowest excited states create the sector $\beta$ with one empty and one doubly occupied site (cf. Figure 8.3). With the finite hopping matrix element, the empty (holon) and the doubly occupied (doublon) site become "mobile". A fraction of the degeneracy ($2^{N-2}N(N-1)$) is herewith lifted and the energy obtains a momentum dependence,

$$E_{\boldsymbol{k},\boldsymbol{k}'} = U + \epsilon_{\boldsymbol{k}} + \epsilon_{\boldsymbol{k}'} > U - 12t. \tag{8.7}$$

Even though ignoring the spin configurations here is a daring approximation, we obtain a qualitatively good picture of the situation.[2] One notices that, with increasing $|t|$, the two energy sectors approach each other, until they finally overlap. In the left panel Figure8.2 the holon-doublon excitation spectrum is depicted by two bands, the lower and upper Hubbard bands, where the holon is a hole in the lower and the doublon a particle in the upper Hubbard band. The excitation gap is the gap between the two bands and we may interpret this system as an insulator, called a Mott insulator. (Note, however, that this band structure depends strongly on the correlation effects (e.g. spin correlation) and is not rigid as the band structure of a semiconductor.) The band overlap (closing of the gap) indicates a transition, after which a perturbative treatment is definitely inapplicable. This is, in fact, the metal-insulator transition.

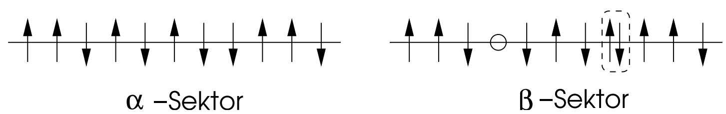



Figure 8.3: Illustration of the two energy sectors, $\alpha$ and $\beta$.

### 8.1.3 The metallic state

On the metallic side, the initial state is better defined since the ground state is a filled Fermi sea without degeneracy. The treatment of the Coulomb repulsion $U$ turns out to become difficult, once we approach the Mott transition, where the electrons suffer a strong impediment in their mobility. In this region, there is no straight-forward way of a perturbative treatment. The so-called Gutzwiller approximation, however, provides a qualitative and very instructive insights into the properties of the strongly correlated electrons.
For this approximation we introduce the following important densities:

- 1: electron density
- $s_\uparrow$: density of the singly occupied lattice sites with spin $\uparrow$
- $s_\downarrow$: density of the singly occupied lattice sites with spin $\downarrow$
- $d$: density of the doubly occupied sites
- $h$: density of the empty sites

It is easily seen, that $h = d$ and $s_\uparrow = s_\downarrow = s/2$, as long as no uniform magnetization is present. Note, that $d$ determines the energy contribution of the interaction term to $Ud$, which we regard as the index of fixed interaction energy sectors. Furthermore,

$$1 = s + 2d \tag{8.8}$$

holds. The view point of the Gutzwiller approximation is based on the renormalization of the probability of the hopping process due to the correlation of the electrons,exceeding restrictions

[2] Note that the motion of an empty site (holon) or doubly occupied site (doublon) is not independent of the spin configuration which is altered through moving these objects. As a consequence, the holon/doublon motion is not entirely free leading to a reduction of the band width. Therefore the band width seen in Figure8.2 (left panel) is smaller than 2D, in general. The motion of a single hole was in detail discussed by Brinkman and Rice (Phys. Rev. B 2, 1324 (1970).

due to the Pauli principle. With this, the importance of the spatial configuration of the electrons is enhanced. In the Gutzwiller approximation, the latter is taken into account statistically by simple probabilities for the occupation of lattice sites.
We fix the density of the doubly occupied sites $d$ and investigate the hopping processes which keep $d$ constant. First, we consider an electron hopping from a singly occupied to an empty site ($i \to j$). Hopping probability depends on the availability of the initial configuration. We compare the probability to find this initial state for the correlated ($P$) and the uncorrelated ($P_0$) case and write

$$P(\uparrow 0) + P(\downarrow 0) = g_t[P_0(\uparrow 0) + P_0(\downarrow 0)]. \tag{8.9}$$

The factor $g_t$ will eventually appear as the renormalization of the hopping probability and, thus, leads to an effective kinetic energy of the system due to correlations. We determine both sides statistically. In the correlated case, the joint probability for $i$ to be singly occupied and $j$ to be empty is obviously

$$P(\uparrow 0) + P(\downarrow 0) = sh = sd = d(1-2d). \tag{8.10}$$

where we used equation (8.8). In the uncorrelated case (where $d$ is not fixed), we have

$$P_0(\uparrow 0) = n_{i\uparrow}(1-n_{i\downarrow})(1-n_{j\uparrow})(1-n_{j\downarrow}) = \frac{1}{16}. \tag{8.11}$$

