

### UNIVERSITÄT ZU KÖLN

Institute for Theoretical Physics

LECTURE NOTES

## Solid State Theory

Prof. Dr. Achim Rosch

## CONTENTS

1	Soli	ds 1						
	1.1	Solids as Crystals						
	1.2	Crystalline Structures						
	1.3	Reciprocal Lattice						
	1.4	Scattering from a Crystal						
2	Solid as Quantum System 9							
	2.1	Adiabatic approximation						
	2.2	Hartree Fock Approximation 11						
	2.3	Density Functional Theory (LDA) 14						
		2.3.1 A theorem						
		2.3.2 Local density approximation and Kohn-Sham equations						
		<b>2.3.3</b> Exchange energy						
3	Electrons in a Periodic Potential 21							
	3.1	Bloch's Theorem						
	3.2	Electrons in a Weak Periodic Potential						
	3.3	Fermi Surfaces						
	Density of States, van-Hove singularities and thermodynamics							
		<b>3.4.1</b> Density of states						
		3.4.2 Van-Hove singularities						
		3.4.3 Thermodynamics of non-interacting electrons and Sommerfeld expansion 34						
	3.5	Almost localized electrons: tight binding model						
		3.5.1 Linear combination of atomic orbitals - LCAO 36						
		3.5.2 Phenomenology with tight-binding approaches						
		<b>3.5.3</b> Graphene						
		3.5.4 Wannier functions						
	3.6	Symmetries and degeneracies in the band-structure						
		3.6.1 Time reversal invariance						
		3.6.2 Band degeneracies due to symmetries						
		3.6.3 Accidental degeneracies and Weyl points						
	3.7	Topological insulators - a case study						
		3.7.1 Insulator with surface states						
		3.7.2 A topological insulator: Quantum spin-Hall effect						
		3.7.3 Outlook: Topology in three dimensions						

4	Sen 4.1 4.2 4.3	hiclassical Dynamics of Bloch Electrons         Equations of Motion & Berry Curvature         4.1.1 Berry phase         4.1.2 Wave-packet dynamics         4.1.3 Anomalous Hall effect and TKNN formula         Bloch Oscillations and Zener Tunneling         Bloch Electrons in a Magnetic Field	<b>57</b> 57 59 64 67 69
5	<b>Tra</b> 5.1 5.2 5.3 5.4	nsport TheoryDrude ModelBoltzmann EquationCollision termsLinearized Boltzmann equation and relaxation time approximation5.4.1Perturbative solution of the Boltzmann equation5.4.2Scattering kernel5.4.3Conservation laws5.4.4Relaxation time approximation5.5.1Conductivity in relaxation time approximation5.5.2Transport scattering rate5.5.3Adding scattering processes – Matthiessen's rule	<b>75</b> 75 77 78 81 81 82 83 84 87 87 89 91
6	Lat 6.1 6.2 6.3 6.4	tice Vibrations Classical Lattice Dynamics Quantum Theory of Lattice Vibration: Phonons Thermodynamics of phonons Electron-Phonon Scattering	<b>93</b> 93 97 99 102
6	Lat 6.1 6.2 6.3 6.4 Inte 7.1 7.2 7.3 7.4 7.5	tice Vibrations         Classical Lattice Dynamics         Quantum Theory of Lattice Vibration: Phonons         Thermodynamics of phonons         Electron-Phonon Scattering         Electron-Phonon Scattering         eracting Phases of Matter         Spontaneous Symmetry Breaking         7.1.1 Symmetry breaking         7.1.2 Order parameter         7.1.3 Goldstone Theorem         Magnetism         7.2.1 Itinerant Magnetism         7.2.2 Local Magnetism         Superconductivity         7.3.1 BCS Theory         7.3.2 Supercurrents and Anderson-Higgs Mechanism         Outlook: Phases of matter	<ul> <li>93</li> <li>93</li> <li>97</li> <li>99</li> <li>102</li> <li>103</li> <li>103</li> <li>105</li> <li>106</li> <li>109</li> <li>109</li> <li>112</li> <li>116</li> <li>117</li> <li>120</li> <li>124</li> <li>126</li> </ul>

# SOLIDS

#### 1.1 Solids as Crystals

A large part of study of solids is devoted to the understanding of crystals. Crystals are characterized by a periodic arrangement of atoms in a regular structure that is repeated through space. The aforementioned idealization also assumes that the crystal is infinite in size. Real crystals are, however, finite and have a definite beginning and an end (and hence a boundary). Furthermore, they are also interspersed with random defects that deform the idealized pattern that one assumes as a first approximation for the sake of simplicity.



Figure 1.1: Left: 2D triangular lattice. Right: 3D simple cubic lattice

It is worthwhile to emphasize that a crystal has a number of remarkable properties. Some of them sound obvious and trivial but actually should be considered as highly surprising. We will discuss only two of them (which are actually closely related). First, the positions of the atoms are highly correlated: when one moves the atoms on the left side of the solid, also atoms on the right side move, despite being a long distance away. This is surprising when one looks at the laws of nature which tell us that the direct interaction between those particles is very weak.

Second, we can build objects like bridges or chairs out of solids. These are structures which can sustain forces without any cost of energy. Let us describe the latter observation is different terms. According to Newton's law,  $\mathbf{F} = \partial_t \mathbf{p}$ , forces are responsible for the change of momentum. Therefore sustaining a force without any cost of energy is equivalent to a dissipationless flow of momentum. This is surprising as usually the flow of something (water, charge, energy,...) is accompanied by friction and dissipation. Dissipationless flow is the exception. One famous exception is a superconductor where charge is flowing without friction. Solids are

another exceptions as momentum is flowing without friction in this case, one could even call a solid a momentum-superconductor. It turns out that the physics and mathematics of the two phenomena are closely related. We will not go into details here but only give a hint to the underlying physics. According to Noether's theorem, momentum conservation arises because the laws of nature are translational invariant, i.e., they are independent of the position in space. While a liquid is translationally invariant and looks the same everywhere, this is not the case in a solid where the atoms are located on a regular lattice. Moving, e.g., by half a lattice spacing one observes locally a different state. We say that the continuous translational symmetry is spontaneously broken. Translational symmetry is still a property of the laws of nature but not of the thermodynamic state of a solid. The periodic arrangement of atoms explains why they move in such a correlated way and, perhaps, it is also plausible that it is linked to the fact that a solid can sustain forces without any energy cost. The story of a superconductor is similar: charge conservation is also related to a symmetry. Here the relevant symmetry is the multiplication of wavefunctions by an arbitrary phase factor  $e^{i\phi}$ , and in a superconductor this symmetry is also spontaneously broken - a story we might came back to at the end of the course.

Another important consequence translational invariance its spontaneous breaking in a solid, is that it does not cost any energy to move a crystal. We will discuss later how this observation leads naturally to the existence of acoustic phonons, lattice vibrations whose energy vanishes in the limit of small momentum.

In passing, we would also like to mention that there a solid objects which cannot be classified as crystals. One such example is *glass*, where the atom arrangement has mainly short-rangled correlations and resembles more a frozen-in liquid. While one can view glasses as liquids with an extremely high viscosity, they behave similar to solids for most practical purposes. Another interesting type of solid are *quasi-crystals*, which are systems with a long-range order which has no well-defined periodicity.

We will, however, focus in the following on crystals, the by far most simple and most important type of solid.

#### **1.2** Crystalline Structures

In the following we will give a short crash course on the basic tools needed to describe a solid, assuming that most of the readers have heard the underlying concepts before, e.g., in a course on solid state physics on the Bachelor level. For this audience, the following section is a short reminder. In case that you have never seen this before, I would recommend to consult any book on solid state physics for an introduction to the subject. Below we simply list the vocabluary which is used in the following. Wikipedia is also an excellent source to look up some of the basic concepts listed below.

**Bravais Lattice**: The most basic structure, which we need to describe periodic arrangements of atoms are so called Bravais lattices. Bravais lattices are simply periodic arrangements of points defined by three linearly independent basis vectors,  $a_i$ , i = 1, 2, 3, multiplied by integers

$$\boldsymbol{R}_n = n_1 \boldsymbol{a}_1 + n_2 \boldsymbol{a}_2 + n_3 \boldsymbol{a}_3, \qquad n_i \in \mathbb{Z}.$$

For a given lattice, the basis vectors are not unique.

There are 14 different types of Bravais lattices, below we discuss a few which we will need later.

Example 1: Simple cubic lattice (rare)

 $a_1 \perp a_2 \perp a_3, |a_1| = |a_2| = |a_3|.$ 

Example 2: Body-centered cubic lattice (BCC)

#### 1.2. Crystalline Structures



Figure 1.2: Simple cubic lattice



Figure 1.3: Body-centered cubic lattice

$$a_1 = a\hat{x}, a_2 = a\hat{y}, a_3 = \frac{a}{2}(\hat{x} + \hat{y} + \hat{z})$$

Example 3: Face centered cubic lattice (FCC)



Figure 1.4: Face-centered cubic lattice

$$a_1 = rac{a}{2}(\hat{m{y}} + \hat{m{z}}), a_2 = rac{a}{2}(\hat{m{x}} + \hat{m{z}}), a_3 = rac{a}{2}(\hat{m{x}} + \hat{m{y}}).$$

Example 4: Triclinic lattice (lowest symmetry)

 $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$  along arbitrary directions. And  $|\boldsymbol{a}_1| = |\boldsymbol{a}_2| = |\boldsymbol{a}_3|$  all different.

**Coordination Number:** Number of nearest neighbors of an atom or a Bravais lattice. The Bravais lattice with the highest coordination number is the FCC lattice.

**Primitive Unit Cell**: The *smallest* volume which fills space after translations by Bravais lattice. It is also not unique! However, its volume is unique.

$\mathbf{SC}$	BCC	FCC	Triclinic
6	8	12	2

Table 1.1: Coordination number (number of nearest neighbors of a point) in different lattice systems

eg.  $\mathcal{UC} = \{x_1 a_1 + x_2 a_2 + x_3 a_3 | 0 \leq x_i \leq 1\}.$ 

It leads to a unique representation of an arbitrary vector, r, in space as  $r = R_n + x$ ,  $x \in \mathcal{UC}$ .

**Wigner-Seitz cell** is defined as the volume:  $\{x \in \mathbb{R}^3 | |x| \leq |x - R_n| \forall n \neq (0, 0, 0)\}$ . It is a special example of a primitive unit cell which has, e.g., the advantage that it has all the symmetries of the underlying lattice.



Figure 1.5: Wigner-Seitz cell for a square/rectangular lattice.

**Basis**: To define a crystal, i.e., a periodic arrangement of atoms, we have to specify (i) the relevant Bravais lattice and (ii) the positions of each atom within the unit cell. The latter is called basis.

Examples: 1. The diamond lattice (diamond, Si, Ge) is an FCC lattice with basis **0**,  $\frac{a}{4}(\hat{x} + \hat{y} + \hat{z})$ ; 2. NaCl structure is also an FCC where the Na atoms have the basis vector **0**, while the Cl atoms are located at the basis vector  $\frac{a}{2}(\hat{x} + \hat{y} + \hat{z})$ .

**Symmetries**: Symmetries which leave a crystal invariant are an important tool to classify crystals (see below). They furthermore determine many of its properties and are heavily used when developing the theory for materials. Besides translations these are rotations (only rotations by multiples of 60, 90, 120 and 180 degrees are possible for periodic structures), mirror transformations, inversion, and combinations thereof. Each of these transformations or inversions and a vector describing a translation.

$$r' = \underbrace{D}_{ ext{rotation,}} r + \underbrace{a}_{ ext{translation}}.$$

Note that a does not necessarily have to be a vector of the Bravais lattice. In this case one talks about non-symmomorphic symmetries. For example, a symmetry transformation can combine a rotation by 180 degrees with a translation by half of a unit vector of the Bravais lattice (screw axis). Similarly, a mirror transformation can be combined with such a translation (glide plane).

**Space groups**: The set of all symmetry operations which leave an (idealized) crystal invariant forms a group, the so-called space group. There are 230 different space groups (if one also considers how time reversal symmetry acts on a magnet, one obtains 1651 magnetic space groups needed for a complete symmetry classification of all types of magnets and antiferromagnets).

**Crystalline point-groups**: If one ignores the translation a in the symmetry transformations of a crystal, one obtains the point group of a crystal (here glide planes and screw axes are simply

replaced by simple mirror transformations and rotations). There are 32 different point-groups, for a list and the relevant notations see the corresponding wikipedia article.

#### **1.3 Reciprocal Lattice**

The theory of solids is mainly formulated in momentum space. Therefore we need to consider the Fourier transformation of quantities like the periodic potential  $V(\mathbf{r})$  arising from the periodic arrangement of atoms. We first expand the potential in Fourier modes,  $V(\mathbf{r}) = \sum_{m} V_{\mathbf{G}_m} e^{i\mathbf{G}_m \cdot \mathbf{r}}$ . where the set of reciprocal lattice vectors  $\mathbf{G}_m$  has not yet been specified. Now we should consider which restrictions arise from the condition that the potential is periodic

$$V(\boldsymbol{r}+\boldsymbol{R}_n)=V(\boldsymbol{r}), \quad \forall \ \boldsymbol{R}_n=\sum_{i=1}^3 n_i \boldsymbol{a}_i, \ n_i\in\mathbb{Z}$$

This condition can only be fulfilled if the  $G_m$  obey the following relation

$$e^{i\boldsymbol{G}_m\cdot\boldsymbol{R}_n} = 1 \Leftrightarrow \boldsymbol{G}_m\cdot\boldsymbol{R}_n = 2\pi m, \ m \in \mathbb{Z} \quad \forall \, \boldsymbol{G}_m, \boldsymbol{R}_n$$

This condition is met for all vectors of the so-called **reciprocal lattice** spanned by the vectors  $G_m$  with

$$G_m = m_1 G_1 + m_2 G_2 + m_3 G_3, \quad m_i \in \mathbb{Z}, m = (m_1, m_2, m_3)$$
 (1.1)

$$G_i \cdot a_j = 2\pi \delta_{i,j}, \quad i, j = 1, 2, 3$$
 (1.2)

This condition is fulfilled by the following basis vectors of the reciprocal lattice

$$G_1 = 2\pi \frac{\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{1.3}$$

$$\boldsymbol{G}_2 = 2\pi \frac{\boldsymbol{a}_3 \times \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{1.4}$$

$$\boldsymbol{G}_3 = 2\pi \frac{\boldsymbol{a}_3 \times \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \tag{1.5}$$

While the Bravais lattice is a lattice in real space, the reciprocal lattice is a lattice in Fourier space.

Example: We previously considered NaCl which forms an fcc lattice where basis vectors can be chosen as

$$\boldsymbol{a}_1 = \frac{a}{2} \begin{pmatrix} 0\\1\\1 \end{pmatrix}, \boldsymbol{a}_2 = \frac{a}{2} \begin{pmatrix} 1\\0\\1 \end{pmatrix}, \boldsymbol{a}_3 = \frac{a}{2} \begin{pmatrix} 1\\1\\0 \end{pmatrix},$$
  
where the Na atom is located at  $\begin{pmatrix} 0\\1\\1 \end{pmatrix} + \boldsymbol{R}_n$  while the Cl can be found at  $\frac{a}{2} \begin{pmatrix} 1\\1\\1 \end{pmatrix} + \boldsymbol{R}_n$ 

From the formula given above, we can compute the basis vectors of the reciprocal lattice

$$\boldsymbol{G}_1 = \frac{2\pi}{a} \begin{pmatrix} -1\\ 1\\ 1 \end{pmatrix}, \ \boldsymbol{G}_2 = \frac{2\pi}{a} \begin{pmatrix} 1\\ -1\\ 1 \end{pmatrix}, \ \boldsymbol{G}_3 = \frac{2\pi}{a} \begin{pmatrix} 1\\ 1\\ -1 \end{pmatrix},$$

these are actually the basis vectors of a body-centered cubic lattice: the reciprocal lattice of an fcc lattice is a bcc lattice (and vice versa).



Figure 1.6: Reciprocal lattice of an fcc lattice is a bcc lattice

An important concept, heavily used in the following is the **First Brillouin zone**, abbreviated as **1**. **BZ**. It is defined as the Wigner-Seitz cell (see definition given in the previous section) of the reciprocal lattice. Using the 1. BZ, any arbitrary momentum vector,  $\boldsymbol{k}$ , can be uniquely written as

$$\boldsymbol{k} = \boldsymbol{G}_n + \boldsymbol{k}_{\mathrm{BZ}}, \text{ with } \boldsymbol{k}_{\mathrm{BZ}} \in 1. \mathrm{BZ}.$$

As a side remark, let me also mention that reciprocal lattice vectors can be used to identify planes in the real-space lattice. Each lattice plane of a Bravais lattice (defined by 3 noncollinear points of the Bravais lattice) is perpendicular to a set of reciprocal lattice vectors  $G_m = m_1 G_1 + m_2 G_2 + m_3 G_3$ . Therefore a set of three integers,  $(m_1, m_2, m_3)$ , (usually chosen such that  $gcd(m_1, m_2, m_3) = 1$ ), can be used to indentify such planes. They are called are the Miller indices of this plane. This is, however, a concept which we will not use in the following.

#### 1.4 Scattering from a Crystal



Figure 1.7: A scattering experiment

Reciprocal lattice vectors are not only central for the theoretical description of solids but also for a large set of experiments. The most direct way to determine them is to do a scattering experiment using X-rays or neutrons. Let us consider elastic scattering, i.e., scattering without energy transfer from a crystal. What can we expect when we scatter waves from a periodic potential? Here we will not try to develop a detailed theory of scattering but instead simply recall a few basic principles of scattering which most of the readers will already be familiar with.



Figure 1.8: Path difference between two interfering rays scattered off of two atoms, one at  $r_0$  and the other at  $r_0 + d$ .

Consider an incoming plane wave described by  $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ ,  $\boldsymbol{k} = \frac{2\pi}{\lambda}\hat{\boldsymbol{n}}$  which is scattered to an outgoing wave,  $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ ,  $\boldsymbol{k}' = \frac{2\pi}{\lambda}\hat{\boldsymbol{n}}'$ . As we consider elastic scattering, ingoing and outgoing waves have the same wavelength  $\lambda$ . The unit vectors  $\hat{\boldsymbol{n}}$  and  $\hat{\boldsymbol{n}'}$  denote the directions in which the waves travel. When a given incoming wave scatters from a period arrangement of atoms, then for most directions  $\hat{\boldsymbol{n}'}$  there will only be destructive interference. Our goal is to find out, for which special situations (called Bragg condition) we can expect constructive interference. For this, we have to consider the path difference of a wave scattered from, e.g., an atom located at position  $r_0 + d$  which is given by  $d \cdot (\hat{\boldsymbol{n}} - \hat{\boldsymbol{n}'})$ . Constructive interference is obtained if the path difference is an integer multiple of the wavelength,  $d \cdot (\hat{\boldsymbol{n}} - \hat{\boldsymbol{n}'}) = 2\pi m$ , or, equivalently,  $d \cdot (\boldsymbol{k} - \boldsymbol{k}') = 2\pi m$ . Therefore, constructive interference from all atoms of a Bravais lattice located at positions  $\boldsymbol{R}_n$  can only be obtained if

$$\boldsymbol{R}_n \cdot (\boldsymbol{k} - \boldsymbol{k}') = 2\pi m, \quad m \in \mathbb{Z}$$

We conclude that scattering from a periodic lattice (Bragg scattering) is only possible if

 $(\boldsymbol{k} - \boldsymbol{k}') = \boldsymbol{G}_n$  is a reciprocal lattice vector

In more complicated systems with several atoms per unit cell the presence of further symmetries can lead to extinctions, implying that not all possible  $G_n$  may be observed in such a scattering experiment.

For an alternative and more quantitative derivation of the scattering rates one can describe the interaction of the scattered particl, e.g., a neutron with the solid by an effective potential  $\Delta H = V(\mathbf{r})$ . As this potential is usually weak compared to the kinetic energy one can use the Golden rule formula,  $\Gamma_{\mathbf{k}',\mathbf{k}} = \frac{2\pi}{\hbar} |\langle \mathbf{k}' | \Delta H | \mathbf{k} \rangle|^2 \delta(E_{\mathbf{k}} - E_{\mathbf{k}'})$ , to calculate the transition rate of the neutron from a state  $\mathbf{k}$  to a state  $\mathbf{k}'$ . The transition matrix element is then computed from

$$\langle \mathbf{k}' | \Delta H | \mathbf{k} \rangle = \int e^{-i\mathbf{k}' \cdot \mathbf{r}} V(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} \, \mathrm{d}\mathbf{r} = V_{\mathbf{k}-\mathbf{k}'}$$

As we argued above, for a periodic potential  $V(\mathbf{r})$  the Fourier transform can only be finite if  $(\mathbf{k} - \mathbf{k}') = \mathbf{G}_n$  is a reciprocal lattice vector. Therefore scattering experiments are ideal to determine reciprocal lattice vectors, the precise structure of a crystal and its symmetries.

This concludes our - very brief - introduction to the some basic notions used to describe solids.

## SOLID AS QUANTUM SYSTEM

To develop a quantum theory of a solid appears to be a challenging problem. Even with the best supercomputer it is only possible to the solve the Schrödinger equation for a handful of electrons. Just storing a generic quantum mechanical wave function for a few dozen electrons (with a reasonable discretization) on a computer is simply not possible because the amount of storage grows exponentially with system size. Surprisingly, it is nevertheless possible to understand the quantum properties of many materials and therefore of quantum systems with about  $10^{23}$  quantum particles both qualitatively and in many cases even quantitatively. The main goal of this course is to discuss some of most important approximations underlying remarkable success story. We start by formulating the many-particle quantum problem which we have to face.