The case for $\downarrow$ follows accordingly. In order to collect the total result for hopping processes which keep $d$ constant, we have to do the same calculation for the hopping process $(\uparrow\downarrow, \uparrow) \to (\uparrow, \uparrow\downarrow)$, which leads to the same result. Processes of the kind $(\uparrow\downarrow, 0) \to (\uparrow, \downarrow)$ leave the sector of fixed $d$ and are ignored.[3] With this, we obtain in all cases the same renormalization factor for the kinetic energy,

$$g_t = 8d(1-2d), \tag{8.12}$$

i.e., $t \to g_t t$. We consider the correlations by treating the electrons as independent but with a renormalized matrix element $g_t t$. The energy in the sector $d$ becomes

$$E(d) = g_t \epsilon_{\text{kin}} + Ud = 8d(1-2d)\epsilon_{\text{kin}} + Ud, \qquad \epsilon_{\text{kin}} = \frac{1}{N}\int_{-D}^{0} d\epsilon\, N(\epsilon)\epsilon. \tag{8.13}$$

For fixed $U$ and $t$, we can minimize this with respect to $d$ (note that this in not a variational calculation in a strict sense, the resulting energy is not an upper bound to the ground state energy), and find

$$d = \frac{1}{4}\left(1 - \frac{U}{U_c}\right) \qquad \text{and} \qquad g_t = 1 - \left(\frac{U}{U_c}\right)^2, \tag{8.14}$$

with the critical value

$$U_c = 8|\epsilon_{\text{kin}}| \approx 25t \sim 4D. \tag{8.15}$$

For $u \geq U_c$, double occupancy and, thus, hopping is completely suppressed, i.e., electrons become localized. This observation by Brinkman and Rice [Phys. Rev. B **2**, 4302 (1970)] provides a qualitative description of the metal-insulator transition to a Mott insulator, but

[3] This formulation is based on plausible arguments. A more rigorous derivation can be found in the literature, e.g., in D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984); T. Ogawa et al., Prog. Theor. Phys. **53**, 614 (1975); S. Huber, *Gutzwiller-Approximation to the Hubbard-Model* (Proseminar SS02, http://www.itp.phys.ethz.ch/proseminar/condmat02).

takes into account only local correlations, while correlations between different lattice sites are not considered. Moreover, correlations between the spin degrees of freedom are entirely neglected. The charge excitations contain contributions between different energy scales: (1) a metallic part, described via the renormalized effective Hamiltonian

$$\mathcal{H}_{\text{ren}} = \sum_{\boldsymbol{k},s} g_t \epsilon_{\boldsymbol{k}} \widehat{c}^{\dagger}_{\boldsymbol{k}s} \widehat{c}_{\boldsymbol{k}s} + Ud, \tag{8.16}$$

and (2) a part with higher energy, corresponding to charge excitations on the energy scale $U$, i.e., to excitations raising the number of doubly occupied sites by one (or more).
We can estimate the contribution to the metallic conduction. Since in the tight-binding description the current operator contains the hopping matrix element and is thus subject to the same renormalization as the kinetic energy, we obtain

$$\sigma_1(\omega) = \frac{\omega_p^{*2}}{4}\delta(\omega) + \sigma_1^{\text{high energy}}(\omega), \tag{8.17}$$

where we have used equation (6.12) for a perfect conductor (no residual resistivity in a perfect lattice). There is a high-energy part which we do not specify here. The plasma frequency is renormalized, $\omega_p^{*2} = g_t\omega_p^2$, such that the $f$-sum rule in equation (6.13) yields

$$I = \int_0^\infty d\omega \sigma_1(\omega) = \frac{\omega_p^2}{8} g_t + I_{\text{high energy}} = \frac{\omega_p^2}{8}. \tag{8.18}$$

For $U \to U_c$, the coherent metallic part becomes weaker and weaker,

$$\frac{\omega_p^2}{8} g_t = \left(1 - \left(\frac{U}{U_c}\right)^2\right) \frac{\omega_p^2}{8}. \tag{8.19}$$

According to the $f$-sum rule, the lost weight must gradually be transferred to the "high-energy" contribution.

### 8.1.4 Fermi liquid properties of the metallic state

The just discussed approximation allows us to discuss a few Fermi liquid properties of the metallic state close to metal-insulator transition in a simplified way. Let us investigate the momentum distribution. According to the above definition,

$$\epsilon_{\text{kin}} = \sum_{\boldsymbol{k}\in\text{FS}} \epsilon_{\boldsymbol{k}}, \tag{8.20}$$

where the sum runs over all $\boldsymbol{k}$ in the Fermi sea (FS). One can show within the above approximation, that the distribution is a constant within ($n_{\text{in}}$) and outside ($n_{\text{out}}$) the Fermi surface for finite $U$, such that, for $\boldsymbol{k}$ in the first Brillouin zone,

$$\frac{1}{2} = \frac{1}{N}\sum_{\boldsymbol{k}\in\text{FS}} n_{\text{in}} + \frac{1}{N}\sum_{\boldsymbol{k}\notin\text{FS}} n_{\text{out}} = \frac{1}{2}(n_{\text{in}} + n_{\text{out}}) \tag{8.21}$$

and

$$g_t\epsilon_{\text{kin}} = \frac{1}{N}\sum_{\boldsymbol{k}\in\text{FS}} n_{\text{in}}\epsilon_{\boldsymbol{k}} + \frac{1}{N}\sum_{\boldsymbol{k}\notin\text{FS}} n_{\text{out}}\epsilon_{\boldsymbol{k}}. \tag{8.22}$$

Taking into account particle-hole symmetry, i.e.,

$$\sum_{\boldsymbol{k}} \epsilon_{\boldsymbol{k}} = \sum_{\boldsymbol{k}\in\text{FS}} \epsilon_{\boldsymbol{k}} + \sum_{\boldsymbol{k}\notin\text{FS}} \epsilon_{\boldsymbol{k}} = 0, \tag{8.23}$$

we are able to determine $n_{\text{in}}$ and $n_{\text{out}}$,

$$\left.\begin{array}{l} n_{\text{in}} + n_{\text{out}} = 1 \\ n_{\text{in}} - n_{\text{out}} = g_t \end{array}\right\} \quad \Rightarrow \quad \left\{\begin{array}{l} n_{\text{in}} = (1+g_t)/2 \\ n_{\text{out}} = (1-g_t)/2 \end{array}\right. . \tag{8.24}$$

With this, the jump in the distribution at the Fermi energy is equal to $g_t$, which, as previously, corresponds to the quasiparticle weight (cf. Figure 8.4). For $U \to U_c$ it vanihes, i.e., the quasiparticles cease to exist for $U = U_c$.

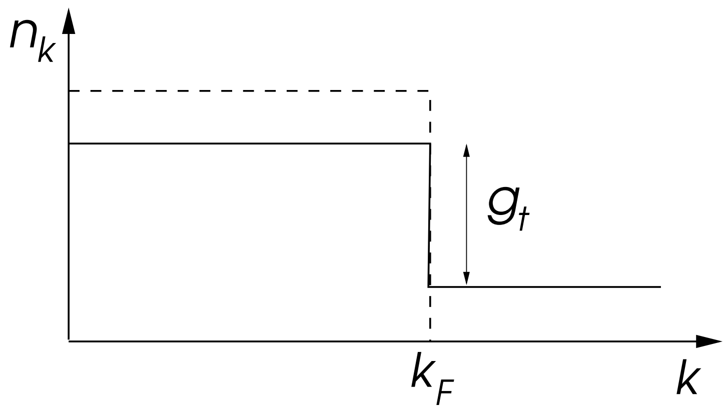



Figure 8.4: The distribution function in the Gutzwiller approximation, displaying the jump at the Fermi energy.

Without going into the details of the calculation, we provide a few Fermi liquid parameters. It is easy to see that the effective mass

$$\frac{m}{m^*} = g_t, \tag{8.25}$$

and thus

$$F_1^s = 3\left(g_t^{-1} - 1\right) = \frac{3U^2}{U_c^2 - U^2}, \tag{8.26}$$

where $t = 1/2m$ and the density of states $N(\epsilon_F)^* = N(\epsilon_F) g_t^{-1}$. Furthermore,

$$F_0^a = -\frac{UN(\epsilon_F)}{4}\frac{2U_c + U}{(U+U_c)^2}U_c, \qquad \Rightarrow \qquad \chi = \frac{\mu_B^2 N(\epsilon_F)^*}{1+F_0^a}, \tag{8.27}$$

$$F_0^s = \frac{UN(\epsilon_F)}{4}\frac{2U_C - U}{(U-U_c)^2}U_c, \qquad \Rightarrow \qquad \kappa = \frac{N(\epsilon_F)^*}{n^2(1+F_0^s)}. \tag{8.28}$$