We start by describing a solid as system consisting of charged ions and electrons. Here "ions" may refer to the nuclei of atoms but sometimes one also defines an ion as the tightly bound state of a nucleus plus some surrounding "core" electrons from the inner shells of the atom. To simplify notations, we will assume in the following that there is just one type of ions and we will ignore some relativistic corrections which are small for most solids.

The total Hamiltonian of the system can then be written down

$$H = H_{\rm el} + H_{\rm ion} + H_{\rm el-ion}$$

$$H_{\rm el} = \sum_{i} \frac{\boldsymbol{p}_{i}^{2}}{2m} + \sum_{i < j} \frac{e^{2}}{|\boldsymbol{r}_{i} - \boldsymbol{r}_{j}|}$$

$$H_{\rm ion} = \sum_{n} \frac{\boldsymbol{P}_{n}^{2}}{2M} + \sum_{n < m} \frac{Z^{2}e^{2}}{|\boldsymbol{R}_{n} - \boldsymbol{R}_{m}|}$$

$$H_{\rm el-ion} = -\sum_{l,n} \frac{Ze^{2}}{|\boldsymbol{r}_{l} - \boldsymbol{R}_{n}|}$$

$$(2.1)$$

Here M and Z are the mass and charge of the ions while m is the electron mass. Besides the kinetic energies of the quantum particles, there is only the Coulomb energy (we use Gaussian units here in which the Coulomb energy is  $\frac{q_1q_2}{r}$  and in the formula written above, we ignored for simplicity corrections arising from the core electrons which lead to a modified potential at close to the ionic core. The simple Hamiltonian (and trivial variants obtained by adding different types of ions) describes a remarkable amount of physics (and chemistry) ranging from molecules to solids and can - in principle - also be applied to such complicated structures as the

human brain. Also many yet undiscovered quantum phenomena and new materials are hidden in this harmless looking Hamiltonian.

The fast that H is well known is of advantage but the problem is that the solution of the corresponding Schrödinger equation remains unknown and - as argued above is beyond the capacity of any supercomputer (perhaps future quantum computers will help to solve such challenging quantum problems directly).

There are many different approaches how one can make progress here. We will first discuss how one may try to solve the problem in a directly: we discuss (rather crude) approximation tailored to solve the problem defined defined by H directly. This set of approaches is called "ab initio" approaches or solutions "from first principles", both referring to the fact that one starts from basic underlying Hamiltonian. Another approach is to use simplified model systems as the starting point (e.g., a Heisenberg or Ising model to describe a magnet). Finally, one can try to "guess" an effective low-energy theory (using, e.g., powerful symmetry principles) often with parameters which then can be fitted to experiments. But we will start with the most straightforward approach, a direct attack on H.

#### 2.1 Adiabatic approximation

To make progress and obtain a simplified description of a solid, we will first make use of the fact that electrons are much, much lighter than ions,  $m/M_{\rm ion} \sim 10^{-3}$ . Therefore electrons also move much faster. As a (very good) approximation, we can assume that on the time-scales relevant for electron motion, the ions do not move at all. This is called "adiabatic approximation". Our starting point is thus the electronic wave function for fixed ion positions  $\mathbf{R}_1, \ldots, \mathbf{R}_N$  written as

$$\psi = \psi_{\boldsymbol{R}_1, \boldsymbol{R}_2, \dots, \boldsymbol{R}_N}(\boldsymbol{r}_1, \dots, \boldsymbol{r}_N)$$

We will never know this wave function exactly, as the underlying many-particle problem is too complicated, but for the moment let us just pretend that we know that we know the eigenstates (or at least the ground state) of the corresponding Schrödinger equation

$$(H_{\rm el} + H_{\rm el-ion})\psi = E^{\rm el}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)\psi$$
(2.2)

where, importantly, the electronic energy depends on the ion positions,  $E^{\text{el}} = E^{\text{el}}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$ . We will develop later approaches to find approximate solutions for  $\psi$ .

The wave function of the total system, including the wave function of the ions is then given by

$$\Phi(\mathbf{r}_1,\ldots,\mathbf{r}_N,\mathbf{R}_1,\ldots,\mathbf{R}_N)\approx\psi_{\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{R}_N}(\mathbf{r}_1,\ldots,\mathbf{r}_N)\varphi(\mathbf{R}_1,\ldots,\mathbf{R}_N)$$
(2.3)

Here the quantum mechanical probability amplitude  $\varphi(\mathbf{R}_1, \ldots, \mathbf{R}_N)$  that ions are located at the positions  $\mathbf{R}_1, \ldots, \mathbf{R}_N$  is multiplied with the corresponding probability amplitude  $\psi_{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$  of the electrons. We now have to check whether this wave function approximately solves the full Schrödinger equation and we have to derive an equation for  $\varphi$ .

We therefore have to investigate  $H\Phi = (H_{\rm el} + H_{\rm ion} + H_{\rm el-ion})\psi\phi$ . Here the most tricky term turns out to arise from the kinetic energy of the ions as the operator  $\boldsymbol{P}_n = i\hbar\partial_{\boldsymbol{R}_n}$  acts on the  $\boldsymbol{R}_n$  coordinate both in  $\varphi$  and in  $\psi$ . This gives rise to three different terms,

$$\frac{\boldsymbol{P}_n^2}{2M_{\text{ion}}}(\psi_{\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N}(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)\varphi(\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N)) = \psi\frac{\boldsymbol{P}_n^2}{2M_{\text{ion}}}\varphi + \frac{1}{2M_{\text{ion}}}(\varphi \boldsymbol{P}_n^2\psi + 2(\boldsymbol{P}_n\varphi)\cdot(\boldsymbol{P}_n\psi))$$

The first term is the kinetic-energy Hamiltonian of ions applied to the ion wave function. We will now try to give a hand-waving argument, why we can neglect the other two terms. Can we estimate how large, e.g.,  $\frac{1}{2M_{\rm ion}}\varphi P_n^2\psi$  is? If we assume that  $\psi$  is a function of the difference of electronic and ionic coordinates  $(\boldsymbol{r} - \boldsymbol{R})$ , then we can replace the derivative by respect to an  $\boldsymbol{R}$  coordinate by a derivative with respect to a  $\boldsymbol{r}$  coordinate. Thus we can make an order of magnitude estimate  $\frac{1}{2M_{\rm ion}}P_n^2\psi \sim O(\frac{m}{M_{\rm ion}} \cdot E_{\rm kin} \cdot \psi)$ . It is therefore suppressed by the tiny factor  $\frac{m}{M_{\rm ion}}$ , i.e., by about 3 orders of magnitudes, compared to other terms which we keep in our analysis. If we neglect those and use that  $(H_{\rm el} + H_{\rm el-ion})\psi = E^{\rm el}\psi$  we obtain

$$H\Phi = H\psi\varphi \approx \psi(H_{\rm ion} + E_n^{\rm el}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N))\phi$$

Therefore we obtain the following effective eigenvalue problem for the ion coordinates only

$$(H_{\text{ion}} + \underbrace{E^{\text{el}}(\boldsymbol{R}_{1}, \dots, \boldsymbol{R}_{N})}_{\text{complicated interaction}})\phi(\boldsymbol{R}_{1}, \dots, \boldsymbol{R}_{N}) = E\phi(\boldsymbol{R}_{1}, \dots, \boldsymbol{R}_{N})$$
(2.4)

Here the kinetic energy of the ions is still small due to the large ion mass. Therefore, we can minimize in a next step only the potential energy  $H_{\text{ion}}^{pot} = \sum_{n < m} V_{\text{ion}}(\mathbf{R}_n - \mathbf{R}_m) + E^{\text{el}}(\mathbf{R}_1, \dots, \mathbf{R}_N)$ 

which includes the ion-ion repulsion and attractive potentials arising from the electronic part of the energy. While it is difficult to make any rigorous statement on such a minimization problem, we expect that the minimum of  $H_{\text{ion}}^{pot}$  is obtained by arranging the ions in a periodic structure, i.e., a crystalline lattice. In practice, one assumes that it is a lattice with a certain symmetry and then one optimizes the lattice constants and - for systems with several atoms per unit cell - the position of the atoms within the unit cell. We denote the value of the ion coordinates at the minimum by  $\mathbf{R}_n^0$ :

$$\boldsymbol{R}_n = \boldsymbol{R}_n^0$$
 minimizes  $H_{\text{ion}}^{pot} = \sum_{n < m} V_{\text{ion}}(\boldsymbol{R}_n - \boldsymbol{R}_m) + E^{\text{el}}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)$  (2.5)

The natural next step is to perform a Taylor expansion around this minimum by setting  $\mathbf{R}_n = \mathbf{R}_n^0 + \Delta \mathbf{R}_n$ . This Taylor expansion is expected to work as long as the displacements from the minimum are not too large, which is usually the case as long as the temperature is sufficiently below the melting transition. The  $\Delta \mathbf{R}_n$  parameterizes the lattice vibrations. To obtain a theory of phonons coupled to electrons one has to Taylor expand both the effective ionic Hamiltonian,  $H_{\text{ion}} + E^{\text{el}} \approx H_{\text{ion}}^0 + \underbrace{H_{\text{phonon}}}_{\text{small}}$  and the electron-ion interaction,  $H_{\text{el-ion}} \approx H_{\text{el-phonon}}^0 + \underbrace{H_{\text{el-phonon}}}_{\text{small}}$ .

But this will be covered in a separate chapter later in the lecture as we will first concentrate on the electronic problem.

#### 2.2 Hartree Fock Approximation

We now want to find a reasonable approximate solution for the electronic problem, Eq. (2.2) assuming that all ion positions have been fixed. Our strategy will be to find the 'best possible' (in a way we will define below) non-interacting particle approximation to the interacting many-particle system.

Therefore we will first consider a set of N non-interacting electrons described by some Hamiltonian,  $H_V = \sum_{i=1}^{N} H_0(\mathbf{r}_i, \mathbf{p}_i)$ . It will be important for the following that we do *not* specify here what  $H_0$  is as we want to find a kind of optimal  $H_0$  which approximates best our interacting system. As we consider Fermions the eigen functions of  $H_V$  are totally antisymmetric and can be written in terms of a so-called Slater determinant.

$$\psi_{\rm sl}(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2, \dots, \boldsymbol{r}_N\sigma_N) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \phi_1(\boldsymbol{r}_1, \sigma_1) & \phi_2(\boldsymbol{r}_1, \sigma_1) & \cdots & \phi_N(\boldsymbol{r}_1, \sigma_1) \\ \phi_1(\boldsymbol{r}_2, \sigma_2) & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \phi_1(\boldsymbol{r}_N, \sigma_N) & \cdots & \cdots & \phi_N(\boldsymbol{r}_N, \sigma_N) \end{pmatrix}, \sigma_i \in \uparrow / \downarrow$$

where  $\boldsymbol{\phi}_i = \begin{pmatrix} \phi_i(r,\uparrow) \\ \phi_i(r,\downarrow) \end{pmatrix}$  is a single-particle eigenfunction of  $H_0$ .

Our goal is to find the ground state of an interacting system with the Hamiltonian

$$H = \sum_{i} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i < j} \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{i} V(\boldsymbol{r}_i) \,.$$

The exact ground state with the exact ground-state energy can be found from a variational principle

$$E_0 = \min_{\text{all } \psi} \langle \psi | H | \psi \rangle$$

Here we minimize over all normalized wave functions  $\psi(\mathbf{r}_1\sigma_1, \mathbf{r}_2\sigma_2, \ldots, \mathbf{r}_N\sigma_N)$  which can be written as some linear combination of slater determinants, an object which is impossible even to store on a computer for any system with more than a few dozens of electrons.

To obtain an approximation, we will, however, not minimize over all many-particle wave functions, but only over those which can be written as a single Slater determinant. We use the ground state of some not-yet-specified non-interacting Hamiltonian as an variational ansatz for our interacting problem. The 'best' independent electron approximation is thereby the one which provides the lowest energy.

$$E_0 \le E_0^{\rm HF} \approx \min_{\psi_{\rm sl} = {\rm slater-det}} \langle \psi_{\rm sl} | H | \psi_{\rm sl} \rangle$$

This scheme is called a Hartree-Fock approximation (or, more precisely, a self-consistent Hartree-Fock approximation). It turns out that it is not too difficult to evaluate the expectation value. Here we just give the result

$$\langle \psi_{\rm sl} | H | \psi_{\rm sl} \rangle = \sum_{j=1}^{N} \sum_{\sigma=\uparrow/\downarrow} \int \phi_j^*(\boldsymbol{r},\sigma) \left( \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r}) \right) \phi_j(\boldsymbol{r},\sigma) \,\mathrm{d}\boldsymbol{r} + \frac{1}{2} \sum_{j'=1}^{N} \sum_{j=1}^{N} \sum_{\sigma=\uparrow/\downarrow} \left( |\phi_j(\boldsymbol{r},\sigma)|^2 |\phi_j(\boldsymbol{r}',\sigma')|^2 - \delta_{\sigma\sigma'} \phi_j^*(\boldsymbol{r},\sigma) \phi_{j'}^*(\boldsymbol{r}',\sigma') \phi_j(\boldsymbol{r}',\sigma') \phi_{j'}(\boldsymbol{r},\sigma) \right) \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \,\mathrm{d}\boldsymbol{r} \,\mathrm{d}\boldsymbol{r}'$$

$$(2.6)$$

We have to take the minimum of this expression under the boundary condition that each wavefunction is normalized. The latter condition can be achieved by using a standard trick, so-called Lagrange multipliers. This means that we are minimizing  $\langle \psi_{\rm sl} | H | \psi_{\rm sl} \rangle - \sum_{j} \epsilon_j (\langle \phi_j | \phi_j \rangle - 1)$  both with respect to the wavefunctions and with respect to the real parameters  $\epsilon_j$  (the Lagrange multiplier). As we search for a minimum, we compute  $\frac{\delta}{\delta \phi_j^*(\mathbf{r}, \sigma)} \left( \langle \psi_{\rm sl} | H | \psi_{\rm sl} \rangle - \sum_j \epsilon_j (\langle \phi_j | \phi_j \rangle - 1) \right) =$ 0. This give rise to the following equation

$$\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V(\boldsymbol{r})\right)\phi_j(\boldsymbol{r},\sigma) - e\int \frac{\rho^{\mathrm{H}}(\boldsymbol{r}') - \rho^{\mathrm{ex}}(\boldsymbol{r},\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\phi_j(\boldsymbol{r},\sigma) \,\mathrm{d}\boldsymbol{r}' = \epsilon_j\phi_j(\boldsymbol{r},\sigma) \quad (2.7)$$

This looks almost like a linear single-particle Schrödinger equation but we have hidden all complications and non-linearities in two densities,  $\rho^{\rm H}(\mathbf{r})$  and  $\rho^{\rm ex}(\mathbf{r}, \mathbf{r}')$ , defined below. The first one, the Hartree density,  $\rho^{\rm H}(\mathbf{r})$ , is easy to compute and to interpret

$$\rho^{\mathrm{H}}(\boldsymbol{r}') = -e \sum_{j',\sigma'} |\phi_{j'\sigma'}(\boldsymbol{r}')|^2$$

It describes the charge density of all electrons. From this charge density one computes in Eq. (2.7) the Coulomb potential. The physics is that each electron is affected by the Coulomb potential from all other electrons.

This is, however, not the full story. There is a correction to this which arises because electrons are indistinguishable. Therefore one cannot say that one electron is affected by the Coulomb potential from the other electrons but one has to take into account that the wave function is antisymmetric. This gives rise to the last term in Eq. (2.6) and the "Fock density"

$$ho_{j,\sigma}^{ ext{ex}}(m{r},m{r}') = -e \sum_{j'} rac{\phi_{j'}^*(m{r}',\sigma)\phi_j(m{r}',\sigma)\phi_{j'}(m{r},\sigma)}{\phi_j(m{r},\sigma)}$$

The Hartree-Fock equations (2.7) can in principle (but not in practice, see discussion below) be solved in the following way:

1. Start with guess for wave functions (eg. plane waves).

2. Determine 
$$U(\mathbf{r}) = \sum_{n} V_{\text{el-ion}}(\mathbf{r}' - \mathbf{R}_n) - e \int \frac{\rho^{\text{H}}(\mathbf{r}') - \rho^{\text{ex}}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r}'$$

- 3. Solve Schrödinger equation in potential  $U(\mathbf{r})$  and obtain new  $\phi_j(\mathbf{r}', \sigma)$ .
- 4. With the new  $\phi_j(\mathbf{r}', \sigma)$  determine the updated  $U(\mathbf{r})$  and repeat.

At first glance it seems that the Hartree term is more important than the Fock term because the total Hartree charge  $\int \rho^{\rm H}(\mathbf{r}') \, \mathrm{d}\mathbf{r}' = -Ne$  is proportional to the total number of electrons

in the system, while the Fock charge  $\int \rho^{\text{ex}}(\boldsymbol{r}, \boldsymbol{r}') \, \mathrm{d}\boldsymbol{r}' = -e$  contributes only a single electron (one can use  $\int \phi_{j'}^*(\boldsymbol{r}', \sigma) \phi(\boldsymbol{r}', \sigma) \, \mathrm{d}\boldsymbol{r}' = \delta_{jj'}$  to derive that). But this is misleading: taking also into accout the positive charge of all the ions both are equally important. The energies  $\epsilon_{j\sigma}$ also have a direct physical interpretation as ionization energies, which is the energy needed to remove a particle. The Hartree-Fock approximation is a very powerful and can be used to get the energy levels of atoms, of small molecules and is also often used in the context of simplified models of electrons where it can describe phenomena like magnetism (this is discussed later). There is, however, a major problem: a solution of the full Hartree-Fock equations for larger molecules or solids turns to be practically impossible due to severe numerical problems. The algorithm sketched above is just not converging and no good way has been found to solve the non-local and non-linear integral equations (2.7). The problem here is related to the highly non-local form of the exchange density. We therefore need to find an approximation which gives equations which on the one hand can be solved in practice and on the other hand contain at least part of the physics underlying the Hartree-Fock equation.

#### 2.3 Density Functional Theory (LDA)

#### 2.3.1 A theorem

The discussion in the following paragraph will rely heavily on the concept of a Legendre transformation. This should be familiar to most readers, e.g., from the field of analytical mechanics where Legendre transformations are used to switch between the Lagrange and the Hamiltonian formalism. In thermodynamics Legendre transformations are used to switch between different thermodynamic potentials (energy, free energy, enthalpy, ...). Let us recall the basic concepts by considering the (internal) energy E(S) as a function of the entropy S. To define the Legendre transform (thus computing the Helmholtz free energy F(T)) one proceeds in the following way:

- 1. Define and compute the temperature  $T = \frac{\partial E(S)}{\partial S}$ .
- 2. Solve this equation for S to compute S(T).
- 3. Define the Legendre transformation of the energy, the (Helmholtz) free energy as function of T: F(T) = E(S(T)) S(T)T.

Importantly, in the definition of F, the entropy S was replaced everywhere by S(T) thus that the free energy is a function of T, not of S. An important property of the Legendre transformation is obtained when one calculates its derivative:

$$\frac{\partial F}{\partial T} = \frac{\partial E}{\partial S} \frac{\partial S}{\partial T} - \frac{\partial S}{\partial T} T - S = -S$$

where we used that  $T = \frac{\partial E(S)}{\partial S}$  to cancel the first two terms. Using this equation one can compute the Legendre transformation of F, which gives back E(S). One can also define the free energy from a minimization principle. Here one starts from E(S) - ST as above but in contrast to above, one treats S and T as *independent* variables and defines

$$F(T) = \min_{S} [E(S) - ST]$$
(2.8)

The thermodynamic potential is obtained by minimizing E(S) - ST with respect to S. We can easily check that this is indeed true: taking the derivative with respect to S, we obtain  $T = \frac{\partial E(S)}{\partial S}$ , the equation defining temperature in an equilibrium state.

A rather straightforward generalization of the principles described above leads to one of the most influential theorems in physics (and chemistry), the Hohenberg-Kohn theorems which are the basis of density functional theory outlined in seminal papers of Hohenberg, Kohn (1964) and

Kohn, Sham (1965). These papers belong to the mostly cited papers in physics and chemistry of all times and Walter Kohn received the Nobel prize in chemistry in 1998.

Here is the theorem:

The ground state energy of electrons in an arbitrary potential is a functional of the local charge density:

$$E_{0} = E_{0}[\rho(\mathbf{r})] = \underbrace{\int V(\mathbf{r})\rho(\mathbf{r}) \,\mathrm{d}\mathbf{r}}_{\text{external potential (ions)}} + \frac{1}{2} \underbrace{\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \,\mathrm{d}\mathbf{r} \,\mathrm{d}\mathbf{r}'}_{\text{Hartree}} + G[\rho(\mathbf{r})]$$
(2.9)

where the exchange functional  $G[\rho(\mathbf{r})]$  is independent of the potential  $V(\mathbf{r})$ . Minimization of  $E_0[\rho]$  gives the ground state energy and its density distribution.