It follows, that the compressibility $\kappa$ vanishes for $U \to U_c$ as expected, since it becomes more and more difficult to compress the electrons or to add more electrons, respectively. The insulator is, of curse, incompressible. The spin susceptibility diverges because of the diverging density of states $N(\epsilon_F)^*$. This indicates, that local spins form, which exist as completely independent degrees of freedom at $U = U_c$. Only the antiferromagnetic correlation between the spins would lead to a renormalization, which turns $\chi$ finite. This correlation is, as mentioned above, neglected in the Gutzwiller approximation. The effective mass diverges and shows that the quasiparticles are more and more localized close to the transition, since the occupation of a lattice site is getting more rigidly fixed to 1.[4] As a last remark, it turns out that the Gutzwiller approximation is well suited to describe the strongly correlated Fermi liquid $^3$He (cf. [D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984)]).

[4] This can be observed within the Gutzwiller approximation in the form of local fluctuations of the particle number. For this, we introduce the density matrix of the electron states on an arbitrary lattice site,

$$\hat{\rho} = h|0\rangle\langle 0| + d|\uparrow\downarrow\rangle\langle\uparrow\downarrow| + \frac{s}{2}\left(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|\right), \tag{8.29}$$

## 8.2 The Mott insulator as a quantum spin system

One of the most important characteristics of the Mott insulator is the presence of spin degrees of freedom after the freezing of the charge. This is one of the most profound features distinguishing a Mott insulator from a band insulator. In our simple discussion, we have seen that the atomic limit of the Mott insulator provides us with a highly degenerate ground state, where a spin-1/2 degree of freedom is present on each lattice site. We lift this degeneracy by taking into account the kinetic energy term $\mathcal{H}_{\text{kin}}$ ($t \ll U$). In this way new physics appears on a low-energy scale, which can be described by an effective spin Hamiltonian. Prominent examples for such spin systems are transition-metal oxides like the cuprates $La_2CuO_4$, $SrCu_2O_3$ or vanadates $CaV_4O_9$, $NaV_2O_5$.

### 8.2.1 The effective Hamiltonian

In order to employ our perturbative considerations, it is sufficient to observe the spins of two neighboring lattice sites and to consider perturbation theory for discrete degenerate states. Here, this is preferably done in real space. There are 4 degenerate configurations, $\{|\uparrow,\uparrow\rangle, |\uparrow,\downarrow\rangle, |\downarrow,\uparrow\rangle, |\downarrow,\downarrow\rangle\}$. The application of $\mathcal{H}_{\text{kin}}$ yields

$$\mathcal{H}_{\text{kin}}|\uparrow,\uparrow\rangle = \mathcal{H}_{\text{kin}}|\downarrow,\downarrow\rangle = 0, \tag{8.31}$$

$$\mathcal{H}_{\text{kin}}|\uparrow,\downarrow\rangle = -\mathcal{H}_{\text{kin}}|\downarrow,\uparrow\rangle = -t|\uparrow\downarrow,0\rangle - t|0,\uparrow\downarrow\rangle, \tag{8.32}$$

where, in the last two cases, the resulting states have an energy higher by $U$ and lie outside the ground state sector. Thus, it becomes clear that we have to proceed to second order perturbation, where the states of higher energy will appear only virtually (cf. Figure 8.5).

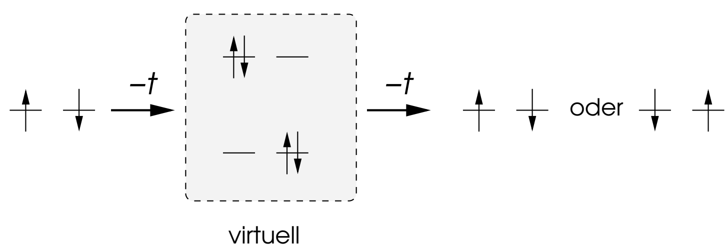



Figure 8.5: Illustration of the origin of the superexchange.

We obtain the matrix elements

$$M_{s_1,s_2;s_1',s_2'} = -\sum_n \langle s_1,s_2|\mathcal{H}_{\text{kin}}|n\rangle \frac{1}{\langle n|\mathcal{H}_{\text{Coul}}|n\rangle} \langle n|\mathcal{H}_{\text{kin}}|s_1',s_2''\rangle, \tag{8.33}$$

where $|n\rangle = |\uparrow\downarrow,0\rangle$ or $|0,\uparrow\downarrow\rangle$, such that the denominator is always $U$. We end up with

$$M_{\uparrow\downarrow;\uparrow\downarrow} = M_{\downarrow\uparrow;\downarrow\uparrow} = -M_{\uparrow\downarrow;\downarrow\uparrow} = -M_{\downarrow\uparrow;\uparrow\downarrow} = -\frac{2t^2}{U}. \tag{8.34}$$

---

from which we deduce the variance of the occupation number,

$$\langle n^2\rangle - \langle n\rangle^2 = \langle n^2\rangle - 1 = \text{tr}(\hat{\rho} n^2) - 1 = 4d + s - 1 = 2d. \tag{8.30}$$

The deviation from single occupation vanishes with $d$, i.e., with the approach of the metal-insulator transition. Note that the dissipation-fluctuation theorem connects $\langle n^2\rangle - \langle n\rangle^2$ to the compressibility.

Note that the signs originates from the anti-commutation properties of the Fermion operators. In the subspace $\{|\uparrow,\downarrow\rangle, |\downarrow,\uparrow\rangle\}$ we find the eigenstates of the respective secular equations,

$$\frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)E = 0, \tag{8.35}$$

$$\frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle)E = -\frac{4t^2}{U}. \tag{8.36}$$

Since the states $|\uparrow,\uparrow\rangle$ and $|\downarrow,\downarrow\rangle$ have energy $E = 0$, the sector with total spin $S = 1$ is degenerate (spin triplet). The spin sector $S = 0$ with the energy $-4t^2/U$ is the ground state (spin singlet).

An effective Hamiltonian with the same energy spectrum for the spin configurations can be written with the help of the spin operators $\widehat{\boldsymbol{S}}_1$ and $\widehat{\boldsymbol{S}}_2$ on the two lattice sites

$$\mathcal{H}_{\text{eff}} = J\left(\widehat{\boldsymbol{S}}_1 \cdot \widehat{\boldsymbol{S}}_2 - \frac{\hbar^2}{4}\right), \qquad J = \frac{4t^2}{U\hbar^2} > 0. \tag{8.37}$$

This mechanism of spin-spin coupling is called *superexchange* and introduced by P.W. Anderson [Phys. Rev. **79**, 350 (1950)].

Since this relation is valid between all neighboring lattice sites, we can write the total Hamiltonian as

$$\mathcal{H}_H = J\sum_{\langle i,j\rangle} \widehat{\boldsymbol{S}}_i \cdot \widehat{\boldsymbol{S}}_j + \text{const.} \tag{8.38}$$

This model, reduced to spins only, is called *Heisenberg model.* The Hamiltonian is invariant under a global $SU(2)$ spin rotation,

$$U_s(\boldsymbol{\theta}) = e^{-i\widehat{\boldsymbol{S}}\cdot\boldsymbol{\theta}}, \qquad \widehat{\boldsymbol{S}} = \sum_j \widehat{\boldsymbol{S}}_j. \tag{8.39}$$

Thus, the total spin is a good quantum number, as we have seen in the two-spin case. The coupling constant is positive and favors an antiparallel alignment of neighboring spins. The ground state is therefore not ferromagnetic.

### 8.2.2 Mean field approximation of the anti-ferromagnet

There are a few exact results for the Heisenberg model, but not even the ground state energy can be calculated exactly (except in the case of the one-dimensional spin chain which can be solved by means of a Bethe Ansatz). The difficulty lies predominantly in the treatment of quantum fluctuations, i.e., the zero-point motion of coupled spins. It is easiest seen already with two spins, where the ground state is a singlet and maximally entangled. The ground state of all antiferromagnetic systems is a spin singlet ($S - tot = 0$). In the so-called thermodynamic limit ($N \to \infty$) there is long-ranged anti-ferromagnetic order in the ground state for dimensions $D \geq 2$. Contrarily, the fully polarized ferromagnetic state (ground state for a model with $J < 0$) is known exactly, and as a state with maximal spin quantum number $\boldsymbol{S}^2$ it features no quantum fluctuations.