We will not give the original proof of the Hohenberg-Kohn theorems (which is not difficult and can be looked up on wikipedia) but instead emphasize that they can be understood as arising from a Legendre transformation. We start from the rather trivial statement, that the ground state energy of interacting electrons in the presence of a potential  $V(\mathbf{r})$  is a functional of this potential,  $E_0 = E_0[V(\mathbf{r})]$ . A 'functional' is defined as a function which maps a function (here: the potential) to a real number (here: the ground-state energy). To compute this functional

one has to determine the groundstate of the Hamiltonian  $H = H(V = 0) + \int V(\mathbf{r})\hat{\rho}(\mathbf{r}) d\mathbf{r}$ .

Following the general strategy for Legendre transformations outlined above, we first calculate the derivative of the ground state energy. How is the derivative  $\frac{\delta E_0[V(\boldsymbol{r})]}{\delta V(\boldsymbol{r}_0)}$  of a functional defined? One modifies the potential at the position  $\boldsymbol{r}_0$  by a small  $\epsilon$  and tracks the change of the total energy. This can be done using standard perturbation theory, which states that to 1. order in perturbation theory one has to calculate the expectation value of the perturbing operator. This gives

$$\frac{\delta E_0[V(\boldsymbol{r})]}{\delta V(\boldsymbol{r}_0)} = \frac{\partial E_0[V(\boldsymbol{r}) + \epsilon \delta(\boldsymbol{r} - \boldsymbol{r}_0)]}{\partial \epsilon} \mathop{=}_{1^{\text{st}} \text{order PT}} \langle \psi | \hat{\rho}(\boldsymbol{r}_0) | \psi \rangle = \rho(\boldsymbol{r}_0)$$
(2.10)

The derivative of  $E_0[V(\mathbf{r})]$  is simply the density. By inverting this relation, one can - at least in principle - obtain the potential as a (highly non-local and complicated) function of the density,  $V[\rho(\mathbf{r})]$ . With this knowledge, one can now define the Legendre transformation, precisely in the same way as we did above when we obtained F(T) from E(S). Taking the analog step yields the Legendre transform  $\tilde{E}[\rho(\mathbf{r})]$  of  $E_0[V(\mathbf{r})]$  defined by

$$\tilde{E}[\rho(\boldsymbol{r})] = E_0[V[\rho(\boldsymbol{r})]] - \int V[\rho(\boldsymbol{r})](\boldsymbol{r}')\rho(\boldsymbol{r}') \,\mathrm{d}\boldsymbol{r}$$

We also have,

$$\frac{\delta \tilde{E}_0}{\delta \rho} = \int \frac{\delta E_0}{\delta V} \frac{\delta V}{\delta \rho} - \int \frac{\delta V}{\delta \rho} \rho - V$$

In direct analogy to Eq. 2.8, we furthermore define

$$E_0[V] = \min_{\rho} \left[ \int V(\boldsymbol{r})\rho(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + \underbrace{\tilde{E}_0[\rho]}_{\text{independent of } V} \right]$$

We can easily check from the definition that at the minimum the equation  $\frac{\delta \tilde{E}_0}{\delta \rho(\mathbf{r}_0)} = V(\mathbf{r}_0)$ holds. Using furthermore the definition of  $\tilde{E}$ , we see that the minimization indeed yields the groundstate energy and the exact electron density in the ground state. Finally, we obtain (2.9) by defining  $G[\rho(\mathbf{r})] = \tilde{E}[\rho(\mathbf{r})] - \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$  which concludes the derivation of the theorem.

#### 2.3.2 Local density approximation and Kohn-Sham equations

The theorem is exact but it has a major problem: the functional  $G[\rho(\mathbf{r})]$  is not known. In its exact form it much be highly non-linear and non-local as it should be able to describe, e.g., the exact properties of phase transitions occuring at zero temperature. But our goal is not to obtain an exact solution but only a reasonable approximation. Here the key idea is that all the truely long-ranged aspect of Coulomb interactions is already captured by the Hartree energy and the remaining physics is 'more or less' local. Therefore our plan will be to make a Taylor expansion in space to approximate the functional  $G[\rho(\mathbf{r})]$ . How do we do that? If there is, for example, a nonlocal term proportional to  $\rho(\mathbf{r}) - \rho(\mathbf{r}')$  we Taylor expand and approximate  $\rho(\mathbf{r}') \approx \rho(\mathbf{r}) + (\mathbf{r}' - \mathbf{r})\nabla\rho(\mathbf{r}) + \dots$  Assuming that such an expansion is well-behaved (which is far from obvious), we can at least in principle write our functional (after subtracting the kinetic energy,  $E_{kin}$ , see below) as a simple integral  $G[\rho] = \int g(\rho(\mathbf{r}), \nabla \rho, \nabla \nabla \rho, \ldots)$  over a (still unknown) function which depends on the density and its derivatives. Now comes the central approximation: we neglect all gradients and assume that we have only a function of the local density (without gradients)

$$G[\rho] = E_{kin} + \int g(\rho(\boldsymbol{r}), \boldsymbol{\nabla}\rho, \boldsymbol{\nabla}\boldsymbol{\nabla}\rho, \dots) \, \mathrm{d}\boldsymbol{r} \approx E_{kin} + \int \underbrace{E_{\mathrm{exch}}(\rho(\boldsymbol{r}))}_{\mathrm{exchange}} d\boldsymbol{r}.$$
(2.11)

This very influential approximation is called **Local Density Approximation** (LDA).

Note that we still do not know the function  $E_{exch}(\rho(\mathbf{r}))$  which is called "exchange" term. But before we investigate this question, we discuss an algorithm how for a given  $g(\rho(\mathbf{r}))$  one can calculate the groundstate energy by minimizing our functional in practice. For this, we first split the energy in a kinetic part and a so-called exchange term is indicated in Eq. (2.11). To implement the variation we use a trick which greatly simplifies the calculation of the kinetic energy contribution. In the spirit of a single-particle approximation we parametrize the density by the wave function of singly occupied single-particle states

$$ho(m{r}) = \sum_{i, ext{occupied}} |\phi_i(m{r})|^2$$

The main adaptage of this that it allows for a simple calculation of the kinetic energy. Now we proceed in the same way as in the previous chapter. We minimize  $E_0$  using the constraint  $\int |\phi_j|^2 d\mathbf{r} = 1$  for each single-particle wave function. This is implemented by a Lagrange multiplier. Therefore we have to calculate

implier increase to have to careatate

$$\frac{\delta}{\delta\phi_j^*}\left(E_0 - \sum_j \epsilon_j(\langle\phi_j|\phi_j\rangle - 1)\right) = 0$$

`

to obtain the minimum. This leads to the famous **Kohn-Sham equations** which read  $\frac{\delta E_0}{\delta \rho} \phi_j(\mathbf{r}) = \epsilon_j \phi_j$  or written out fully

$$\begin{pmatrix} -\frac{\hbar^2}{2m} \nabla^2 + V_{\text{eff}}(\boldsymbol{r}) \end{pmatrix} \phi_j(\boldsymbol{r}) = \epsilon_j \phi_j(\boldsymbol{r}), \\ V_{\text{eff}}(\boldsymbol{r}) = V(\boldsymbol{r}) - e \int \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} e^2 \, \mathrm{d}\boldsymbol{r}' + \frac{\partial E_{\text{exch}}(\rho(\boldsymbol{r}))}{\partial \rho}$$

Here the first term arises from the variation of  $E_{kin}$ . Above, we cheated a bit because we did not give a proper definition of  $E_{kin}$  and did not explain how this is related to the kinetic energy in the Kohn-Sham equation. Indeed, some extra work is required to show that the Kohn-Sham equation do give the correct value of the kinetic energy within the spirit of the LDA approximation.

The Kohn-Sham equation can be solved in a rather straightforward way. One can start with some guess for the wave functions and use this guess to calculate  $\rho(\mathbf{r})$  and the effective potential. Then one solves the resulting single-particle Schrödinger equation and determines from the new wavefunctions a revised potential. Iterating this a few times, one obtains a converged solution of the Kohn-Sham equations. Importantly, the convergence properties of these local equations is much better compared to the non-local Hartree Fock equations. In practice, the Kohn-Sham equations (and variants thereof) can be solved even for complex molecules and solids with complicated unit cells.

Formally, our starting point was a theory formulated only for the electron density  $\rho(\mathbf{r})$ . The wavefunctions  $\phi_i(\mathbf{r})$  and the paremeters  $\epsilon_i$  were only introduced as a mathematical trick which no obvious direct physical meaning. Nevertheless, the form of the Kohn-Sham equation suggests to interpret them directly as wavefunctions and single-particle energies. While one should keep in mind, that this is an interpretation not really justified and thus has its limitations, this interpretation turns out to be extremely useful and quantitatively successful to describe a broad range of experiments which measure quantities like the electronic dispersion of the shape of the Fermi surface (see below). Partially this success can be understood from the relation of the LDA approximation to the Hartree-Fock approch which we will consider next.

#### 2.3.3 Exchange energy

As a last step in our derivation of density functional theory, we have to derive a good approximation for the exchange energy  $E_{\text{exch}}(\rho(\mathbf{r}))$  as function of the local density. This turns out to be more simple than it sounds because we can use that - by construction - the density functional is the same for each external potential  $V(\mathbf{r})$ . Therefore we can determine it by considering the simplest possible potential, namely  $V(\mathbf{r}) = 0$  (more precisely, one assumes that the positive charge density of the ions is constant in space). In this case the electronic density is also constant in space,  $\rho(\mathbf{r}) = \rho$ . Our task is therefore to calculate for such a model the total energy. After subraction of the kinetic energy and the Hartree term (which cancels exactly with the constant positive background charge), one obtains the exchange energy as function of  $\rho$ .

To obtain an analytical formula, we will use the self-consistent Hartree-Fock approximation. While we argued that the self-consistent Hartree-Fock equations extremely difficult to evaluate in general, it is actually very simple in the absence of an external potential. We can use that we know in this case already the single-particle wavefunctions. They are given by plane waves

$$\phi_n(\boldsymbol{r}) = \frac{1}{\sqrt{V}} e^{i\boldsymbol{k}_n \cdot \boldsymbol{r}}$$

where V is the volume of the system. To make everything well defined, we use periodic boundary conditions  $\phi_i(\mathbf{r} + \mathbf{L}) = \phi_i(\mathbf{r})$  which implies that only a discrete set of lattice vectors

are allowed

$$\boldsymbol{k}_n = rac{2\pi}{L} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}, \; n_i \in \mathbb{Z}, V = L^3.$$

In the ground state all  $\mathbf{k}_n$  up to a maximal value are occupied,  $|\mathbf{k}_n| \leq k_{\rm F}$ . The maximal momentum is called Fermi momentum. The first step is to calculate the local density for such a setup.

$$\rho = \underbrace{\sum_{|\mathbf{k}_n| < k_{\rm F}}}_{|\mathbf{k}_n| < k_{\rm F}} \left| \frac{1}{\sqrt{V}} e^{i\mathbf{k}_n \cdot \mathbf{r}} \right|^2 = 2\left(\frac{1}{2\pi}\right)^3 \left(\frac{2\pi}{L}\right)^3 \sum_{|\mathbf{k}_n| < k_{\rm F}} 1$$
$$= 2\int_{|\mathbf{k}| < k_{\rm F}} \frac{1}{(2\pi)^3} \, \mathrm{d}\mathbf{k} = \frac{2 \times 4\pi}{(2\pi)^3} \int k^2 \, \mathrm{d}k = \frac{k_{\rm F}^3}{3\pi^2}.$$
(2.12)

In the second line, we used that the volume is large, thus the sum can be replaced by an integral using that the distance of neighboring k-values is given by  $\frac{2\pi}{L}$  in x,y and z directions.

The next step is to calculate the correction to the groundstate energy arising from the interactions beyond the Hartree term, which is within the Hartree-Fock approximation the Fock term, defined by the second term in the second line of Eq.(2.6). Therefore the exchange energy is given by

$$\begin{split} \int & E_{\text{exch}} \, \mathrm{d}\boldsymbol{r} = -\sum_{n,n'} \frac{2}{2} \iint \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \phi_n^*(\boldsymbol{r}) \phi_{n'}^*(\boldsymbol{r}) \phi_n(\boldsymbol{r}') \phi_{n'}^*(\boldsymbol{r}') \, \, \mathrm{d}\boldsymbol{r} \, \mathrm{d}\boldsymbol{r}' \\ &= -\int_{|\boldsymbol{k}| < k_{\mathrm{F}}} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \int_{|\boldsymbol{k}'| < k_{\mathrm{F}}} \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} \iint \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} e^{i(\boldsymbol{k} + \boldsymbol{k}') \cdot (\boldsymbol{r} - \boldsymbol{r}')} \, \, \mathrm{d}\boldsymbol{r} \, \mathrm{d}\boldsymbol{r}' \\ &= -V \int_{|\boldsymbol{k}| < k_{\mathrm{F}}} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \int_{|\boldsymbol{k}'| < k_{\mathrm{F}}} \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} \frac{e^{24\pi}}{(k'+k')^2} \\ &= -V k_{\mathrm{F}}^4 e^2 \left( \left(\frac{1}{2\pi}\right)^6 \int_0^1 \mathrm{d}x' \int_0^1 \mathrm{d}x \frac{(4\pi)^3}{2} \int \mathrm{d}(\cos\phi) \frac{1}{x^2 + x'^2 - 2x'x\cos\phi} \right) \\ &= V e^2 k_{\mathrm{F}}^4 \frac{1}{4\pi^3} \,. \end{split}$$

We can now use Eq. (2.12) to replace  $k_F$  by the density and find

$$E_{\text{exch}} = -e^2 \frac{1}{4\pi^3} k_{\text{F}}^4 = -\frac{e^2}{4\pi} (3\pi^2 \rho)^{4/3} = -\frac{3}{4\pi} \rho (3\pi^2 \rho)^{1/3}.$$

The corresponding potential which shows up in the Kohn-Sham equations is therefore given by

$$V_{\text{exch}}(\rho) = \frac{\partial E_{\text{exch}}}{\partial \rho} = -\frac{1}{\pi} (3\pi^2 \rho)^{1/3}$$
(2.13)

This concludes our derivation of density functional theory. With the local density approximation as defined above one already has a remarkably powerful approach which despite its simplicity can be used for quantitative predictions in a broad set of materials. The non-analytic exchange potential is - somewhat suprisingly - able to capture at least semi-quantitative the most important part of interaction effects which go beyond the simple Hartree approximation. LDA can be improved in various ways, for example by including corrections due to gradients proportional to  $\frac{(\boldsymbol{\nabla}\rho)^2}{\rho^{\frac{4}{3}}}$  or by considering corrections arising from relativistic effects. LDA and its variants are implemented in ready-to-use programs with names like Quantum Espresso, VASP or Wien2K, some powerful codes are developed at the research center in Jülich, http://www.judft.de/. The input for these programs can, e.g., be an approximate crystalline structures with approximate positions of the relevant ions, from which one can compute not only the ground state energy (which is then used to optimze the crystalline structure) but also the band structure of the material (dicussed in the next section).

Despite its success in calculating ground-state energies of many systems, LDA remains a crude approximation and there are many cases where it fails also completely. These are typically systems where interactions are really strong (often related to atoms with partially occupied d or f shells). There are, for example, some insulators with very large gaps (vanadium oxide being one example) which are predicted to be good metals by LDA. A modern development is therefore the combination of LDA-style methods with other field-theoretical techniques like dynamical mean-field theory (DMFT) tailored to treat strong local interactions.

## ELECTRONS IN A PERIODIC POTENTIAL

When studying the physics of electrons in a crystal, the first task is to find the eigenstates of electrons in a periodic potential  $U(\mathbf{r})$ ,

$$H\psi = E\psi, H = -rac{oldsymbol{
abla}^2}{2m} + U(oldsymbol{r})$$

with

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}_n) \quad \forall \mathbf{R}_n \in \text{ Bravais lattice}$$
(3.1)

For simplicity, we neglect relativistic corrections like spin-orbit interactions here, which can, however, be easily added in the following discussions.

For example, when we solve the Kohn-Sham equations for a crystal, the effective potential  $U(\mathbf{r})$  obtains contributions from the ion potentials, from the Coulomb interaction of all the other electrons, and from the exchange potential calculated, e.g., within the local density approximation.

#### 3.1 Bloch's Theorem

The discrete translational invariance of a potential strongly restricts the form of eigenstates  $\psi$  of the Hamilton operator. This is the content of an important theorem going back to Bloch (1928) but similar theorems had been discovered before (e.g., by Hill 1877 or by Floquet in 1883).

**Bloch's theorem:** The eigenstates of  $H_0 = -\frac{\hbar^2 \nabla^2}{2m} + U(\mathbf{r})$  with a periodic potential  $U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}_n) \quad \forall \mathbf{R}_n \in \text{Bravais lattice, can be chosen as}$ 

$$\psi_{m\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{m\mathbf{k}}(\mathbf{r}) \quad \text{with} \quad u_{m\mathbf{k}}(\mathbf{r} + \mathbf{R}_n) = u_{m\mathbf{k}}(\mathbf{r})$$
(3.2)

Here  $\mathbf{k}$  is in the 1. Brillouin zone,  $\mathbf{k} \in 1$ . BZ, and m is a band index. The eigenenergies  $\epsilon_m(\mathbf{k})$  with  $H\psi_{m\mathbf{k}} = \epsilon_m(\mathbf{k})\psi_{m\mathbf{k}}$  are periodic functions in reciprocal space

$$\varepsilon_m(\mathbf{k} + \mathbf{G}_n) = \varepsilon_{m'}(\mathbf{k}), \tag{3.3}$$

where  $G_n$  is a reciprocal lattice vector.

We have used two different indices, m and m' in Eq. (3.3) as the scheme how bands are numbered may be different at the two edges of the 1.BZ.

The theorem is so important that we provide two different proof. The first one is constructive and can actually be used to compute in a rather efficient way the eigenstates. The second one emphasizes more the underlying group theory.

The first proof starts from a simple Fourier decomposition of the wavefunction,  $\psi(\mathbf{r}) = \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}}c_{\mathbf{k}}$ , and the potential,  $U(\mathbf{r}) = \sum_{\mathbf{G}_n} e^{iG_n\cdot\mathbf{r}}U_{\mathbf{G}_n}$ , where importantly only reciprocal lattice

vectors  $G_n$  occur. We plug this into the Schrödinger equation,  $\left(-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right)\psi(\mathbf{r}) = E\psi(\mathbf{r})$ , and obtain

$$\sum_{\boldsymbol{q}} \frac{\hbar^2 \boldsymbol{q}^2}{2m} c_{\boldsymbol{q}} e^{i\boldsymbol{q}\cdot\boldsymbol{r}} + \sum_{\boldsymbol{q},\boldsymbol{G}_n} U_{\boldsymbol{G}_n} c_{\boldsymbol{q}} e^{i(\boldsymbol{q}+\boldsymbol{G}_n)\cdot\boldsymbol{r}} = E \sum_{\boldsymbol{q}} c_{\boldsymbol{q}} e^{i\boldsymbol{q}\cdot\boldsymbol{r}}$$

Now we want to pick up the components oscillating with the momentum  $q_0 = G_m + k, k \in$ 1. BZ,  $G_m \in$  reciprocal lattice. Formally, this is done multiplying this equation with  $e^{-i(G_m+k)r}$ and integrating over space. From the first and last term only  $q = G_m + k$ , from the second term only  $q + G_n = G_m + k$  survives. This leads to the Schrödinger equation in momentum space,

$$\left[\frac{\hbar^2 (\boldsymbol{k} + \boldsymbol{G}_m)^2}{2m} - E\right] c_{\boldsymbol{k} + \boldsymbol{G}_m} + \sum_{\boldsymbol{G}_n} U_{\boldsymbol{G}_n} c_{\boldsymbol{k} + \boldsymbol{G}_m - \boldsymbol{G}_n} = 0$$
(3.4)

for each  $\mathbf{k} \in 1.$  BZ fixed and for each  $\mathbf{G}_m$ . Importantly, the equation connects only momenta of the form  $\mathbf{k} + \mathbf{G}_{m'}$  which are connected by reciprocal lattice vectors. Therefore, for fixed  $\mathbf{k} \in 1.$  BZ, we can cast the Schrödinger equation into a matrix equation

$$\sum_{n} M_{mn}^{\boldsymbol{k}} c_{\boldsymbol{k}+\boldsymbol{G}_{n}} = E c_{\boldsymbol{k}+\boldsymbol{G}_{m}} \quad \text{with} \quad M_{mn}^{\boldsymbol{k}} = \delta_{mn} \frac{\hbar^{2}}{2m} (\boldsymbol{k}+\boldsymbol{G}_{m})^{2} + U_{\boldsymbol{G}_{m}-\boldsymbol{G}_{m}}$$

Diagonal entries are given by the kinetic energy, offdiagonal ones by the Fourier transformation of the potential.

The rest of the proof is simple: the Hermitian matrix  $M^{k}$  has discrete eigenvalues  $\varepsilon_{j}(k)$ ,  $j \in \mathbb{N}$ , and we call j the "band-index". The various elements of the corresponding eigenvectors are denoted by  $c_{k+G_{n}}^{j}$  in the following. From them, we can easily reconstruct the eigenfunctions in real space

$$\psi_{j\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}_n} e^{i(\boldsymbol{k}+\boldsymbol{G}_n)\cdot\boldsymbol{r}} c^j_{\boldsymbol{k}+\boldsymbol{G}_n} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{j\boldsymbol{k}}(\boldsymbol{r})$$
  
where  $u_{j\boldsymbol{k}}(\boldsymbol{r}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \psi_{j\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}_n} e^{i\boldsymbol{G}_n\cdot\boldsymbol{r}} c^j_{\boldsymbol{k}+\boldsymbol{G}_n} = u_{j\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}_n)$ 

Thus we have successfully constructed the eigenfunctions with the desired property. Also the fact that  $\varepsilon_m(\mathbf{k} + \mathbf{G}_{n_0}) = \varepsilon_m(\mathbf{k})$  follows directly from our construction as the matrices  $M_{mn}^{\mathbf{k}}$  and  $M_{mn}^{\mathbf{k}+\mathbf{G}_{n_0}}$  coincide after a trivial shift of indices which map  $\mathbf{G}_m + \mathbf{G}_{n_0}$  back to  $\mathbf{G}_m$ . The proof is constructive and for relatively smooth potentials with only a few Fourier components one can directly implement it numerically by approximating the nominally infinite matrix  $M_{mn}^{\mathbf{k}}$  by

a finite matrix where one takes into account reciprocal lattice vector up to a maximal length,  $|G_m| < g_{max}$ .