In order to describe the antiferromagnetic state anyway, we apply the mean field approximation again. We can characterize the equilibrium state of the classical Heisenberg model (spins as simple vectors without quantum properties) by splitting the lattice into two sublattices $A$ and $B$, where each $A$-site has only $B$-sites as neighbors, and vice-versa.[5] On the $A$-($B$-)sublattice, the spins point up (down). This is unique up to a global spin rotations. Note, that this spin

[5]Lattices which allow for such a separation are called *bipartite*. There are lattices, where this is not possible, e.g., triangular or cubic face centered lattices. There, frustration phenomena appear, a further complication of anti-ferromagnetically coupled systems.

configuration doubles the unit cell.
We introduce the respective mean field,

$$\widehat{S}_i^z = \begin{cases} m + (\widehat{S}_i^z - m) & i \in A \\ \\ -m + (\widehat{S}_i^z + m) & i \in B \end{cases}. \tag{8.40}$$

This leads to the mean field Hamiltonian

$$\mathcal{H}_{\text{mf}} = \mathcal{H}_A + \mathcal{H}_B = -Jzm\sum_{i\in A}\widehat{S}_i^z + Jzm\sum_{i\in B}\widehat{S}_i^z + Jz\frac{m^2}{2}N + \cdots, \tag{8.41}$$

with the coordination number $z$, the number of nearest neighbors ($z = 6$ in a simple cubic lattice). It is simple to calculate the partition sum of this Hamiltonian,

$$Z = \text{tr}\left(e^{-\beta\mathcal{H}_{\text{mf}}}\right) = \left[\left(e^{\beta Jmz\hbar/2} + e^{-\beta Jmz\hbar/2}\right)e^{-\beta Jzm^2/2}\right]^N. \tag{8.42}$$

The free energy per spin is consequently given by

$$F(m,T) = -\frac{1}{N}k_BT\ln Z = Jz\frac{m^2}{2} - k_BT\ln\left(2\cosh(\beta Jzm\hbar/2)\right). \tag{8.43}$$

At fixed temperature, we minimize the free energy with respect to $m$ to determine the thermal equilibrium state,[6] i.e., set $\partial F/\partial m = 0$ and find

$$m = \frac{\hbar}{2}\tanh\left(\frac{Jzm\hbar}{2k_BT}\right). \tag{8.44}$$

This is the self-consistency equation of the mean field theory. It provides a critical temperature $T_N$ (Nel temperature), below which the mean moment $m$ is finite. For $T \to T_{N-}$, $m$ approaches 0 continuously. Thus, $T_N$ can be found from a linearized self-consistency equation,

$$m = \left.\frac{Jzm\hbar^2}{4k_BT}\right|_{T=T_N}, \tag{8.45}$$

and thus

$$T_N = \frac{Jz\hbar^2}{4k_B}. \tag{8.46}$$

This means, that $T_N$ scales with the coupling constant and with $z$. The larger $J$ and the more neighbors are present, the more stable is the ordered state.[7] For $T$ close to $T_N$, we can expand the free energy in $m$,

$$F(m,T) = F_0 + \frac{Jz}{2}\left[\left(1 - \frac{T_N}{T}\right)m^2 + \frac{2}{3\hbar^2}\left(\frac{T_N}{T}\right)^3 m^4\cdots\right]. \tag{8.47}$$

This is a Landau theory for a phase transition of second order, where a symmetry is spontaneously broken. The breaking of the symmetry (from the high-temperature phase with high symmetry to the low-temperature phase with low symmetry) is described by the order parameter $m$. The minimization of $F$ with respect to $m$ yields (cf. Figure 8.6)

$$m(T) = \begin{cases} 0, & T > T_N, \\ \\ \frac{\hbar}{2}\sqrt{3(T_N/T - 1)}, & T \le T_N. \end{cases} \tag{8.48}$$

[6] Actually, a magnetic field pointing into the opposite direction on each site would be another equilibrium variable (next to the temperature). We set it to zero.
[7] At infinite $z$, the mean field approximation becomes exact.

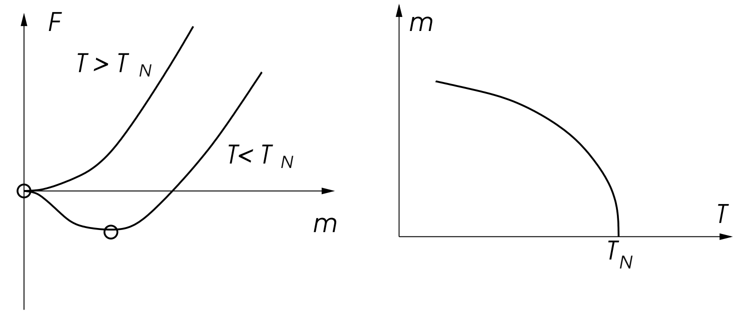



Figure 8.6: The free energy and magnetization of the anti-ferromagnet above and below $T_N$.