The **second proof** uses the group of translations. Consider the unitary operator  $T_{\mathbf{R}_n}$  which introduces translations by a vector of the Bravais lattice,  $T_{\mathbf{R}_n}\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}_n)$ . Obviously the discrete translations form an abelian group with  $T_{\mathbf{R}_n}T_{\mathbf{R}_m} = T_{\mathbf{R}_n+\mathbf{R}_m}$ . As

 $[H_0, T_{\mathbf{R}_n}] = 0$  and  $[T_{\mathbf{R}_n}, T_{\mathbf{R}_m}] = 0$  one can simultaneously diagonalize all these operators. Consider one common eigenfunction of these operators

 $\psi$  and let us study its eigenvalues for translations by a basis vector,  $T_{a_i}\psi = \tau_i\psi$ . For periodic boundary conditions, translations are unitary operators and one can therefore expect that eigenvalues have the property that  $|\tau_i| = 1$  (for periodic boundary conditions there cannot be wavefunctions which are exonentially growing or decaying). Thus we can write without loss of generality  $\tau_i = e^{\mathbf{k} \cdot \mathbf{a}_i}$ . Doing this for the three basis vectors fixes the vector  $\mathbf{k}$ uniquely within the first Brillouin zone. Now we can easily define  $u(\mathbf{r}) = e^{-i\mathbf{k} \cdot \mathbf{r}}\psi(\mathbf{r})$  and check that  $T_{a_i}u = e^{-i\mathbf{k} \cdot \mathbf{a}}u = u$  is indeed a periodic function,  $T_{\mathbf{R}_n}u = u$ .

Let us add a few remarks:

- $\varepsilon_{mk}$  is an analytic functions of k at least for most values of k as it arises from the discrete eigenvalues of an analytic matrix  $M^k$ . Only when degeneracies occur the energy could be a non-analytic function.
- $\varepsilon_{mk}$  is an analytic functions of k at least for most values of k as it arises from the discrete eigenvalues of an analytic matrix  $M^k$ . Only when degeneracies occur the energy could be a non-analytic function.
- One can either define  $\varepsilon_{nk}$  as a function of k only in in the 1. BZ, or, alternatively, as a periodic function in reciprocal space,  $\varepsilon_{nk} = \varepsilon_{n,k+G_n}$ .
- The quantum number  $\mathbf{k}$  is **not** the momentum,  $\mathbf{p}\psi_{m\mathbf{k}} = -i\hbar\nabla\psi_{m\mathbf{k}} = \hbar\mathbf{k}\psi_{m\mathbf{k}} + e^{i\mathbf{k}\cdot\mathbf{r}}\frac{\hbar}{i}\nabla u_{m\mathbf{k}}(\mathbf{r})$ . One calls  $\hbar\mathbf{k}$  the "crystalline momentum" which turns out to be conserved only modulo reciprocal lattice vectors,  $\mathbf{G}_n$ .

#### 3.2 Electrons in a Weak Periodic Potential

In the general case, we have to use numerical methods to calculate the bandstructure. Only in certain limits can we expect to make analytical progress. One such limit is to consider weak periodic potentials. In this limit one can use perturbation theory. Most importantly, the qualitative insights we get from this calculation can also be used in cases where perturbation theory is not possible. We consider a periodic potential:  $U(\mathbf{r}) = \sum_{\mathbf{G}_n} e^{\mathbf{G}_n \cdot \mathbf{r}} U_{\mathbf{G}_n}$ , which is "weak" (in a sense defined below). To simplify notations, we consider the case with  $U_{\mathbf{G}=\mathbf{0}} = \int U(\mathbf{r}) \, \mathrm{d}\mathbf{r} = 0$  which we can always achieve by subtracting a trivial constant from the potential.

As we plan to do perturbation theory, let us start with a short reminder how this works. For a Hamiltonian  $H = H_0 + \Delta H$ , one first assumes that eigenfunctions and eigenvectors of  $H_0$  are known,  $H_0|m^0\rangle = E^0|m^0\rangle$ . Then one can derive perturbation theory simply doing a Taylor expansion in powers of the perturbation using  $(H_0 + \Delta H)(|m^0\rangle + |m^1\rangle + ...) = (E^0 + E^1 + ...)(|m^0\rangle + |m^1\rangle + ...)$ . This gives rise the following shift of the energies

$$E_m = E_m^0 + \langle m^0 | \Delta H | m^0 \rangle - \sum_{i \neq m} \frac{|\langle i^0 | \Delta H | m^0 \rangle|^2}{E_i^0 - E_m^0} + \mathcal{O}((\Delta H)^3).$$

This will work as long as the perturbation is small and as long as there are no degeneracies. In the latter case, one has to use degenerate perturbation theory (see below).

Back to our original problem which we want to analyze using the Schrödinger equation (3.4) in momentum space. We have to determine the eigenvalues of the matrix

$$M = \delta_{mn} \frac{\hbar^2}{2m} (\boldsymbol{k} + \boldsymbol{G}_m)^2 + U_{\boldsymbol{G}_m - \boldsymbol{G}_m}$$

for  $\mathbf{k} \in 1.BZ$ . For U = 0, the matrix is diagonal and has eigenvalues  $E^0 = \frac{\hbar^2}{2m} (\mathbf{k} + \mathbf{G}_m)^2$ with trivial eigenfunctions:  $c^0_{\mathbf{k}+\mathbf{G}_n} = \delta_{nm}$  corresponding to plane waves,  $\psi = e^{i(\mathbf{k}+\mathbf{G}_m)\cdot\mathbf{r}}$ . All formulas above are given for a three-dimensional case, but it is useful to think about the more simple one-dimensional periodic system with lattice constant a, where reciprocal lattice vectors  $G_n = \frac{2\pi}{a}n$ . In this case the 1. BZ is given by  $-\frac{\pi}{a} \leq k < \frac{\pi}{a}$  and we obtain the following Fig. 3.1.



**Figure 3.1:** Bandstructure of a free particle in 1d described using  $k \in 1$ . BZ and the band-index (1, 2, 3, 4 etc.). The size of the 1. BZ is determined by the size, a, of the unit cell.

Now let us consider finite but small U and first try standard perturbation theory in U. As above, we use a Taylor expansion of energies ,  $E = E^0 + E^1 + E^2 + \ldots$ , and the wave function denoted by  $c = c^0 + c^1 + \ldots$ . We perturb around  $c^0_{\mathbf{k}+\mathbf{G}_i} = \delta_{im}$  collecting the terms linear in U in the Schrödinger equation  $M\mathbf{c} = E\mathbf{c}$ . We obtain the equation  $E^1c^0_{\mathbf{k}+\mathbf{G}_i} + E^0_{\mathbf{k}+\mathbf{G}_m}c^1_{\mathbf{k}+\mathbf{G}_i} = \delta_{ij}E^0_{\mathbf{k}+\mathbf{G}_j}c^1_{\mathbf{k}+\mathbf{G}_j} + U_{\mathbf{G}_i-\mathbf{G}_j}c^0_{\mathbf{k}+\mathbf{G}_j}$ . For  $i \neq m$  the  $E_1$  term drops and we find  $c^1_{\mathbf{k}+\mathbf{G}_i} = \frac{U_{\mathbf{G}_i-\mathbf{G}_m}}{E^0_{\mathbf{k}+\mathbf{G}_i} - E^0_{\mathbf{k}+\mathbf{G}_i}}$ . This result can also be used to calculate the shift of energy to second order in the potential which gives

$$E = E_{k+G_n}^0 - \sum_{G_i \neq G_n} \frac{|U_{G_i-G_n}|^2}{E_{k+G_i}^0 - E_{k+G_n}^0} + \mathcal{O}(U^3)$$

Importantly, we see that the perturbation theory breaks down when  $|U_{\mathbf{G}_i-\mathbf{G}_n}| \gtrsim |E_{\mathbf{k}+\mathbf{G}_i}^0 - E_{\mathbf{k}+\mathbf{G}_n}^0|$ . Let us assume that for a given  $\mathbf{k}$  and m this happens for a set of n reciprocal

vectors  $G_i \in {\{\tilde{G}_1, \ldots, \tilde{G}_n\}}$ . We will use **degenerate perturbation theory** for these cases but ordinary perturbation theory for the rest.

How does one do degenerate perturbation theory? One treats those modes exactly, keeping in (3.4) only the *n* 'dangerous' modes  $c_{\mathbf{k}+\tilde{\mathbf{G}}_i}$ . For all the other modes we can either use perturbation theory or just ignore them. Keeping them, we obtain *n* equations for the *n* modes

$$(E - E_{\mathbf{k}+\mathbf{G}_{i}}^{0})c_{\mathbf{k}+\mathbf{G}_{i}} = \sum_{j=1}^{n} U_{\tilde{\mathbf{G}}_{i}-\tilde{\mathbf{G}}_{j}}c_{\mathbf{k}+\tilde{\mathbf{G}}_{j}} + \sum_{\substack{\mathbf{G}\neq\tilde{\mathbf{G}}_{1},\ldots\tilde{\mathbf{G}}_{n}}} U_{\tilde{\mathbf{G}}_{i}-\mathbf{G}}c_{\mathbf{k}+\mathbf{G}}$$

$$\approx \sum_{j=1}^{n} U_{\tilde{\mathbf{G}}_{i}-\tilde{\mathbf{G}}_{j}}c_{\mathbf{k}+\tilde{\mathbf{G}}_{j}} + \underbrace{\sum_{\substack{\mathbf{G}\neq\tilde{\mathbf{G}}_{1},\ldots\tilde{\mathbf{G}}_{n}}} U_{\tilde{\mathbf{G}}_{i}-\mathbf{G}}U_{\mathbf{G}-\tilde{\mathbf{G}}_{j}}}_{\text{can be neglected if smaller than } U_{\tilde{\mathbf{G}}_{i}-\tilde{\mathbf{G}}_{j}}} + \mathcal{O}(U^{3})$$

One can also view this as a  $n \times n$  matrix equation which has to be diagonalized to treat the coupling of the *n* almost degenerate levels exactly.

Under what condition do we obtain a degeneracy? Obviously when  $(\mathbf{k} + \mathbf{G}_i)^2 = (\mathbf{k} + \mathbf{G}_j)^2$ , i.e. on a plane (called *Bragg-plane*) which has the equal distance from two vectors in reciprocal space. This is the case on all boundaries of 1. BZ where n = 2. At edges and corners of the 1.BZ one obtains larger values of n.

Let us focus on the most simple and most important case, n = 2. We have to consider a  $2 \times 2$  matrix of the form

$$M = \begin{pmatrix} E_{\boldsymbol{k}}^0 & U_{\boldsymbol{G}} \\ U_{\boldsymbol{G}}^* & E_{\boldsymbol{k}+\boldsymbol{G}}^0 \end{pmatrix}$$

which has eigenvalues

$$E_{\pm}^{0} = \frac{1}{2} \left( E_{k}^{0} + E_{k+G}^{0} \right) \pm \sqrt{\left( \frac{E_{k}^{0} - E_{k+G}^{0}}{2} \right)^{2} + |U_{G}|^{2}}$$

At the degeneracy point, one has  $E^0_{\mathbf{k}} = E^0_{\mathbf{k}+\mathbf{G}}$  and the splitting of bands is **linear** in U due to the Bragg reflection

$$E_{\pm}^{0} = E_{k}^{0} \pm |U_{G}| \tag{3.5}$$

An example of such a band in one dimension can be found in Fig. 3.2.



**Figure 3.2:** Splitting of bands in a monoatomic 1d lattice. The band gap opens to linear order in  $U_k$  due to Bragg reflection from  $k = \pm G, \pm 2G$  etc. where  $|G| = \frac{2\pi}{a}$ , a being the lattice spacing.

Above, we used a "reduced zone scheme", considering only momenta in the 1. BZ. Sometimes it is usefull to continue all bands periodically, in this case one talks about a "periodic" or "repeated" or "extended" zone scheme.



**Figure 3.3:** Repeated Zone Scheme for describing bandstructure is obtained by a periodic extension of the band energies in the first Brillouin zone.

To visualize band structures in more than one dimension, one often plots the bands  $E_{k}^{m}$  along certain high-symmetry lines in the 1. BZ.

#### 3.3 Fermi Surfaces

We continue by discussing basic properties of non-interacting electron using the fact that within Hartree-Fock or within the LDA approximation we indeed obtain an effective description

in terms of non-interacting particles. Later we will give more precise arguments why noninteracting electrons work remarkably well to describe many properties of solids even in the presence of interactions. But for the moment, we will just assume that we have done some bandstructure calculation and have obtained the single-particle energies  $\varepsilon_{mk}$  where  $m \in \mathbb{N}$  is the band index and  $k \in 1$ . BZ.

To find the groundstate of the many-particle system we use Pauli's principle. As an electron carries spin, each state with quantum numbers m and k can be occupied by an up and a down-electron. The many-particle groundstate is obtained by occupying all states up to a maximal energy, called the Fermi energy.

occupied states:  $\varepsilon_{m\mathbf{k}} \leq \varepsilon_{\mathrm{F}}, \mathbf{k} \in 1.\mathrm{BZ}$ Fermi energy:  $\varepsilon_{\mathrm{F}}$ electron spin: each  $m, \mathbf{k}$  occupied with  $\uparrow, \downarrow$ 

Let us do this more precisley. For this we use periodic boundary conditions

$$\psi_{m\boldsymbol{k}}(\boldsymbol{r}+N_j\boldsymbol{a}_j)=\psi_{m\boldsymbol{k}}(\boldsymbol{r}), j=1,2,3, N_j\in\mathbb{N}, N_j\gg 1$$

Here  $N_j$  is the number of unit cells in the direction of  $\boldsymbol{a}_j$ . From Bloch's theorem, we know that  $\psi_{m\boldsymbol{k}}(\boldsymbol{r}+N_j\boldsymbol{a}_j) = e^{i\boldsymbol{k}\cdot N_j\boldsymbol{a}_j}\psi(\boldsymbol{r})$  which allows to conclude that  $\boldsymbol{k}\cdot\boldsymbol{a}_j = \frac{m_j}{N_j}2\pi$ ,  $m_j \in \mathbb{Z}$ . This condition can rewritten using our definition for the basis vector,  $\boldsymbol{G}_1, \boldsymbol{G}_2, \boldsymbol{G}_3$  of the reciprocal lattice,  $\boldsymbol{G}_j \cdot \boldsymbol{a}_i = \delta_{ij}2\pi$ . We find that the following discrete momenta are allowed

$$\boldsymbol{k_m} = \sum_{j=1}^{3} \frac{m_j}{N_j} \boldsymbol{G}_j, \quad \mathbf{m} \in \mathbb{Z}, \quad k \in 1.BZ,$$

In the limit of large  $N_j$  the grid of momenta becomes denser and denser. We will therefore often rewrite the sum over the discrete momenta  $\sum_{k} \ldots$  as an integral. For this we need the volume around each  $k_m$  point which is given by

$$\frac{\boldsymbol{G}_1 \cdot (\boldsymbol{G}_2 \times \boldsymbol{G}_3)}{N_1 N_2 N_3} \tag{3.6}$$

Let us calculate this using that  $G_i = 2\pi \epsilon_{ijk} \frac{(a_j \times a_k)}{a_i \cdot (a_j \times a_k)}$ . After a bit of algebra we obtain that

$$\boldsymbol{G}_1 \cdot (\boldsymbol{G}_2 \times \boldsymbol{G}_3) = \frac{(2\pi)^3}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}$$
(3.7)

Using the volume per discrete k-point (3.6), we can approximate

$$\sum_{\boldsymbol{k}} \cdots \approx \frac{N_1 N_2 N_3}{\boldsymbol{G}_1 \cdot (\boldsymbol{G}_2 \times \boldsymbol{G}_3)} \int_{1.BZ} \dots d\boldsymbol{k} = N_1 N_2 N_3 \underbrace{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)}_{\text{volume of unit cell}_{1.BZ}} \int_{1.BZ} \dots \frac{d\boldsymbol{k}}{(2\pi)^3} = V \int_{1.BZ} \dots \frac{d\boldsymbol{k}}{(2\pi)^3}$$

Identifying  $N_1 N_2 N_3 \underbrace{a_1 \cdot (a_2 \times a_3)}_{\text{volume of unit cell}}$  with the total volume V of the system, we obtain the following frequently used formula

$$\sum_{\boldsymbol{k_m}} f(\boldsymbol{k_m}) = V \int_{1.BZ} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} f(\boldsymbol{k}).$$
(3.8)

where f is some arbitrary function.

As a first example, let us compute the total number of particles  $N_e$  and the density of electrons  $n = \frac{N_e}{V}$  at zero temperature by counting all occupied states,  $N_e = 2 \sum_{\varepsilon_{mk} < \varepsilon_{\rm F}} 1$ , where the factor 2 arises from the spin.

$$\sum_{\substack{m \\ \boldsymbol{k} \in 1. \text{BZ} \\ \varepsilon_{m \boldsymbol{k}} < \varepsilon_{\text{F}}}} \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} = \frac{N_e}{V} = n \quad \text{density of electrons}$$
(3.9)

This formula can also be used to determine the Fermi energy  $\varepsilon_{\rm F}$  from the electron density n.

We can generalize the formula to finite temperature using that states are occupied with the probability  $f(\varepsilon_{m\mathbf{k}} - \mu)$  where f is the Fermi function

$$f(\varepsilon) = \frac{1}{e^{\varepsilon/k_{\rm B}T} + 1}$$

Her  $\mu$  is the chemical potential with  $\mu(T \to 0) = \varepsilon_{\rm F}$ .



**Figure 3.4:** Plot of the Fermi function,  $f(\varepsilon_{m\mathbf{k}} - \mu)$ . The drop at  $\varepsilon = \mu$  becomes progressively sharper as  $T \to 0$ .

At finite temperature,  $\mu$  is therefore determined from the equation

$$2\sum_{m} \int_{1.BZ} f(\varepsilon_{m\mathbf{k}} - \mu) \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3} = n$$

Note that within the Hartree-Fock approximation  $\varepsilon_{mk}$  does in general depend on both on T and n and one has to solve such an equation self-consistently.

But let us focus now first on zero temperature. Here we want to consider two cases

Case 1: All bands either completely filled or completely empty.



Figure 3.5: Case 1: All bands either fully filled or fully empty at T = 0.

This describes typically an insulator or a semiconductor (a counterexample is graphene, where one obtains a semimetal). In this case the total electron density is just given by integrals over the full BZ.

$$n = 2m_0 \int_{\boldsymbol{k} \in 1. \text{ BZ}} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} = 2m_0 \frac{1}{V_{\text{unit-cell}}}$$

where  $m_0$  is the number of filled bands and we used that according to(3.7) the volume of the 1.BZ is  $(2\pi)^3/V_{\text{unit-cell}}$ . If one has only filled or empty bands, then the density of electrons is given by an even number of electrons per unit cell. This is the case for all semiconductors and (non-magnetic) insulators.

**Case 2:** In metals, in contrast, one or more bands cross the Fermi energy and bands are only partically occupied.



Figure 3.6: Case 2: One or more bands only partially occupied at T = 0.

In this case the equation  $\varepsilon_{m\mathbf{k}} = \varepsilon_{\rm F}$  defines one or several d-1 dimensional manifolds which are called Fermi surfaces.

Fermi surface: manifold with 
$$\varepsilon_{m\mathbf{k}} = \varepsilon_{\rm F}$$
 (3.10)

As an example one can, e.g., consider the case of a single band in two dimension,  $\varepsilon_{\mathbf{k}} = -2t(\cos(k_x a) + \cos(k_y a))$ , where several Fermi surfaces are shown in Fig3.7.

To obtain an analytic understanding it is useful to consider various Taylor expansions. For small  $\mathbf{k}$ :  $\epsilon_{\mathbf{k}} = -2t\left(1 - \frac{1}{2}(k_x a)^2 + 1 - \frac{1}{2}(k_y a)^2\right) = t\left(\mathbf{k}^2 a^2 - 4\right)$  giving rise to spherical Fermi surfaces when the Fermi energy is slightly above -4t. Another special point is  $\varepsilon_F = 0$  where



**Figure 3.7:** Different Fermi surfaces in 2d for a band with dispersion  $\varepsilon_{\mathbf{k}} = -2t(\cos(k_x a) + \cos(k_y a))$ .

the Fermi surface has a square shape (why?). Finally for a Fermi energy close to the band maximum 4t one can again use a Taylor expansion, this time around the edges of the BZ where  $\varepsilon_{\boldsymbol{k}+(\pm\frac{\pi}{a},\pm\frac{\pi}{a})} \approx t \left(4 - \boldsymbol{k}^2 a^2\right)$ .

We conclude with a few remarks.

– In typical metals the Fermi energy is of the oder of typical atomic binding energies and thus given by  $\varepsilon_{\rm F} \sim 1...10 \, {\rm eV}$ . Importantly this coresponds to a giant temperature scale of 10,000 -100,000K, many orders of magnitude larger than room temperature. Therefore thermal excitations mainly exists very close to the Fermi energy.

- We defined the Fermi surface for a non-interacting system but it turns out that it can also be exactly defined even for strongly interacting metals.

- Fermi surfaces come with different shapes and topologies.

- Fermi volume, defined by the volume within the Fermi surface is, according to Eq. (3.9) fixed by the electron density (at T = 0). This turn out to be true **even in the presence of interactions**. This relation is also called(Luttinger's theorem)

Luttinger's theorem: 
$$n = \text{spin-degeneracy} \times \sum_{m} \int_{\varepsilon_{mk} < \varepsilon_{\text{F}}} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}$$

Interaction can, however, change the shape and topology of Fermi surface and also the velocity of electrons at Fermi surface which turns out to be a decisive factor for many properties as we will discuss next.

#### 3.4 Density of States, van-Hove singularities and thermodynamics

#### 3.4.1 Density of states

When discussing the physical properties of solids, we often have to do sums of the form  $\sum_{n,\boldsymbol{k},\sigma=\uparrow/\downarrow} g(\varepsilon_{m\boldsymbol{k}})$ . For example, the total number of noninteracting particles is given by N =

 $\sum_{\substack{n,\boldsymbol{k},\sigma=\uparrow/\downarrow}} f(\varepsilon_{n\boldsymbol{k}}), \text{ the total energy by } E = \sum_{\substack{n,\boldsymbol{k},\sigma=\uparrow/\downarrow}} \varepsilon_{n\boldsymbol{k}} f(\varepsilon_{n\boldsymbol{k}}). \text{ From the latter we will be able to compute, for example, thermodynamic properties.}}$ 

The basic trick is to rewrite the sum as an integral over energy

$$\sum_{n,\boldsymbol{k},\sigma=\uparrow/\downarrow} g(\varepsilon_{m\boldsymbol{k}}) = \sum_{n,\boldsymbol{k}} \int g(\varepsilon) \delta(\varepsilon - \varepsilon_{m\boldsymbol{k}}) \, \mathrm{d}\epsilon = V \int g(\varepsilon) \mathcal{N}(\varepsilon) \, d\varepsilon$$

Here we defined the (total) density of states

$$\mathcal{N}(\varepsilon) = \frac{2}{V} \sum_{n, \mathbf{k}} \delta(\varepsilon - \epsilon_{n\mathbf{k}}) \tag{3.11}$$

where the factor 2 in front arises from the sum over spins. Sometimes it is also usefuld to define the density of states for a single band,

$$\mathcal{N}_m(\varepsilon) = 2 \int_{\mathbf{k} \in 1. \text{ BZ}} \delta(\varepsilon - \varepsilon_{m\mathbf{k}}) \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}$$

The integral over the density of states of a single band is 2 electrons per unit-cell or

$$\int_{-\infty}^{\infty} \mathcal{N}_m(\varepsilon) \, \mathrm{d}\varepsilon = \frac{2}{V_{\text{unit-cell}}}$$

In some cases one also uses the density of states per spin, omitting the prefactor 2 in the equations above. The density of states counts the number of states per volume and per energy and has the corresponding units. With these definitions, we can easily compute the density of electrons

$$n = \frac{N_e}{V} = \int f(\varepsilon - \mu) \mathcal{N}(\varepsilon) \,\mathrm{d}\varepsilon,$$

To compute the density of states, Eq. (3.11) (and similar k integrals), it is useful to split the k integral into an integral over energy and a contribution along surfaces of constant energy (the most important of such surfaces is the Fermi surface)

$$\int d\boldsymbol{k}... = \int d\boldsymbol{k}... \int \delta(\varepsilon - \varepsilon_{m\boldsymbol{k}}) d\epsilon = \int d\epsilon \underbrace{\int dS}_{\substack{\text{integral over surface} \\ \text{of constant energy}}} \frac{1}{|\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_{m\boldsymbol{k}}|}...$$

Here  $\nabla_k \varepsilon_{mk} = v_{mk}$  is the group velocity of electrons, evaluated at the energy  $\epsilon$ . For the density of states of band m, we obtain

$$\mathcal{N}_m(\omega) = \frac{2}{(2\pi)^d} \int \frac{\mathrm{d}\varepsilon \, dS}{|\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_{m\boldsymbol{k}}|} \delta(\omega - \varepsilon) = \frac{2}{(2\pi)^d} \int \frac{1}{|\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_{m\boldsymbol{k}}|_{\varepsilon_{m\boldsymbol{k}}=\omega}} \,\mathrm{d}S \tag{3.12}$$

The density of states is therefore an integral of the inverse velocity integrated over surfaces in momentum space where  $\varepsilon_{m\mathbf{k}} = \omega$ .

#### 3.4.2 Van-Hove singularities

The density of states is not a smooth function of  $\omega$  but is characterized by a set of so-called **van-Hove singularities**. According to Eq. (3.12) we might obtain something singular when the velocity  $\nabla_{k} \varepsilon_{mk}$  vanishes somewhere on the surface with energy  $\omega$ . Let us assume that we have identified such a point  $k_0$  with  $v_{mk_0} = 0$ . Then we can Taylor-exand the energy around that point and obtain

$$\varepsilon_{m\mathbf{k}} \approx \varepsilon_{m\mathbf{k}_0} + \frac{1}{2} \frac{\partial^2 \varepsilon}{\partial k_i \partial k_j} \bigg|_{\mathbf{k}=\mathbf{k}_0} (\mathbf{k}-\mathbf{k}_0)_i \cdot (\mathbf{k}-\mathbf{k}_0)_j + \dots$$

Such points arise at minima, maxima or - most interesting - at saddle points of  $\varepsilon_{mk}$ . To analyze what is happing in a *d*-dimensional system, we start with a simple variable transformation:  $\mathbf{k} = \mathbf{k}_0 + \sqrt{(\varepsilon - \varepsilon_{mk_0})} \mathbf{x}$ . Taking only small  $\mathbf{x}$  into account, we obtain a contribution to the density of states which reads

$$\mathcal{N}(\varepsilon) = 2\int \sqrt{(\varepsilon - \varepsilon_{m\mathbf{k}_0})^d} \delta \left[ (\varepsilon - \varepsilon_{m\mathbf{k}_0}) \left( 1 + \frac{1}{2} \frac{\partial^2 \varepsilon}{\partial x_i \partial x_j} x_i x_j \right) \right] \frac{\mathrm{d}\boldsymbol{x}}{(2\pi)^d} \\ = (\varepsilon - \varepsilon_{m\mathbf{k}_0})^{\frac{d}{2} - 1} 2\int \delta \left[ 1 + \frac{1}{2} \frac{\partial^2 \varepsilon}{\partial x_i \partial x_j} x_i x_j \right] \frac{\mathrm{d}\boldsymbol{x}}{(2\pi)^d} \propto (\varepsilon - \varepsilon_{m\mathbf{k}_0})^{\frac{d}{2} - 1}$$

Depending on the spatial dimension d, we obtain different results.

d = 1: Both at the minima and maxima of bands, there is a  $\frac{1}{\sqrt{\varepsilon}}$  divergence.



**Figure 3.8:** van-Hove singularity in d = 1 with  $\frac{1}{\sqrt{\varepsilon}}$  divergence in DoS.

d = 2: Here there is a jump to a constant value at the band edge. In d = 2 the bands have not only maxima and minima but also at least one saddle point. Closer inspection of the equation above reveals that there is a jump at logarithmic singularity  $(-\ln |\varepsilon - \varepsilon_0|)$  at the saddle point.


**Figure 3.9:** van-Hove singularity in d = 2 with  $-\ln |\varepsilon - \varepsilon_0|$  divergence in DoS.

d = 3: Here there are  $\sqrt{\varepsilon}$  non-analyticity both at minima, maxima and saddle points.



Figure 3.10: van-Hove singularity in d = 3 displays  $\sqrt{\varepsilon}$  non-analytic behavior in DoS.

The fact that  $\varepsilon_{mk}$  is periodic in k enforces that each band has to have at least one maximum, at least one minimum and in d > 1 several saddle points.

A modern interpretation of van-Hove singularities is that they are associated to a special type of "topological quantum phase transitions". Here, at T = 0 as function of the chemical potential  $\mu$  (or, equivalently, as function of the Fermi energy) **the topology of the Fermi surface changes**. For example, upon crossing the band minimum a new Fermi surfaces emerges, at a maximum it vanishes. Most interesting are saddle point where typically two Fermi surfaces merge. This can, e.g., be seen in a 2*d* example, with  $\varepsilon_{m\mathbf{k}} \approx k_x^2 - k_y^2$ , where for  $\mu = 0$  two Fermi surfaces touch at the origin.



Figure 3.11: van-Hove singularity in d = 2 – an example of a topological phase transition.

## 3.4.3 Thermodynamics of non-interacting electrons and Sommerfeld expansion

We finish this paragraph with a few results on the thermodynamic properties of non-interacting electron. We start with the observation that - as discussed above - both the density of electrons n and the energy per volume u can be written as simple integrals over the density of states:

$$n = \sum_{m} \int \mathcal{N}_{m}(\varepsilon) f(\varepsilon - \mu) \, \mathrm{d}\varepsilon$$
$$u = \sum_{m} \int \mathcal{N}_{m}(\varepsilon) \varepsilon f(\varepsilon - \mu) \, \mathrm{d}\varepsilon$$

From the energy, one can, for example compute the specific heat, which measures how much energy per Kelvin is needed to heat a sample:

$$C_V = \frac{T}{V} \frac{\partial f}{\partial T} \bigg|_V = \frac{\partial U}{\partial T} \bigg|_V$$

To compute the integrals, we can use that for typical metals the bandwidth and the Fermi energy are gigantic compared to the temperature,  $1-10 \text{ eV} = 10,000 - 100,000 \text{ K}, T \ll \mu$ . We are therefore always in the low-temperature limit and can calculate the integrals above using a Taylor-expansion trick called **Sommerfeld expansion**. These are rules to perform integrals over Fermi functions in the low-temperature limit which read

$$\int_{-\infty}^{\infty} f(\varepsilon - \mu) n(\varepsilon) \approx \int_{-\infty}^{\mu} n(\varepsilon) + n'(\mu) \frac{\pi^2}{6} (k_{\rm B}T)^2 + \mathcal{O}(T^4)$$
(3.13)

where  $n(\varepsilon)$  is some arbitrary well-behaved function which varies only smoothly on the energy scale set by the temperature T.

For a proof it is useful to introduce the integral  $H(\varepsilon) = \int_{-\infty}^{\varepsilon} n(\varepsilon') d\varepsilon'$  which allows to rewrite

by partial integration

$$\int f(\varepsilon - \mu)n(\varepsilon) = \int_{-\infty}^{\infty} H(\varepsilon)(-f'(\varepsilon - \mu))$$

Now we use that  $-f'(\varepsilon) = \frac{\beta e^{\beta\varepsilon}}{(e^{\beta\varepsilon}+1)^2} = \frac{\beta}{(2\cosh(\beta\epsilon/2))^2}$  is a sharply peaked function of width T which is even in  $\epsilon$ . It can be viewed as a broadened version of a  $\delta$ -function. As we assume that  $h(\varepsilon)$  is smooth on the scale T, we can use a Taylor expansion of  $H(\varepsilon) = H(\mu) + \sum_{n=1}^{\infty} \frac{(\varepsilon-\mu)^n}{n!} \frac{\mathrm{d}^n H}{\mathrm{d}\varepsilon^n}$ . The remaining task is to evaluate integrals of the typ  $\int (-f'(\varepsilon-\mu)) \frac{(\varepsilon-\mu)^n}{n!} \,\mathrm{d}\varepsilon = \sum_{x=\frac{\varepsilon-\mu}{k_{\mathrm{B}}T}} (k_{\mathrm{B}}T)^n \int_{-\infty}^{\infty} \frac{1}{n!} \frac{x^n}{4\cosh^2(x/2)} \,\mathrm{d}x = (k_{\mathrm{B}}T)^n S_n$ 

$$S_{2n+1} = 0, \quad S_2 = \frac{\pi^2}{6}, \quad S_4 = \frac{\pi^4}{90}$$

Using  $S_2$ , we recover (3.13), with  $S_4$  we can easily calculate also the next-order correction proportional to  $T^4$ .

As a useful example, let us consider the electron density for a system with a single band crossing th eFermi energy (for simplicity we set  $k_{\rm B} = 1$ , measuring temperatures in units of energy). More precisely, we will need below to calculate how the chemical potential changes as function of T for low T. Using

$$n \approx \int_{-\infty}^{\mu} \mathcal{N}(\varepsilon) + \mathcal{N}'(\mu) \frac{\pi^2}{6} T^2$$

we keep n fixed and use the equation to determine  $\mu(T)$ . At low T we define,  $\mu(T) - \mu(T = 0) = \mu(T) - \varepsilon_{\rm F} = \Delta \mu$  and Taylor exand in  $\Delta \mu$  and T to obtain

$$\Delta\mu\mathcal{N}(\varepsilon_{\rm F}) + \mathcal{N}'(\varepsilon_{\rm F})\frac{\pi^2}{6}T^2 \approx 0 \qquad \Longrightarrow \qquad \Delta\mu \approx -\frac{\pi^2}{6}\frac{\mathcal{N}'(\varepsilon_{\rm F})}{\mathcal{N}(\varepsilon_{\rm F})}T^2$$

This result we use for the Sommerfeld expansion of the energy per volume

$$u(T) \approx \int_{-\infty}^{\mu} \varepsilon \mathcal{N}(\varepsilon) + \frac{\pi^2}{6} (\mathcal{N}(\mu) + \mu \mathcal{N}'(\mu)) T^2$$

Here we have to set  $\mu = \epsilon_F + \Delta \mu$  and Taylor expand in T. From the first term we obtain a correction  $\mu \mathcal{N}(\mu) \Delta \mu$  which cancels the last term and we obtain

$$u(T) \approx u(T=0) + \frac{\pi^2}{6} \mathcal{N}(\varepsilon_{\mathrm{F}}) T^2$$

From this we can obtain the important specific heat formula

$$C_v \approx \frac{\pi^2}{3} \mathcal{N}(\varepsilon_{\rm F}) k_{\rm B}^2 T$$
 (3.14)

Consider, for example, a quadratic band,  $\varepsilon_{\mathbf{k}} = \frac{\mathbf{k}^2}{2m}$  (setting  $\hbar = 1$ ) which gives rise to the density of state density of state  $N(\varepsilon) = \frac{2 \cdot 4\pi}{(2\pi)^3} m \sqrt{2m\varepsilon}$ . As T = 0 we can easily compute the electron density and obtain  $n = \frac{1}{\pi^2} m \sqrt{2m} \frac{2}{3} \varepsilon_{\mathrm{F}}^{\frac{3}{2}}$ . This allows to express the Fermi energy in terms of the electron density. To be able to compare the classical and quantum result for the specific heat, it is convenient to express  $\mathcal{N}(\varepsilon_{\mathrm{F}}) = \frac{3n}{2\varepsilon_{\mathrm{F}}}$ . Thus we find

$$C_v \approx \frac{\pi^2}{2} \left(\frac{k_{\rm B}T}{\varepsilon_{\rm F}}\right) k_{\rm B} n$$

Classically, the specific heat of particles with quadratic dispersion is given by  $C_v = \frac{3}{2}k_{\rm B}n$ . We see that this value is reduced by a factor proportional to  $\frac{k_{\rm B}T}{\varepsilon_{\rm F}}$ . This is a consequence of the Pauli principle, only states in a window of width T around the Fermi energy contribute to thermodynamics.

# 3.5 Almost localized electrons: tight binding model

### 3.5.1 Linear combination of atomic orbitals - LCAO

We come back to the problem of calculating band structures. We will start by discussing the tunneling of almost localized electrons. The main reason why we are interested in this limit is that it does not only provide a simple approximation but turns out to be extremely useful to obtain an intuition of how electrons move in a lattice.

Our starting point is, however, a crystal where the atoms are placed in a considerable distance from each other. Therefore we can - in first approximation - neglect any tunneling and consider simply the atomic orbitals. In this case choosing just a linear combination of atomic orbitals (LCAO=Linear Combination of Atomic Orbitals) which all have the same energy will be a good approximation to the wavefunction of the crystal.

We start with a single atom (e.g., hydrogen) assuming that we know all single particle eigenstates and energies

$$H_0\Phi_m^n = E_n\Phi_m^n, m = 1, \dots, N$$

Here we use two quantum numbers: n for the energy and an extra quantum number m assuming that we have N degenerate states which all have the same energy. In the following we will suppress the index n, just focussing on the degenerate states. Now we consider a crystal of those atoms located at positions  $\mathbf{R}_n$  (for simplicity we assume that the atoms form a simple Bravais lattice and discuss generalizations later) and assume that the wave function will be a linear combination of the atomic orbitals  $\Phi_m(\mathbf{r} - \mathbf{R}_n)$ . To guess the form of the wave function at a fixed momentum  $\mathbf{k}$ , we use the constraint of Bloch's theorem. This implies that the wavefunction has to take the form

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_n} e^{i\boldsymbol{k}\cdot\boldsymbol{R}_n} \sum_m b_m \Phi_m(\boldsymbol{r}-\boldsymbol{R}_n)$$

We can easily check that the wavefunction fulfills the necessary condition  $\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_i) = e^{i\mathbf{k}\cdot\mathbf{R}_i}\psi_{\mathbf{k}}(\mathbf{r})$ . As a next step, we would like to evaluate  $H\psi_{\mathbf{k}}$  with the goal to obtain eigenfunctions of the Hamiltonian

$$H\psi_{\boldsymbol{k}}(\boldsymbol{r}) \approx E_{\boldsymbol{k}}\psi_{\boldsymbol{k}}(\boldsymbol{r})$$

To obtain such an equation, we use two tricks. We focus on the atom at the origin,  $\mathbf{R}_n = 0$ , write  $H = \underbrace{H_{\text{atom}}}_{\text{at the origin}} + \Delta U(\mathbf{r})$  and project the equation onto orbital m of the atom at the origin. This gives

ino giveo

$$\int \Phi_m^*(\boldsymbol{r})(H_{\text{atom}} + \Delta U(\boldsymbol{r}))\psi_{\boldsymbol{k}}(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} - E_{\boldsymbol{k}} \int \Phi_m^*(\boldsymbol{r})\psi_{\boldsymbol{k}}(\boldsymbol{r}) \,\mathrm{d}\boldsymbol{r} = 0$$

We have to evaluate 3 terms here. Let us start by combining the first and the last term. Using that  $\Phi_m^*(\mathbf{r})H_{\text{atom}} = E_n \Phi_m^*(\mathbf{r})$ , we obtain

$$(E_n - E_{\boldsymbol{k}}) \int \Phi_m^*(\boldsymbol{r}) \psi_{\boldsymbol{k}}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} = (E_n - E_{\boldsymbol{k}}) \left( b_m + \sum_{m'} b_{m'} \underbrace{\sum_{\boldsymbol{R}_n \neq \boldsymbol{0}} \int \Phi_m^*(\boldsymbol{r}) \Phi_{m'}(\boldsymbol{r} - \boldsymbol{R}_n)}_{\text{small, only relevant for close}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}_n} \right) \\ = (E_{\boldsymbol{k}} - E_n) \left( \delta_{mm'} + M_{mm'}^1 \right) b_{m'}$$
(3.15)

The remaining term reads

$$\int \Phi_{m}(\mathbf{r})\Delta U(\mathbf{r})\psi_{\mathbf{k}}(\mathbf{r}) \,\mathrm{d}\mathbf{r}$$

$$= \sum_{m'} b_{m'} E_{n} \left( \underbrace{\int \Phi_{m}^{*}(\mathbf{r})\Delta U(\mathbf{r})\Phi_{m'}(\mathbf{r})}_{\text{relatively small for small }\mathbf{r}} + \underbrace{\sum_{\mathbf{R}_{n}\neq\mathbf{0}} \int \Phi_{m}^{*}(\mathbf{r})\Delta U(\mathbf{r})\Phi_{m'}(\mathbf{r}-\mathbf{R}_{n})}_{\text{small, only relevant for close}} \right)$$

$$= \sum_{m'} M_{mm'}^{2} b_{m'} \qquad (3.16)$$

In total, we therefore obtain the following k-dependent equation for the N unknowns  $b_m$ 

$$M_{mm'}^{2}(\boldsymbol{k})b_{m'} = (E_{\boldsymbol{k}} - E_{n})\left(\delta_{mm'} + M_{mm'}^{1}\right)b_{m'}$$

N eigenvalues of matrix 
$$(\mathbb{I} + M^1(\mathbf{k}))^{-1} M^2(\mathbf{k})$$
 determine N bands  $\varepsilon_m(\mathbf{k})$  (3.17)

Here the matrix  $M^1$  is a correction term which arises because the atomic orbitals at different sites are not orthogonal to each other.