## 8.3 Collective modes – spin wave excitations

Besides its favorable properties, the mean field approximation also has a number of insufficiencies. Quantum and some part of thermal fluctuations are neglected, and the insight into the low-energy excitations remains vague. As a matter of fact, as in the case of the ferromagnet, collective excitations exist here. In order to investigate these, we write the Heisenberg model in its spin components, i.e.,

$$\mathcal{H}_H = J \sum_{\langle i,j \rangle} \left( \widehat{S}_i^z \widehat{S}_j^z + \frac{1}{2} \left( \widehat{S}_i^+ \widehat{S}_j^- + \widehat{S}_i^- \widehat{S}_j^+ \right) \right). \tag{8.49}$$

In the ordered state, the moments shall be aligned along the $z$-axis.
To observe the dynamics of a flipped spin, we apply the operator $\widehat{S}_l^-$ on the ground state $|\Phi_0\rangle$, and determine the spectrum, by solving the resulting eigenvalue equation

$$(\mathcal{H}_H - E_0)\widehat{S}_l^- |\Phi_0\rangle = [\mathcal{H}_H, \widehat{S}_l^-]|\Phi_0\rangle = \hbar\omega \widehat{S}_l^- |\Phi_0\rangle, \tag{8.50}$$

with the ground state energy $E_0$. Using the spin-commutation relations

$$\left[\widehat{S}_j^+, \widehat{S}_j^-\right] = 2\widehat{S}_j^z \delta_{ij}, \tag{8.51}$$

$$\left[\widehat{S}_j^z, \widehat{S}_j^\pm\right] = \pm\widehat{S}_j^\pm \delta_{ij}, \tag{8.52}$$

then yields the equation

$$\left[ -J{\sum_j}' \widehat{S}_j^z \widehat{S}_l^- + J{\sum_j}' \widehat{S}_j^- \widehat{S}_l^z - \hbar\omega \widehat{S}_l^- \right] |\Phi_0\rangle = 0, \tag{8.53}$$

where $\sum_j'$ runs over all neighbors of $l$. We decouple this complicated problem by replacing the operators $\widehat{S}^z$ by their mean fields. Therefore, we have to distinguish between $A$ and $B$ sublattices, such that we end up with two equations,

$$\left( Jmz\widehat{S}_l^- + Jm \sum_{\boldsymbol{a}} \widehat{S}_{l+\boldsymbol{a}}^- - \hbar\omega \widehat{S}_l^- \right) |\Phi_0\rangle = 0, \qquad l \in A, \tag{8.54}$$

$$\left( -Jmz\widehat{S}_{l'}^- - Jm \sum_{\boldsymbol{a}} \widehat{S}_{l'+\boldsymbol{a}}^- - \hbar\omega \widehat{S}_{l'}^- \right) |\Phi_0\rangle = 0, \qquad l' \in B. \tag{8.55}$$

We introduce the operators

$$\widehat{S}_l^- = \sqrt{\frac{2}{N}} \sum_{\boldsymbol{q}} \widehat{a}_{\boldsymbol{q}}^\dagger e^{i\boldsymbol{q}\cdot\boldsymbol{r}_l}, \tag{8.56}$$

$$\widehat{S}_{l'}^- = \sqrt{\frac{2}{N}} \sum_{\boldsymbol{q}} \widehat{b}_{\boldsymbol{q}}^\dagger e^{i\boldsymbol{q}\cdot\boldsymbol{r}_{l'}}, \tag{8.57}$$

with $l \in A$ and $l' \in B$, and, vice versa,

$$\widehat{a}_{\boldsymbol{q}}^\dagger = \sqrt{\frac{2}{N}} \sum_{l \in A} \widehat{S}_l^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_l}, \tag{8.58}$$

$$\widehat{b}_{\boldsymbol{q}}^\dagger = \sqrt{\frac{2}{N}} \sum_{l' \in B} \widehat{S}_{l'}^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_{l'}}, \tag{8.59}$$

and insert them into the equation and obtain,

$$\left( (Jmz - \hbar\omega) \sum_{l \in A} \widehat{S}_l^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_l} + Jm \sum_{\boldsymbol{a}} e^{i\boldsymbol{q}\cdot\boldsymbol{a}} \sum_{l' \in B} \widehat{S}_{l'}^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_{l'}} \right) |\Phi_0\rangle = 0, \tag{8.60}$$

$$\left( (-Jmz - \hbar\omega) \sum_{l' \in B} \widehat{S}_{l'}^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_{l'}} - Jm \sum_{\boldsymbol{a}} e^{i\boldsymbol{q}\cdot\boldsymbol{a}} \sum_{l \in A} \widehat{S}_l^- e^{-i\boldsymbol{q}\cdot\boldsymbol{r}_l} \right) |\Phi_0\rangle = 0. \tag{8.61}$$