Let us consider the most important example, N = 1, relevant for non-degenerate *s*-electrons. In this case, there is no need to diagonalize any matrix but we can read off

$$E_{\mathbf{k}} = E_n - \frac{\beta + \sum_{\mathbf{R}_n} t(\mathbf{R}_n) e^{i\mathbf{k}\cdot\mathbf{R}_n}}{1 + \sum_{\mathbf{R}_n} \alpha(\mathbf{R}_n) e^{i\mathbf{k}\cdot\mathbf{R}_n}}$$

where  $\beta = -\int \Phi^*(\mathbf{r})\Delta U(\mathbf{r})\Phi(\mathbf{r})$  is the local change of the potential due to neighboring atoms, while  $t(\mathbf{R}_n) = -\int \Phi^*(\mathbf{r})\Delta U(\mathbf{r})\Phi(\mathbf{r}-\mathbf{R}_n)$  is interpreted as the rate with which one can hop from one atom to a neighboring atom. Finally,  $\alpha(\mathbf{R}_n) = -\int \Phi^*(\mathbf{r})\Phi(\mathbf{r}-\mathbf{R}_n)$  is a correction factor due to non-orthogonality of atoms. As  $\phi(\mathbf{r})$  is real, we have  $t(\mathbf{R}_n) =$  $t(-\mathbf{R}_n), \alpha(\mathbf{R}_n) = \alpha(-\mathbf{R}_n)$ .

Often we have the situation that the dominant contribution comes only from nearest neighbors contribute. Furthermore, the overlap of nearest neighbors is small,  $\alpha \ll 1$ , and we can also ignore the term  $\beta \alpha \ll t$ . In this case the dispersion is given by

$$E_{\boldsymbol{k}} \approx E_n - \beta - t \sum_{\text{n.n.}} \cos(\boldsymbol{k} \cdot \boldsymbol{R}_n)$$

where n.n. indicates that the sum is only over vectors  $\mathbf{R}_n$  which connect the origin to nearest neighbors. For example, for an fcc lattice, one has 12 nearest neighbors located at  $\mathbf{R}_n = \frac{a}{2}(\pm 1, \pm 1, 0), \frac{a}{2}(0, \pm 1, \pm 1), \frac{a}{2}(\pm 1, 0, \pm 1)$  and one finds

$$E_{\mathbf{k}} = \operatorname{const} - 4t \left( \cos \left( \frac{k_x a}{2} \right) \cos \left( \frac{k_y a}{2} \right) + \operatorname{cyclic} \right)$$



Figure 3.12: Tight-binding dispersion for an FCC lattice leading to two different kinds of fermi surfaces.

The approach discussed above can easily be generalized to more complex situation. Consider, for example, a case where one has M atoms per unit cell located at  $\mathbf{R}_n + \mathbf{b}_i$ ,  $i = 1, \ldots, M$ . In this case one includes those in the ansatz

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{n} \sum_{i=1}^{M} \sum_{m_i=1}^{N_i} \Phi_{m_i}^i \left(\boldsymbol{r} - (\boldsymbol{R}_n + \boldsymbol{b}_i)\right) \cdot b_{m_i}^i \cdot e^{i\boldsymbol{k} \cdot (\boldsymbol{R}_n + \boldsymbol{b}_i)}$$

In total one uses here  $N = N_1 + N_2 + ... N_M$  variables. Proceeding as above, one then has to solve for each momentum  $\mathbf{k}$  an  $N \times N$  eigenvalue problem from which one calculates N bands.

#### 3.5.2 Phenomenology with tight-binding approaches

One of the most important aspect of the tight-binding description of electron motion is that it provides an **intuitive picture** for electron motion in a crystal. Electrons **hop from orbital to orbital**. Nearest orbital give the dominant contribution but one can easily also include next-nearest or next-next nearest orbitals. Most frequently, the corresponding hopping matrix elements are not calculated from  $\int \Phi_m^*(\mathbf{r})\Delta U(\mathbf{r})\Phi_{m'}(\mathbf{r}-\mathbf{R}_n) = t_{mm'}(\mathbf{R}_n)$  in the way we have described it above but used as **phenomenological fitting parameters** which can either be fitted to some experiment or to some full-fledged band-structure calculation using the methods outlined in the previous chapter. Also features like spin-orbit coupling arising from relativistic corrections in heavy elements (where electrons move faster due to the higher ion charge) can easily be included in a tight-binding approach. With the proper generalization of orbitals (discussed below) this fitting-parameter approach is frequently used even in situations where a tight-binding approximation cannot not justified. When using this approach, one also 'ignores' the non-orthogonality of orbitals at different sites (which can always be avoided by using Wannier functions introduced below). In this case, equation (3.17) simplifies

$$M_{\boldsymbol{k}}\boldsymbol{b} = \varepsilon_{\boldsymbol{k}}\boldsymbol{b}$$
(3.18)  
with  $N \times N$  matrix  $M_{\boldsymbol{k}mm'} = \sum_{\boldsymbol{R}_n} t_{mm'}(\boldsymbol{R}_n) e^{i\boldsymbol{k}\cdot\boldsymbol{R}_n}$ 

How can we limit the number of fitting parameters  $t_{mm'}(\mathbf{R}_n)$ ? First, we can use that  $t_{mm'}(\mathbf{R}_n)$  is only sizable for small  $\mathbf{R}_n$ . Second, one has to perform a proper symmetry analysis to find out which hopping parameters are allowed by symmetry. To give an example, the hopping between

an s-orbital and a  $p_x$  orbital will vanish, if the two orbitals are connected by, e.g., by rotation axis with a 180 degree rotation symmetry. In formulas, the following integral will vanish

$$\int \Phi_s^*(\boldsymbol{r}) \Delta U(\boldsymbol{r}) \Phi_{p_y}\left(\boldsymbol{r} + (a, 0, 0)\right) = 0 \implies t_{sp_y}\left((a, 0, 0)\right) = 0$$

if a symmetry exist which maps y to -y which changes the sign of  $\Phi_y$  but not of the other terms in the integral. In contrast, the integral  $\int \Phi_s^*(\mathbf{r}) \Delta U(\mathbf{r}) \Phi_{p_y}(\mathbf{r} + (0, a, 0))$  will be finite, see Fig. 3.13.



Figure 3.13: Between an s and a p orbital, the relative orientation determines whether hopping is possible or not.

#### 3.5.3 Graphene

Let us do perhaps the - by now - most famous example of a band-structure calculation. We consider **graphene**, a two-dimensional sheet of carbon forming a honeycomb lattice with 2 atoms/unit-cell.



Figure 3.14: Honeycomb lattice structure of Graphene.

The (triangular) Bravais lattice is spanned by the two vectors

$$a_1 = a(1,0), \qquad a_2 = a\left(\cos\frac{\pi}{3}, \sin\frac{\pi}{3}\right) = a\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

There are two atoms per unit cell located at

$$\boldsymbol{R_n} = n_1 \boldsymbol{a}_1 + n_2 \boldsymbol{a}_2$$
 and  $\boldsymbol{R_n} + \boldsymbol{\Delta}$ , with  $\boldsymbol{\Delta} = a \left( 0, \frac{1}{\sqrt{3}} \right)$ 

The reciprocal lattice vectors are spanned by

$$\boldsymbol{G}_1 = \frac{2\pi}{a} \left( 0, \frac{2}{\sqrt{3}} \right), \boldsymbol{G}_2 = \frac{2\pi}{a} \left( 1, -\frac{1}{\sqrt{3}} \right)$$

and the 1.BZ is a hexagon with edges located at  $(\mp \frac{4\pi}{3a}, 0), (\pm \frac{2\pi}{3a}, \pm \frac{2\pi}{\sqrt{3a}}).$ 

The bands close to the Fermi energy are formed by  $p_z$  orbitals sticking out of the graphene plane. We want to use the tight-binding ansatz in its phenomenological version (3.18) (ignoring issues related to the non-orthogonality of bare atomic orbitals). The tight-binding ansatz reads

$$\psi_{k} = \sum_{\boldsymbol{R}_{n}} \left( \alpha_{k} \Phi(\boldsymbol{r} - \boldsymbol{R}_{n}) + \beta_{k} e^{i \boldsymbol{k} \cdot (\boldsymbol{R}_{n} + \boldsymbol{\Delta})} \Phi(\boldsymbol{r} - (\boldsymbol{R}_{n} + \boldsymbol{\Delta})) \right)$$

To set up the matrix, we need the vectors connecting nearest neighbors. Each carbon atom has three nearest neighbors located at at

$$\boldsymbol{n}_1 = a\left(0, \frac{1}{\sqrt{3}}\right), \ n_2 = R_{\frac{2\pi}{3}}\boldsymbol{n}_1 = a\left(-\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right), \ \boldsymbol{n}_3 = R_{\frac{4\pi}{3}}\boldsymbol{n}_1 = a\left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}\right)$$

relative to the Carbon position.



Figure 3.15: Position vectors of nearest neighbors relative to a site in the honeycomb lattice.

These vector connect the A sublattice parameterized by  $\alpha_{\mathbf{k}}$  and the B sublattice parameterized by  $\beta_{\mathbf{k}}$  and thus the hopping matrix elements show up in the off-diagonal elements of the matrix equation (see Eq. (3.18))

$$\varepsilon_{\mathbf{k}} \begin{pmatrix} a_{\mathbf{k}} \\ b_{\mathbf{k}} \end{pmatrix} = \begin{pmatrix} 0 & -t \sum_{i=1}^{3} e^{i\mathbf{k} \cdot \mathbf{n}_{i}} \\ -t \sum_{i=1}^{3} e^{i\mathbf{k} \cdot \mathbf{n}_{i}} & 0 \end{pmatrix} \begin{pmatrix} a_{\mathbf{k}} \\ b_{\mathbf{k}} \end{pmatrix}$$
(3.19)

The eigenvalues are given by  $\pm \sqrt{|t\sum_{i=1}^{3} e^{i \mathbf{k} \cdot \mathbf{n}_i}|^2}$  but most of the interesting physics occurs at the edges of 1. BZ, e.g.  $\mathbf{k}_0 = (\frac{4\pi}{3a}, 0)$ .



Figure 3.16: Bandstructure of Graphene in the tight-binding approximation.

At this momentum we have  $\sum_{i=1}^{3} e^{i\mathbf{k}_{0}\cdot\mathbf{n}_{i}} = 1 + e^{-i\frac{2\pi}{3}} + e^{i\frac{2\pi}{3}} = 0$  and therefore  $\varepsilon_{\mathbf{k}_{0}} = 0$ . Let us do a Taylor expansion around this point by evaluating the Hamilton matrix (3.19) at the momentum  $\mathbf{k}_{0} + \mathbf{k}$  for small k. Using  $-t\sum_{i=1}^{3} e^{i(\mathbf{k}_{0}+\mathbf{k})\cdot\mathbf{n}_{i}} \approx t\frac{\sqrt{3}}{2}a(k_{x}-ik_{y})$  we obtain a new k-dependent Hamiltonian for the Schrödinger equation  $\varepsilon_{\mathbf{k}_{0}+\mathbf{k}} \begin{pmatrix} a_{\mathbf{k}} \\ b_{\mathbf{k}} \end{pmatrix} = H_{\mathbf{k}_{0}+\mathbf{k}} \begin{pmatrix} a_{\mathbf{k}} \\ b_{\mathbf{k}} \end{pmatrix}$  with

$$H_{\mathbf{k}_0+\mathbf{k}} \approx \hbar v \begin{pmatrix} 0 & k_x - ik_y \\ k_x + ik_y & 0 \end{pmatrix} = v(k_x \sigma_x + k_y \sigma_y)$$

where  $v = t\frac{\sqrt{3}}{2}a/\hbar$  is an effective velocity of the order of  $10^6$  m/s. The energy close to the K and K' points is therefore linear in momentum,  $\varepsilon_{k_0+k} = \pm t\frac{\sqrt{3}}{2}\sqrt{k_x^2 + k_y^2}$ . As there are two electron per unit cell, i.e., one per orbital, the states with negative energy are filled while the ones with positive energy are empty. The chemical potential therefore vanishes,  $\mu = 0$ .

A bit of care is required when counting the number of Fermi points: the energy vanishes at the 6 edges of the hexagonal 1. BZ, see Fig. 3.16. However, one edge is connected by reciprocal lattice vectors to 2 other edges. Due to the perodicity of the 1. BZ it thus describes the same point. Therefore there are just 2 non-equivalent Fermi points, which we can denote, e.g., by  $\mathbf{k}_0$  and  $-\mathbf{k}_0$  (also referred to as the K and K' points) the with

$$H_{-\boldsymbol{k}_0+\boldsymbol{k}} \approx \hbar v (-k_x \sigma_x + k_y \sigma_y)$$

The effective low-energy Schrödinger equation takes both at the K and K' points the form of a two-dimensional massless Dirac equation. If we include the spin-index (which we omitted above), we end up with 4 copies of a Dirac equation, which can be described by four twocomponent fields  $\psi_{\pm \mathbf{k}_0,\uparrow/\downarrow}(\mathbf{r})$ , where the two components reflect the A and B sublattice as introduced above. The real-space Hamiltonian for  $\psi_{\mathbf{k}_0,\uparrow/\downarrow}(\mathbf{r})$  reads, for example,

$$H = v(i\hbar\partial_x\sigma_x + i\hbar\partial_y\sigma_y).$$

The special linear dispersion and the physics of the Dirac equation governs many of the properties of graphene. Let me just give one example: In 1929 Klein made a surprising observation. Massless fermions described by the Dirac equation can tunnel through a potential without getting reflected. This has not only been observed for graphene but also is of technological importance as it makes it much harder to make transistors out of graphene.

#### 3.5.4 Wannier functions

The concept of hopping between localized orbitals can be made precise even in cases when atoms are not far apart from each other. Here one has, however, to replace the concept of atomic orbitals by a more general one, the so-called Wannier functions.

The starting point for the construction of Wannier functions are exact eigenstates of a periodic Hamiltonian for fixed momentum  $\mathbf{k}$ , the the Bloch functions  $\psi_{m,\mathbf{k}}(\mathbf{r})$  with  $\psi_{m,\mathbf{k}}(\mathbf{r} + \mathbf{R}_n) = e^{i\mathbf{k}\mathbf{R}_n}\psi_{m,\mathbf{k}}(\mathbf{r})$ . Let us try to construct a wave function localized around the site  $\mathbf{R}_i$  by using a formula motivated from the inverse-Fourier transform. We define the Wannier functions

$$\phi_{m\boldsymbol{R}_{i}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k} \in 1.BZ} \psi_{m,\boldsymbol{k}}(\boldsymbol{r} - \boldsymbol{R}_{i}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k} \in 1.BZ} e^{-i\boldsymbol{k} \cdot \boldsymbol{R}_{i}} \psi_{m,\boldsymbol{k}}(\boldsymbol{r})$$

As usually, N is the number of unit-cells, or equivalently, the number of k vectors and we assume that  $\psi_{m,k}(r)$  is normalized to 1.

Let us start by calculating

$$\langle \phi_{m\mathbf{R}_{i}} | \phi_{m'\mathbf{R}_{j}} \rangle = \int \phi_{m\mathbf{R}_{i}}^{*}(\mathbf{r}) \phi_{m'\mathbf{R}_{j}}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k},\mathbf{k}'} e^{i\mathbf{k}\cdot\mathbf{R}_{i}} e^{-i\mathbf{k}'\cdot\mathbf{R}_{j}} \underbrace{\int \psi_{m,\mathbf{k}}^{*}\psi_{m',\mathbf{k}'}}_{\delta_{\mathbf{k},\mathbf{k}'}\delta_{m,m'}} = \delta_{mm'}\delta_{\mathbf{R}_{i},\mathbf{R}_{j}}$$
(3.20)

The Wannier functions are by construction orthonormal. Furthermore, assuming that the  $\psi_{m,k}(\mathbf{r})$  are smooth function of  $\mathbf{k}$ , one finds that

 $\phi_{m\mathbf{R}_i}(\mathbf{r})$  is localized around  $\mathbf{R}_i$ 

which follows from the construction as a Fourier transform. A remarkable exception are socalled topological insulators (discussed below). For those it is impossible to find for a *single* band m eigenstates  $\psi_{m,\mathbf{k}}(\mathbf{r})$  which are smooth functions of  $\mathbf{k}$ .

Some other properties of Wannier functions follow from their definition

$$\phi_{mR_i}(r) = \phi_{m,R=(0,0,0)}(r - R_i)$$

$$\psi_{m,\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}_l} \phi_m(\boldsymbol{r}-\boldsymbol{R}_l)$$

If you compare this with the tight-binding ansatz, you will recognize that we used exactly the same formulas. Using Wannier functions instead of atomic orbitals makes these equations, however, exact. Another difference is that atomic orbitals at different sites are not orthogonal to each other. Wannier functions instead have this property, see Eq. (3.20).

Starting from a band structure calculation based, e.g., on LDA, one can now construct Wannier functions. Furthermore, using  $\varepsilon_{mk} = \sum_{\mathbf{R}_n} t_m(\mathbf{R}_n) e^{i\mathbf{k}\cdot\mathbf{R}_n}$  one can obtain the hopping rate from one Wannier orbital to the next. This allows for an intuitive interpretation of Wannier functions - at least as long as they are nicely localized objects.

If one encounters the situation that a Wannier function as defined above is not local but spreads over many lattice sites (as is always the case for topological insulators), one can obtain more local (and thus better to interpret) objects by choosing suitable linear combinations of wave functions from different bands. With a rather straightforward optimization code one can obtain maximally localized Wannier functions for a given set of bands. This program works even for topological insulators.

# 3.6 Symmetries and degeneracies in the band-structure

The symmetries of a Hamiltonian have also important consequences for their band structure. For example, they can enforce degeneracies of bands. A good example is graphene, where we found that two bands cross at a Dirac point and the question arises whether this is an artifact of an approximation or an exact statement.

#### 3.6.1 Time reversal invariance

Time reversal invariance is arguably the 'most tricky' symmetry and some care is required for its discussion.

First, we will ignore the electronic spin and switch off any magnetic field. In this case the Hamilitonian is obviously real and we find by complex conjugation

$$i\hbar\partial_t\psi(t) = H\psi(t) \implies -i\hbar\partial_t\psi^*(t) = H\psi^*(t) \implies i\hbar\partial_t\psi^*(-t) = H\psi^*(-t)$$

Apparently, we can find a (new) solution of the Schrödinger equation by setting t to -t if we simultaneously perform a complex conjugation operation. Thus, we define the time-reversal transformation T by

$$T\psi(\boldsymbol{x},t) = \psi^*(\boldsymbol{x},-t)$$
 for spinless fermions

Let us also define a complex conjugation operator C by

$$C\psi = \psi^*$$

With  $T\psi(\boldsymbol{x},t) = C\psi(\boldsymbol{x},-t)$ . Please note that C and T are *not* linear operators, i.e.,  $C(a\psi) = a^*C(\psi) \neq aC(\psi)$ , Operators with this property are called **antilinear**. The fact that timereversal is implemented via antilinear operators makes it different from other symmetry transformations which are described by linear (and unitary) operators.

We can now easily check what happens to various operators and physical quantities upon time-reversal

$$i\hbar\partial_t \to i\hbar\partial_t, \quad E \to E, \quad p \to -p, \quad L = r \times p \to -L$$

From the transformation properties of the angular momentum, it is clear that we should demand that for a particle with spin that  $\mathbf{S} \to -\mathbf{S}$  under time reversal. This requires some extra thought as the Pauli matrix  $\sigma_y$  is purely imaginary while  $\sigma_x$  and  $\sigma_z$  have real entries. This issue is solved by setting

$$T = i\sigma_y C \tag{3.21}$$

This defines how time-reversal acts on a two-component spinor describing spin-up and spin-down electrons

$$T\begin{pmatrix}\psi_{\uparrow}(t)\\\psi_{\downarrow}(t)\end{pmatrix} = i\begin{pmatrix}0 & -i\\i & 0\end{pmatrix}\begin{pmatrix}\psi_{\uparrow}^{*}(-t)\\\psi_{\downarrow}^{*}(-t)\end{pmatrix} = \begin{pmatrix}\psi_{\downarrow}^{*}(-t)\\-\psi_{\uparrow}^{*}(-t)\end{pmatrix}$$

Note that the spin-components are exchanged and there is an extra minus sign for the lower component. An important consequence of the definitions (3.20) is that

$$T^2 = i\sigma_y C i\sigma_y C = i\sigma_y i\sigma_y = -1$$
 for a spin 1/2 particle (3.22)

while  $T^{\dagger}T = C(-i)\sigma_y i\sigma_y C = 1$ . We can now check the transformation properties of the Pauli matrices to check whether we achieve that  $S \to -S$ :

$$T^{\dagger}\sigma_{x}T = C\sigma_{y}\sigma_{x}\sigma_{y}C = -C\sigma_{x}C = -\sigma_{x}$$

$$T^{\dagger}\sigma_{y}T = C\sigma_{y}\sigma_{y}\sigma_{y}C = C\sigma_{y}C = -\sigma_{y}$$

$$T^{\dagger}\sigma_{x}T = C\sigma_{y}\sigma_{z}\sigma_{y}C = -C\sigma_{z}C = -\sigma_{z}$$

We will now consider a solution  $\psi$  of the stationary Schrödinger equation,  $H\psi = E\psi$ . Assuming a time-reversal invariant Hamiltonian  $THT^{\dagger} = H$  we obtain

$$H T\psi = E T\psi$$

 $T\psi$  therefore has the same energy as  $\psi$ . From  $T^2 = -1$ , Eq. (3.22), we can furthermore conclude that  $T\psi$  and  $\psi$  are linearly independent from each other

(assume:  $T\psi = a\psi \implies T^2\psi = a^*T\psi = |a|^2\psi$  in contradiction to  $T^2\psi = -\psi$ ).