From this follows that

$$\left( (Jmz - \hbar\omega)\widehat{a}_{\boldsymbol{q}}^\dagger + Jm\gamma_{\boldsymbol{q}}\widehat{b}_{\boldsymbol{q}}^\dagger \right) |\Phi_0\rangle = 0, \tag{8.62}$$

$$\left( (-Jmz - \hbar\omega)\widehat{b}_{\boldsymbol{q}}^\dagger - Jm\gamma_{\boldsymbol{q}}\widehat{a}_{\boldsymbol{q}}^\dagger \right) |\Phi_0\rangle = 0, \tag{8.63}$$

with $\gamma_{\boldsymbol{q}} = \sum_{\boldsymbol{a}} e^{i\boldsymbol{q}\cdot\boldsymbol{a}} = 2(\cos q_x a + \cos q_y a + \cos q_z a)$. This eigenvalue equation is easily solved leading to the description of spin waves in the antiferromagnet. The energy spectrum is given by

$$\hbar\omega_{\boldsymbol{q}} = \pm Jm\sqrt{z^2 - \gamma_{\boldsymbol{q}}^2}. \tag{8.64}$$

Note, that only the positive energies make sense. It is interesting to investigate the limit of small $\boldsymbol{q}$,

$$z^2 - \gamma_{\boldsymbol{q}}^2 \to z^2\boldsymbol{q}^2 + O(q^4), \tag{8.65}$$

where

$$\hbar\omega_{\boldsymbol{q}} = Jmz|\boldsymbol{q}| + \cdots . \tag{8.66}$$

This means that, in contrast to the ferromagnet, the spin waves of the antiferromagnet have a linear low-energy spectrum (cf. Figure 8.7). The same applies here if we expand the spectrum around $\boldsymbol{Q} = (1,1,1)\pi/a$ (folding of the Brillouin zone due to the doubling of the unit cell).
After a suitable normalization, the operators $\widehat{a}_{\boldsymbol{q}}$ and $\widehat{b}_{\boldsymbol{q}}$ are of bosonic nature; this comes about since, due to the mean field approximation, the $\widehat{S}_l^\pm$ are bosonic as well,

$$[\widehat{S}_l^+, \widehat{S}_j^-] = 2\widehat{S}_l^z \delta_{lj} \approx \pm 2m\delta_{lj}, \tag{8.67}$$

where the sign depends on the sublattice. The zero-point fluctuations of these bosons yield quantum fluctuations, which reduce the moment $m$ from its mean field value. In a one-dimensional

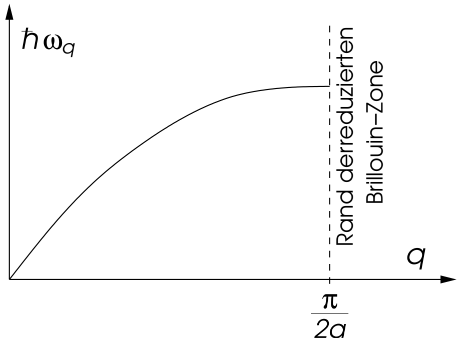



Figure 8.7: Spectrum of the spin waves in the antiferromagnet.

spin chain these fluctuations are strong enough to suppress antiferromagnetically order even for the ground state. The fact that the spectrum starts at zero has to do with the infinite degeneracy of the ground state. The ordered moments can be turned into any direction globally. This property is known under the name *Goldstone theorem*, which tells that each ordered state that breaks a continuous symmetry has collective excitations with arbitrary small (positive) energies. The linear spectrum is normal for collective excitations of this kind; the quadratic spectrum of the ferromagnet has to do with the fact that the state breaks time-inversion symmetry.

These spin excitations show the difference between a band and a Mott insulator very clearly. While in the band insulator both charge and spin excitations have an energy gap and are inert, the Mott insulator has only gapped charge excitation. However, the spin degrees of freedom for a low-energy sector which can even form gapless excitations as shown just above.