From this follows the famous **Kramers degeneracy**: bound states of time-reversal symmetric Hamiltonians for spin-1/2 particles are always doubly degenerate.

What consequence does time-reveral symmetry have on band structures? Are there also Kramers degeneracies? Up to now we have mainly considered systems without any spin-orbit coupling where the band-structure was completely independent of spin. In reality, however, the spin couples to the motion of electrons via spin-orbit coupling, e.g., by terms proportional to  $p \cdot \sigma$ . In this case, single particle Hamiltonians are promoted to  $2 \times 2$  matrices and the bandstructure does depend on the spin-index. From time-reversal symmetry, it follows that

$$\varepsilon_{k\uparrow} = \varepsilon_{-k\downarrow}$$

Here I use a sloppy notation, where  $\uparrow, \downarrow$  denote 2 opposite spin directions which may, however, not point in the z direction. For the momentum  $\mathbf{k} = (0, 0, 0)$ , this leads to a Kramers degeneracy. The same applies also to vectors where  $-\mathbf{k} = \mathbf{k} + \mathbf{G}_n$ , e.g.,  $\mathbf{k} = (\pi/a, 0, 0)$  in a cubic lattice with reciprocal lattice vector  $(2\pi/a, 0, 0)$ .

Without further symmetries and for a generic momentum k, we do, however, not obtain any degeneracies. In the presence of spin-orbit coupling bands do generically split.

A very important exception are materials which are not only time-reversal symmetric but also possess an inversion symmetry I which maps k to -k. In this case we find

$$\varepsilon_{\boldsymbol{k}\uparrow} \stackrel{T}{=} \varepsilon_{-\boldsymbol{k}\downarrow} \stackrel{I}{=} \varepsilon_{\boldsymbol{k}\downarrow}$$

Therefore all bands are doubly degenerate (as in the case without spin-orbit interactions). Indeed, as a rule of thumb spin-orbit interactions can change the physics qualitatively in materials without inversion symmetry but have often only minor implications in the presence of this symmetry.

With and without spin-orbit interactions, external magnetic fields break time-reversal symmetry and lift the spin-degeneracy of the bands.

#### 3.6.2 Band degeneracies due to symmetries

Above, we have shown that the combination of time-reversal and inversion symmetry leads to degenerate bands for arbitrary momenta. Ignoring this type of degeneracy, we can ask are there further points where bands cross? Such points are of interest for many reasons. A good example are the special properties of graphene associated to its Dirac points describing the crossing of two bands. But one can also think about cases, where, e.g., an extra quantum number is used to build novel types of transistors.

Generically, bands do not cross easily. The reason is called level repulsion: as we discuss in more detail below, one has to fine-tune 3 parameters to have two coinciding eigenvalues in a

Hermition  $2 \times 2$  matrix. Thus band crossings are always special. We will first discuss band degeneracies related to symmetries and then 'accidential' band degeneracies unrelated to any symmetries.

Let us first investigate how an arbitrary symmetry operation acts on a solid. Let us denote the group of symmetries of a crystal by  $\mathbb{G}$  where we denote groupelements by  $A_i = (\overleftrightarrow{S}, a) \in \mathbb{G}$ , where  $\overleftrightarrow{S}$  is an element of point group (rotation, reflection, inversion), and a is a translation vector with

$$A_i \boldsymbol{r} = \overleftrightarrow{S} \boldsymbol{r} + \boldsymbol{a}$$
$$\hat{A}_i \psi(\boldsymbol{r}) = \psi(A_i^{-1} \boldsymbol{r})$$

The use of the inverse in the equation above is a bit counterintuitive but one can easily check that it is correct (e.g., translating a wave packet by  $\boldsymbol{a}$  achieved by  $\psi(\boldsymbol{x}) \to T_a \psi(\boldsymbol{x}) = \psi(\boldsymbol{x} - \boldsymbol{a})$ ). In case of a symmetry, we can directly conclude

$$[\hat{A}_i, H] = 0 \implies \psi_{\mathbf{k}, n} \text{ and } \hat{A}_i \psi_{\mathbf{k}, n} \text{ have the same energy } \forall i.$$
 (3.23)

Let us first discuss translations by lattice vectors. Here nothing interesting happens as  $\psi_{n,k}(\mathbf{r} - \mathbf{R}_n) = e^{-i\mathbf{k}\cdot\mathbf{R}_n}\psi_{n,k}(\mathbf{r})$ , the translated wave function is proportional to the original wave function. The reason why noting is happening is that we already used the translational symmetry when we setup Bloch eigenstates of the Hamiltonian.

More relevant are elements of the point-group:  $\psi_{n,k}(\mathbf{r}) \to \psi_{n,k}(S^{-1}\mathbf{r}), S \in \text{rotation, mirror,}$ inversion. Here map a wave function at one momentum to a wave function at a new momentum obtained by the symmetry transformation. Let us check this by computing

$$\psi_{n,\boldsymbol{k}}(S^{-1}(\boldsymbol{r}+\boldsymbol{R}_n)) = e^{i\boldsymbol{k}\cdot S^{-1}\boldsymbol{R}_n}\psi_{n,\boldsymbol{k}}(S^{-1}\boldsymbol{r}) = e^{i\boldsymbol{k}'\cdot\boldsymbol{R}_n}\psi_{n,\boldsymbol{k}}(S^{-1}\boldsymbol{r}) \quad \text{for } \boldsymbol{k}' = S\boldsymbol{k}$$

Thus we find that  $\psi_{n,k}(S^{-1}r)$  is a Bloch wavefunction with momentum k' = Sk. Using Eq. (3.23), we can conclude that

$$\varepsilon_{n,\mathbf{k}} = \varepsilon_{n,S\mathbf{k}}$$

which is again a hardly a supprising observation. The bands inherit the point-group symmetries of the lattice.

Our main goal is, however, to find out whether such symmetries can enforce band degeneracies at a fixed momentum k (instead of relating two different momenta). Here we can only hope to make progress, at some special high-symmetry points in the crystal where

$$\boldsymbol{k}' = S\boldsymbol{k} = \boldsymbol{k} + \boldsymbol{G}_n,$$

where  $G_n$  is a reciprocal lattice vector (including  $G_0 = (0, 0, 0)$ ). For example, all momenta of the form  $(k_x, 0, 0)$  are mapped upon each other by, e.g., all rotations around the x-axis, by relections on the xy or xz planes or by some more fancy symmetrie, e.g., time-reversal in combination with a  $\pi$  around the y-axis. Another example is the edge of the 1.BZ in a simplecubic lattice,  $(\frac{\pi}{a}, \frac{\pi}{a}, \frac{\pi}{a})$ , which is mapped by elements of the cubic group to another edge of the 1. BZ. Those edges are, however, all connected by reciprocal lattice vectors and should therefore be identified with each other.

Let us discuss the resulting mathematical structure. For a fixed momentum k, we first define a subgroup  $\mathbb{G}_k$  of  $\mathbb{G}$  by the set of all symmetry transformations which obey  $S\mathbf{k} = \mathbf{k} + \mathbf{G}_n$  with some reciprocal lattice vector  $\mathbf{G}_n$ .

Each element  $A_{\mathbf{k}}^{i} \in \mathbb{G}_{\mathbf{k}}$ ,  $i = 1 \dots n_{\mathbf{k}}$ , will map a given Bloch eigenstate to a linear combination of other eigenstates

$$\psi_{m\mathbf{k}}(\mathbf{x}) \to A^{i}_{\mathbf{k}}\psi_{m\mathbf{k}}(\mathbf{x}) = \sum_{n} C^{mn}_{\mathbf{k}i}\psi_{n\mathbf{k}}(\mathbf{x})$$

How symmetry operators acts on the wave functions at fixed k is thus encoded in  $n_k$  matrices  $C_{ki}$  which act on the band-index space. In math, such a construction is called a **representation** of a group. A representation is a set of matrices (the  $C_{ki}$ ) which describe how the group operators (the elements of  $\mathbb{G}_k$ ) affect some vector space (the wave functions  $\psi_{nk}(x)$  in our case).

To put our discussion in the right group-theoretical context, we need two more concepts: the notion of a reducible representation and of an irreducible representation.

Physically, this is motivated by the observation that the symmetry transformation will link only a small number of bands to each other (it will be those bands which have the same energy, according to (3.23)). We should therefore focus only on groups of such bands which we will identify with "irreducible representations".

But let us first define a **reducible representation**. This is a representation where the matrices representing the group elements can be brought simultaneously to a block-diagonal form (with at least two blocks) by chosing a suitable basis. Group elements in this basis take, e.g., the following form

$\begin{pmatrix} A_{n_1 \times n_1} \end{pmatrix}$	0	0
0	$B_{n_2 \times n_2}$	0
0	0	$C_{n_3 \times n_3}$

An irreducible representation is a representation which cannot be splitted in such sub-blocks. Each subblock of the example given above defines a irreducible representation. Within a subblock each basis vector is linked by group transformations to each other basis vector of the same irreducible representations. In our example, these states therefore have to have the same energy. We conclude

For each group, one can look up the relevant irreducible representations and thus identify the group theoretical origin of degeneracies. For example the symmetry group of a cube (the octahedral group  $O_h$ ) has 48 elements and 10 different irreducible representations which can be 1, 2 or 3 dimensional and have names like  $T_{2g}$  or  $E_g$ . For example, the three functions x, y, z form a three-dimensional irreducible representation called  $T_{1u}$ . The functions  $x^2 - y^2$  and  $2z^2 - x^2 - y^2$  form the two-dimensional representation  $E_g$ . This means that when you apply all symmetries of a cube to the function  $x^2 - y^2$  you will always get a linear combination of  $x^2 - y^2$  and  $2z^2 - x^2 - y^2$ . And if the symmetry transformation acting on two wavefunctions have the same transformation properties as these two functions, then one can identify them as belonging to the same representation.

We will not work out the math needed to construct and identify all irreducible groups (you can look this up under *character theory*) but instead do a simple example. We will again consider **graphene** and we want to study the band-degeneracies arising in the edge of the 1.BZ at the momentum  $\mathbf{k}_0 = \left(\frac{4\pi}{3a}, 0\right)$ . This momentum is equivalent (i.e. connected by reciprocal lattice vectors) to  $\mathbf{k}'_0 = \left(-\frac{2\pi}{3a}, \pm \frac{2\pi}{\sqrt{3a}}\right)$ . The group  $\mathbb{G}_{\mathbf{k}_0}$  thus consists of rotations by  $n\frac{2\pi}{3}$ , reflection  $k_y \to -k_y$  plus combinations of these operations. FIGURE Now we have to find out how these symmetries act on our wavefunctions which we had been writing in the form

$$\psi_{\boldsymbol{k}_0} = \sum_n \alpha_{\boldsymbol{k}_0} e^{i\boldsymbol{k}_0 \cdot \boldsymbol{R}_n} \Phi(\boldsymbol{r} - \boldsymbol{R}_n) + \beta_{\boldsymbol{k}_0} e^{i\boldsymbol{k}_0 \cdot (\boldsymbol{R}_n + \boldsymbol{\Delta})} \Phi(\boldsymbol{r} - (\boldsymbol{R}_n + \boldsymbol{\Delta}))$$

for C atoms located at  $\mathbf{R}_n$  and  $\mathbf{R}_n + \mathbf{\Delta}$  with  $\mathbf{\Delta} = \left(0, \frac{1}{\sqrt{3}}\right) a$ . We have to study how the symmetry transformations act on the two-component vector

$$\mathbf{\Psi} = \left( egin{array}{c} lpha_{m{k}_0} \ eta_{m{k}_0} \end{array} 
ight)$$

The mirror transformation  $k_y \to -k_y$  simply maps an atom on the A sublattice to the B sublattice, thus transforming  $\alpha_{\mathbf{k}} \to \beta_{\mathbf{k}}$  and vice versa. This it is implemented by the Pauli matrix  $\sigma_x$ . A rotation by 120° maps  $\mathbf{k}_0 = \left(\frac{4\pi}{3a}, 0\right)$  to  $\mathbf{k}'_0 = \left(-\frac{2\pi}{3a}, \frac{2\pi}{\sqrt{3a}}\right)$ . This does not affect the term  $e^{i\mathbf{k}_0\cdot\mathbf{R}_n}$  but

$$e^{i\boldsymbol{k}_0\cdot\boldsymbol{\Delta}} \to e^{i\boldsymbol{k}_0'\cdot\boldsymbol{\Delta}} = e^{i\frac{2\pi}{3}}e^{i\boldsymbol{k}_0\cdot\boldsymbol{\Delta}}$$

Therefore, the impact of the rotation on  $\Psi$  is described by the matrix  $\begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{2\pi}{3}} \end{pmatrix}$ . We conclude that the relevant representation of the group  $\mathbb{G}_{k}$  is build up from the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{2\pi}{3}} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{4\pi}{3}} \end{pmatrix}}_{\text{rotations}}, \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{\text{mirror}} + \text{ combinations}$$

It is not possible to find a basis where all those matrices are block-diagonal and this therefore defines a two-dimensional irreducible representation. The group is non-abelian as mirror transformation and rotation do not commute with each other. Only such non-abelian groups can have irreducible representations with dimension larger than 1 (as one can simultaneously diagonalize matrices which commute with each other).

We conclude that the 2-fold degeneracy at  $k_0$  is enforced by the combination of a mirror symmetry and a rotation symmetry. The knowledge of the representation is useful for all types of physical questions which involve low-energy electrons of graphene as it puts strong constraints on how the electrons couple to all types of perturbations.

#### 3.6.3 Accidental degeneracies and Weyl points

There is a second type of band degeneracies which is not linked to symmetries but arises accidentially by fine-tuning of some parameters. These parameters can be something like pressure or an external magnetic field. We will discuss in this section, however, a more generic version where the fine-tuning occurs via the momenta. Let us study a simple two-band Hamiltonian in momentum space which has the generic form

$$H_{k} = \begin{pmatrix} \varepsilon_{k,1} & \Delta_{k}^{*} \\ \Delta_{k} & \varepsilon_{k,2} \end{pmatrix} = \frac{\varepsilon_{k,1} + \varepsilon_{k,2}}{2} \mathbb{I} + g_{k} \cdot \boldsymbol{\sigma}$$

with the real three-component vector

$$\boldsymbol{g}_{\boldsymbol{k}} = \left( \operatorname{Re}\Delta_{\boldsymbol{k}}, \operatorname{Im}\Delta_{\boldsymbol{k}}, \frac{\varepsilon_{\boldsymbol{k},1} - \varepsilon_{\boldsymbol{k},2}}{2} \right)$$

The corresponding eigenenergies are given by

$$\varepsilon_{\pm} = \frac{1}{2} \left( \varepsilon_{\boldsymbol{k}_1} + \varepsilon_{\boldsymbol{k}_2} \right) \pm \sqrt{|\boldsymbol{g}_{\boldsymbol{k}}} = \frac{1}{2} \left( \varepsilon_{\boldsymbol{k}_1} + \varepsilon_{\boldsymbol{k}_2} \right) \pm \sqrt{\left(\frac{\varepsilon_{\boldsymbol{k}_1} - \varepsilon_{\boldsymbol{k}_2}}{2}\right)^2 + |\Delta|^2}$$

The two energies are degenerate if and only if all three components of  $g_k$  vanish ( $\varepsilon_{k_1} = \varepsilon_{k_2}$  &  $\operatorname{Re}(\Delta_k) = 0$  &  $\operatorname{Im}(\Delta_k) = 0$ ). For most points in the 1. BZ these conditions will obviously be never met.

It is, however, useful to ask whether momenta can be found where  $g_k = 0$ . Here we have three equations (one for each component of  $g_k$ ) and three unknown  $(k_x, k_y, k_z)$ , and it may happen "by accident" that such a solution exist where exactly two bands are degenerate. More precisely, one should search for such a case only in materials where either inversion or time-reversal symmetry are broken as otherwise any band is two-fold degenerate.

Let us assume that we have found a point  $k_0$  with  $g_{k_0} = 0$ . Let us do a Taylor expansion around that point

$$H_{\boldsymbol{k}_0+\boldsymbol{q}} \approx \sum_{i,\alpha} q_i A_{i,\alpha} \sigma^{\alpha}$$

We naturally obtain a Hamiltonian linear in momentum. This is called a **Weyl Hamiltonian**, and  $k_0$  is called a **Weyl point**.

To simplify notations, I omitted above the contribution proportional to the idendity matrix which is irrelevant for our discussion. By choosing a suitable basis in spin and momentum space, one can diagonalize the matrix A and obtain

$$H_{\boldsymbol{k}_0+\boldsymbol{q}} \approx v_x q_x \sigma^x + v_y q_y \sigma^y + v_z q_z \sigma^z \tag{3.25}$$

One year after Dirac published his famous Dirac equation, Weyl pointed out in 1929 that a simplified version of it exist which describes massless particle in three space dimensions. While the Dirac equation in 3d is formulated for a 4-component spinor, the Weyl equation contains only half of the degrees of freedom and has thus been formulated only for a two-component spinor. If in the equation above one identifies the three velicities  $v_i$  with the speed of light, one indeed obtains the Hamiltonian corresponding to the Weyl equation. If one formulates, e.g., a theory for left-handed massless neutrinos alone (those couple weakly to matter), one uses Weyl equations.

Weyl point are "topological". This means that they are robust against small perturbations (see also next chapter). Let us add a small hermitian perturbation to Eq. (3.25) of the form

$$p_0 \mathbb{I} + p_x \sigma^x + p_y \sigma^y + p_z \sigma^z$$

The first term does not affect the degeneracy, all other terms can be absorbed by a shift of the momenta  $q_i$  in Eq. (3.25): The Weyl point is shifted by a perturbation of the solid (e.g., by applying pressure) but is not destroyed. Winding numbers are defined in the next section, here I just want to mention that they can be identified with

$$w = \operatorname{sign}(\det A) = \operatorname{sign}(v_x v_y v_z)$$

The Weyl equation has the remarkable property (the so-called "chiral anomaly", not covered in this lecture) that by a combination of parallel aligned electric and magnetic fields, one can pull electrons out of the Fermi sea (w = 1) or push them into the Fermi sea (w = -1). As total particle number is conserved, this means that Weyl points can only come in pairs with opposite winding number. Another consequence is that the only way to get rid of a Weyl point is to move it in momentum space until it meets another Weyl point of opposite winding number.

Above, we discussed only three-dimensional materials. But a similar physics can also occur in two dimensions. Coming back to graphene, we argued in the previous chapter that a combination of mirror and rotation symmetry enforces the band crossing at the K and K' points in graphene. What will happen when one breaks these symmetries, e.g., by smoothly distorting graphene? Here we assume that the distortion does not break the symmetry between A and B sublattice. Starting from

$$H \approx v(q_x \sigma^x + q_y \sigma^y)$$

such distortions are only proportional to  $\sigma^x$  or  $\sigma^y$  (we can, e.g., demand that  $H = \sigma_x H^* \sigma_x$ which follows from demanding both time-reversal invariance and inversion symmetry). The effect of distortion can therefore be aborbed by a shift of momentum

$$H \approx v((q_x - A_x)\sigma^x + (q_y - A_y)\sigma^y)$$

From this we conclude that inhomogeneous deformations of graphene, which lead to a space dependent  $A_i$ , may act very similar to a vector potential arising from an inhomogeneous magnetic field perpendicular to the graphene layer (the sign of the vector potential and thus of the effective magnetic field is, however, opposite for the two Dirac points in graphene).

## 3.7 Topological insulators - a case study

#### 3.7.1 Insulator with surface states

At a surface of a material one can sometimes find - depending on the precise chemistry of the surface - electronic states bound to the surface. For a class of materials, so-called **topological insulators**, metallic surface states are enforced by a topological property of the material. **Topology** here describes properties like winding numbers which are robust under small perturbations (more precisely, perturbations which respect certain symmetries like time-reversal). An insulator is a topological insulators if it possesses a **topologically robust gapless surface states**, which cannot be removed, e.g., by changing the surface chemistry, without closing the bulk gap of the system.

Arguably, the most influential topological state of matter has been discovered in 1980 by von Klitzing: the quantum Hall effect. Two-dimensional electrons in a strong magnetic field exhibit a quantized Hall resistivity,  $R_H = \frac{2\pi\hbar}{e^2n}$ , with an integer *n*. These properties can be traced back to the existence of a *chiral surface state*, where electrons move around the sample as shown and explained in Fig. 3.17. The quantized Hall effect can be measured with an exceptional precision and is thus nowadays used to define the Ohm.



Figure 3.17: Schematic semiclassical picture of a quantum Hall effect. Due to the large magnetic field electrons are localized in the bulk. On the edge, there is, however, a delocalized electronic orbit where electrons move in only one direction (a chiral channel) around the sample.

For a long time the quantum Hall effect was considered an exotic state of matter which only exists in strong magnetic field. Only much later it was realized that it is just an example of a large class of topological systems, topological insulators and topological superconductors. Within this lecture, we will not try to give a general classification of such topological state but instaed discuss one examle in detail with the goal to explain the notion of topology and how it relates to the existence of a surface state.

### 3.7.2 A topological insulator: Quantum spin-Hall effect

We want to investigate a simple theoretical model, which is motivated by the physics of HgTe, the first new topological insulator which was realized in 2007 by the Molenkamp group in Würzburg, following theoretical predictions by Kane and Mele (2005) and Bernevig and Zhang (2006).

To obtain a quantum spin Hall effect one necessarily needs spin-orbit coupling. In our example, this is hidden in the choice of four local orbitals which are used to set up the tight-binding model

$$|s,\uparrow\rangle, |p_x+ip_y,\uparrow\rangle, |s,\downarrow\rangle, |p_x-ip_y,\downarrow\rangle$$

The states  $p_x + ip_y$  state carries angular momentum  $L_z = +\hbar$  and is aligned parallel to the spin and we assume that the state  $|p_x + ip_y, \downarrow\rangle$  needs not to be considered as it has a higher energy.

We assume that all hopping processes conserve spin. Therefore, the  $4 \times 4$  matrix describing the dispersion of the system splits into two  $2 \times 2$  matrices. For fixed momentum we obtain

$$H(\mathbf{k}) = \begin{pmatrix} \tilde{H}_{\mathbf{k}} & 0\\ 0 & T\tilde{H}_{-\mathbf{k}}T \end{pmatrix}$$

In the following, we will focus on the upper  $2 \times 2$  matrix (the lower one is the time-reversed partner of the upper one). Like any Hermitian  $2 \times 2$  matrix, we can write it in terms of the identity matrix I, and the three Pauli matrices defined as usually by

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

Here these matrices do, however, act not on spin space but describe the wave function in the space spanned by the two states  $|s,\uparrow\rangle$ ,  $|p_x + ip_y,\uparrow\rangle$ .

$$\tilde{H}_{k} = \varepsilon_{k} \cdot \mathbb{I} + d_{k} \cdot \sigma_{k}, \quad d_{k} \in \mathbb{R}^{3}$$
(3.26)

The corresponding eigenenergies are given by

$$\varepsilon_{k} \pm |d_{k}|$$

The bands are split and we obtain a band insulators if

$$|\boldsymbol{d}_{\boldsymbol{k}}| > 0 \qquad \forall \, \boldsymbol{k} \in 1. \, \mathrm{BZ}$$

We would like to discuss the "topology" of  $d_k$  interpreted as a function on the 1. BZ. We will only give a hand-waving physics discussion here, the corresponding math goes under the name homotopy theory.

Let us first discuss what we want to call a trivial insulator. An example is obtained when there is no hopping between the orbitals and the two orbitals have a different energy, which would correspond to  $d_{\mathbf{k}} = (0, 0, \Delta E)$ . We define a trivial insulator if we can smoothly deform  $d_{\mathbf{k}}$  while keeping the band gap intact,  $|d_{\mathbf{k}}| > 0 \forall \mathbf{k} \in 1$ . BZ.



Figure 3.18: A configuration of vectors  $d_k$  which cannot be smoothly deformed to the trivial state (all vectors up).

A configuration which does not have this property is shown in Fig. 3.18, where spins point in all possible directions. Sometimes one uses a trick to map the 1.BZ to the surface of a sphere. In our example this can be done by first deforming the vectors on the edge of the 1.BZ so that they all point in parallel direction (which is always possible) and then one projects the 1.BZ on the surface of a sphere in such a way that the center of the 1.BZ maps to the north pole, the edge to the south pole of the sphere. The one obtains the state indicated on the left side of Fig. 3.19, which cannot be smoothly deformed to the trivial state without any winding. This "hedgehog" style configuration is a topological state.

How can we characterize this topological state? Here we can use that the spins in Fig. 3.18 point in all possible directions. Therefore they cover the surface of a sphere and we would like to define a quantity which captures this aspect. Here one can use that the triple product of three neighboring unity vectors can be interpreted as the area covered by those three vectors on the surface of a unit sphere. Using the unit vectore  $\hat{d} = d/|d|$  we thus define the winding number

$$w = \frac{1}{4\pi} \int d^2 \mathbf{k} \ \hat{d}_{\mathbf{k}} \cdot (\partial_{k_x} \hat{d}_{\mathbf{k}} \times \partial_{k_y} \hat{d}_{\mathbf{k}}) \in \mathbb{Z}$$
(3.27)

In a trivial state w vanishes, while it is 1 when the spins cover the full sphere as shown in Fig. 3.18.



Figure 3.19: Illustration of two states with different winding number.

Let us considere a more concrete model by setting

$$d_{\mathbf{k}}^{x} = a \sin k_{x}, \quad d_{\mathbf{k}}^{y} = a \sin k_{z}, \quad d_{\mathbf{k}}^{z} = m - 2b(2 - \cos k_{x} - \cos k_{y})$$

To get some intuition for this model, let us compute

$$d_{k=(0,0)} = (0,0,m), \qquad d_{k=(\pi,\pi)} = (0,0,m-8b)$$

We can expect a non-trivial winding (+1 or -1) for 0 < m < 8b (b > 0) as in this case the two vectors point in opposite directions. In Fig. 3.20 the resulting winding number is shown, which indeed always takes an integer value with the exception of the transition points. At each of these transition points, the gap has to close somewhere.



Figure 3.20: Winding number for b=1.

For the following discussion, we investigate the physics close to m = 0. If we set m = 0, the band gap closes at k = 0. Expanding around this point (now for finite m) we obtain a Dirac equation with a mass m

$$H_{\mathbf{k}} = m\sigma_z + ak_x\sigma_x + ak_y\sigma_y$$

$$= \begin{pmatrix} m & (k_x - ik_y)a \\ (k_x + ik_y)a & -m \end{pmatrix}$$
(3.28)

This is not an accident, but will happen generically at a transition where the winding number changes by 1 (can you see why?). At m = 4, where the winding number changes by 2, one obtains actually two such Dirac points.

A finite winding number has also an important consequences for the eigenstates  $\psi_{\mathbf{k}} = (\psi_{\mathbf{k}}^1, \psi_{\mathbf{k}}^2)$  of the Hamiltonian (3.26). The wavefunction depends only on the direction of  $d_{\mathbf{k}}$  which we can parametrize by the usual angles of polar coordinates,  $\theta_{\mathbf{k}}$  and  $\phi_{\mathbf{k}}$ . With this convention, the eigenstates can, e.g., be parametrized by

#### 3.7. Topological insulators - a case study

$$\psi_{\mathbf{k}} = \left(\cos(\theta_{\mathbf{k}}/2), ie^{-i\phi_{\mathbf{k}}}\sin(\theta_{\mathbf{k}}/2)\right)$$

I would like to stress one aspect of this formula. Apon approaching the south pole ( $d_k$  pointing down) the angle  $\theta_k$  approaches  $\pi$  but the value of  $\phi_k$  depends on how this point is approached. We therefore see that the wave function at this point is a *singular function* of the direction of  $d_k$  and thus on the momentum k. One can remove the singularity at the south pole, e.g., by multiplying the wave function by  $e^{i\phi_k}$  but then the singularity has moved to the north pole. It turns out (this is related to the fact that SU(2) and SO(3) are different groups) that one cannot find any parametrization where  $\psi_k$  is a non-singular function of k. One important consequence concerns, see Sec. 3.5.4. Whenever one tries to construct a Wannier function from a topological band with a finite winding number one does never obtain a well-localized object. Only when one combines several bands in a way that the total winding number vanishes, one can find well-localized Wannier functions.

Our main goal is not to find out what happens at the edge of the topological insulator. Here it is useful to embed the topological insulator (m > 0) into a trivial insulator (m < 0). Doing that (instead of modeling the vacuum) has the advantage that we can do a purely analytical calculation using the Dirac equation describing the system. Let us assume that one edge of the topological insulator is located at x = 0. We now model the boundary by some space-dependent m(x) with

$$m(x) < 0 \text{ for } x < 0, \quad m(x) > 0 \text{ for } x > 0.$$
 (3.29)

and we have to search for surface states of the space-dependent Dirac equation

$$H \approx m(x)\sigma_z - ia\frac{\partial}{\partial x}\sigma_x - ia\frac{\partial}{\partial y}\sigma_y \tag{3.30}$$

which describes the physics of spin  $\uparrow$  electrons in our topological insulator.

Linear differential equations  $H\psi = E\psi$  are much easier to solve than the usual Schrödinger equation. Using the translational invariance in the y direction, one obtains as a solution

$$\psi(x,y) = e^{-\frac{1}{a} \int\limits_{0}^{x} m(x') \,\mathrm{d}x'} \begin{pmatrix} 1\\ i \end{pmatrix} e^{ik_y y}.$$

Importantly, using Eq. (3.29) we can easily show that  $\psi$  is localized around x = 0 as it decays exponentially both for positive and negative x. The energy of this state depends on  $k_y$  and is simply given by

$$E_{\uparrow}(k_y) = a \, k_y$$

This means that the  $\uparrow$  electron propagate with constant velocity  $a = dE/dk_y$  in the +y direction. We could repeat the same calculation for  $\downarrow$  electrons, but this is not necessary as we can obtain the corresponding solution by using the time-reversal symmetry of our initial model. It is obvious that the  $\downarrow$  electrons will also move along a channel localized at the boundary, but their velocity has to be opposite as they are the time-reversed partners of the  $\uparrow$  electrons. Thus their energy is given by

$$E_{\downarrow}(k_y) = -a \, k_y$$



Figure 3.21: Sketch of the edge modes of a 2d topological insulator showing the quantum spin-Hall effect.

We have done the calculation only for a boundary parallel to the y-axis. But as our Dirac equation is rotationally invariant it is obvious that the same physics will occur for an arbitarily oriented boundary. In Fig. 3.21 we give a sketch of the resulting edge states. One can also view this as two time-reversed copies of a quantum-Hall edge channel as depicted in Fig. 3.17. This type of two-dimensional topological insulator thus realises the so-called quantum spin-Hall effect.

The edge is "topologically protected" as long as the system is time-reversal symmetric. The topological robustness is already obvious from our derivation, the existence of the bound state was, for example, independent of the precise functional form of m(x). More importantly, the edge state is robust against scattering from impurities as long as the impurity is a potential scatterer which is invariant under time-reversal symmetry. Under this condition, it cannot scatter an electron (e.g. a spin up) to its time-reversed partner (spin down). Note, however, that the story is different for magnetic impurities which can scatter an up electron into a down electron. Another process which may occur is that induced by the interplay of interactions and disorder a pair of up electrons is scattered into a pair of down electrons moving in the opposite direction, which is not forbidden by any symmetry. We can therefore not expect the same high-precision quantization of transport known from the ordinary Hall effect. Nevertheless, approximately quantized spin transport was observed in experiments of the Molenkamp group in Mainz.

Above, we considered a simplified model. Under the assumption that the spin  $S_z$  is conserved, we were able to reduce the 4x4 matrix problem two 2x2 matrices. For each of these matrices we obtained an integer winding number,  $w \in \mathbb{Z}$ . In real systems with spin-orbit coupling, however,  $S_z$  is not conserved. One can use this extra degree of freedom to unwind some of the states. It turns out that  $w \mod 2$  remains topologically protected and thus two-dimensional insulators with spin-orbit coupling are classified by  $\mathbb{Z}_2$ , a trivial state 0 and a topological state with edge channel 1. For example, one can easily see that states with winding numbers 1 and -1 (spin-up spins running to the left or to the right, respectively) can be smoothly transformed into each other by a simple spin rotation.

#### 3.7.3 Outlook: Topology in three dimensions

The physics which we discussed in a simple two-dimensional example above can be translated to three dimensional system. Here the discussion of winding numbers gets a bit more complicated and less intuitive. But as in the two-dimensional case, see Eq. 3.30, one can use a Dirac equation (a 4-component equation in 3 space dimensions) to describe the surface of a three

dimensional topological insulator. In this case, one finds that the surface is described by 2component 2-dimensional Dirac equation quite similar to the one discussed for graphene. An important difference is, however, that the number of Dirac points is always odd (while it was 4 in the case of graphene). Such surface states have by now been observed in a large class of materials using angular resolved photoemission.

Let me come back to another interesting topological state of matter: a metal with Weyl points introduced in Sec. 3.6.3. There we obtained pairs of Weyl points, where a single Weyl point was described by Eq. (3.25) which we reproduce here

$$H_{\boldsymbol{k}_0+\boldsymbol{q}} \approx v_x q_x \sigma^x + v_y q_y \sigma^y + v_z q_z \sigma^z \tag{3.31}$$



Figure 3.22: A Fermi arc connects the projection of two Weyl points onto the surface of Weyl semimetal.

For simplicity, let us assume that we are interested in a system with 2 Weyl points orientied both on the  $k_z$ -axis and located at the momenta  $(0, 0, \pm k_0)$ . Let us now study the physics of the system for different fixed values of  $k_z$  or, equivalently,  $q_z$ . For each fixed  $q_z$  we can view this as a 2d Hamiltonian of precisely the form which we studied to investigate the topological transition of a quantum spin-Hall system, Eq. (3.28). Thus, we can expect that we can define a 2d winding number which changes when we change  $q_z$ . Our system, for example, may have a winding number w = 1 for  $-k_0 < k_z < k_0$ . For those values of  $k_z$  for which we find that the 2d system is topological, we furthermore expect a topologically protected surface mode. Consider, for example, a surface parallel to the xz plane such that the surface mode is localized in the y direction. In the  $k_x k_z$  plane describing the surface modes, one will thus find at the chemical potential a Fermi surface connecting the momenta  $(0, \pm k_0)$ . This structure is called a Fermi arc, see Fig. 3.22, and it can only exist at the surface of a three-dimensional Weyl metal. In two-dimensional materials in contrast, Fermi surfaces have always to be closed. In a Weyl metal the wave function gets more and more delocalized upon approaching the edge of the Fermi arc and finally vanishes into the bulk of the material.

# SEMICLASSICAL DYNAMICS OF BLOCH ELECTRONS

# 4.1 Equations of Motion & Berry Curvature

## 4.1.1 Berry phase

As a warm-up problem, we consider the slow "adiabatic" change of a quantum system induced by a **slowly** changing Hamiltonian  $H(\lambda_1(t), \lambda_2(t), \lambda_3(t), \ldots)$ . Here  $\lambda(t) = (\lambda_1(t), \lambda_2(t), \lambda_3(t), \ldots)$ are some parameters controlling the Hamiltonian (e.g., 3 components of a magnetic field). Let us denote the ground state of the Hamiltonian for a fixed time by  $|\lambda\rangle$  with

$$H(\boldsymbol{\lambda})|\boldsymbol{\lambda}\rangle = E_{\boldsymbol{\lambda}}|\boldsymbol{\lambda}\rangle$$

Now we use that for a slow change of parameters (changing on a time scale much larger than the gap separating ground state from first excited state) the system always remains in the ground state

$$|\psi(t)\rangle \approx \alpha(t)|\boldsymbol{\lambda}(t)\rangle$$

Using the Schrödinger equation  $i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\psi\rangle = H(t) |\psi\rangle$  we find

$$i\hbar(\dot{\alpha}(t)|\boldsymbol{\lambda}(t)\rangle + \alpha(t)|\partial_t \dot{\boldsymbol{\lambda}}(t)\rangle) \approx E_{\boldsymbol{\lambda}(t)}\alpha(t)|\boldsymbol{\lambda}(t)\rangle$$

Multiplying this from the right-side with  $\langle \boldsymbol{\lambda} |$  we find

$$i\hbar\dot{\alpha} = (E_{\lambda(t)} - i\hbar\langle\lambda|\partial_t\lambda\rangle)\alpha$$

which can easily be solved (assuming  $\alpha(t_0) = 1$ ) to give

$$\alpha(t) = \exp\left[-\frac{i}{\hbar}\left(\int_{t_0}^t E(t') - i\hbar \langle \boldsymbol{\lambda}(t') | \partial_{t'} \boldsymbol{\lambda}(t') \rangle \, \mathrm{d}t'\right)\right] = e^{-i(\phi_d + \phi_B)}$$

The first term is the well-known dynamical phase  $\phi_d = \frac{1}{\hbar} \int_{t_0}^t E_{\lambda(t')} dt'$ . The focus in this section is, however, the second term, the so-called Berry phase

$$\phi_B = -i \int_{t_0}^t \langle \boldsymbol{\lambda}(t') | \partial_{t'} \boldsymbol{\lambda}(t') \rangle \, \mathrm{d}t' = \int_{t_0}^t \mathrm{Im} \left( \langle \boldsymbol{\lambda}(t') | \partial_{t'} \boldsymbol{\lambda}(t') \rangle \right) \, \mathrm{d}t' = i \int_{\lambda(t_0)}^{\lambda(t)} \mathrm{Im} \left( \langle \boldsymbol{\lambda} | \mathrm{d} \boldsymbol{\lambda} \rangle \right) \quad (4.1)$$



Figure 4.1: A trajectory in parameter space.

For the second equality we used that the norm of the wave function is time independent,  $\frac{d}{dt}\langle \boldsymbol{\lambda} | \boldsymbol{\lambda} \rangle = \langle \boldsymbol{\lambda} | \partial_t \boldsymbol{\lambda} \rangle + \langle \partial_t \boldsymbol{\lambda} | \boldsymbol{\lambda} \rangle = 2 \operatorname{Re}(\langle \boldsymbol{\lambda} | \partial_t \boldsymbol{\lambda} \rangle) = 0.$  An important aspect of the Berry phase is that it is a **geometric** property. It is a property of the path  $\boldsymbol{\lambda}(t)$  taken in parameter space, Fig. 4.1, but it does not matter of how this path is parametrized and how fast one is moving along the path (assuming that changes are sufficiently slow to justify the adiabatic approximation). For example  $\boldsymbol{\lambda}(f(t'))$  with an arbitrary smooth function f(t') with  $f(t_0) = t_0$  and f(t) = t will produce just the same Berry phase (as you can check using integration by substitution rules).

Geometric phases play an important role in physics. The most prominent example is the Aharonov-Bohm phase: a charged particle moving on a closed loop in a magnetic field, picks up the geometric phase

$$\oint \frac{q}{\hbar} \boldsymbol{A}(\boldsymbol{r}) d\boldsymbol{r} = \frac{\phi}{\phi_0} 2\pi$$

where  $\phi$  is the magnetic flux going through the path of the charged particle and  $\phi_0 = \frac{2\pi\hbar}{e}$  is the flux quantum. This phase also depends only on a geometric property, the flux enclosed by the path of the particle.

We will use the language of electromagnetism to describe Berry phases. To make the connection clear it is useful to track what happens when one redefines the groundstate wavefunctions in the spirit of a gauge transformation

$$|\boldsymbol{\lambda}
angle 
ightarrow e^{i arphi_{\boldsymbol{\lambda}}} |\boldsymbol{\lambda}
angle$$

with an arbitrary function  $\varphi_{\lambda}$ . This changes  $\langle \lambda(t') | \partial_{t'} \lambda(t') \rangle$  by

$$\langle \boldsymbol{\lambda}(t') | \partial_{t'} \boldsymbol{\lambda}(t') \rangle \rightarrow \langle \boldsymbol{\lambda}(t') | \partial_{t'} \boldsymbol{\lambda}(t') \rangle + i \partial_{t'} \varphi_{\boldsymbol{\lambda}(t')}$$

reminiscent of a gauge transformation. It does **not** affect the Berry phase when it is calculated for a loop with  $\lambda(t) = \lambda(t_0)$ . Only such "gauge invariant" Berry phases can be measured and have experimental consequences (as in the example of the Aharonov-Bohm effect which can only measured when electrons encircle a flux completely).

We finish this paragraph with a classical analog of a Berry phase. Consider in Fig. 4.2the movement of a vector on a sphere using the rule that vectors should remain always parallel

when it moves. As shown in a figure this "parallel" transport leads to a net rotation of the vectors. The net rotation angle thereby corresponds to the solid angle enclosed by the path of the vector,  $4\pi/8 = \pi/2$  in the example shown in Fig. 4.2.



Figure 4.2: Parallel transport on sphere.

#### 4.1.2 Wave-packet dynamics

A classical particle can be described by the equation of motion

$$\dot{\boldsymbol{r}} = \frac{\boldsymbol{p}}{m}, \ \dot{\boldsymbol{p}} = \boldsymbol{F} = e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}).$$

These equations are only valid, if the length scale of the particle is much smaller than the length scale on which external field vary,

 $\lambda \ll a.$ 

Thus it is obvious that it cannot be valid in a solid where  $\lambda \sim a$ . Thus we need quantum mechanics to calculate the band structure  $\varepsilon_n(\mathbf{p})$ .

In the following, we want to consider a slightly different problems. We spit the fields acting onto electrons into internal fields arising from the solid and some "external fields" controlled by the experimentalist, which vary on a much longer length scale  $a_{\text{external}}$ .

$$\lambda \sim a \ll a_{\text{external}}$$

and they also change very slowly on a time scale small compared to  $\hbar/\Delta$  where  $\Delta$  is the bandwidth of the system.

In this case, we might hope to be able to develop a so-called **semiclassical equation of motion** which describes the dynamics on length scales much larger than  $\lambda$ . We can start our analysis by guessing the appropriate classical Hamiltonian. It should include the information from the band structure and thus should read

$$H_{\rm cl}(\boldsymbol{x},\boldsymbol{p}) = \varepsilon_n(\boldsymbol{p}) + V_{\rm ext}(\boldsymbol{x}).$$

From this Hamiltonian we can obtain a first guess for the semiclassical equations of motion using, e.g., the Poisson bracket formulation of classical physics

$$\dot{\boldsymbol{x}} = \{H_{\mathrm{cl}}, \boldsymbol{x}\} = rac{\partial arepsilon_n(\boldsymbol{p})}{\partial \boldsymbol{p}}, \qquad \dot{\boldsymbol{p}} = \{H_{\mathrm{cl}}, \boldsymbol{p}\} = -rac{\partial V_{\mathrm{ext}}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})$$

This first guess is not fully correct because it misses possible extra contributions arising from Berry phases as we will show now. How can one derive classical equations of motion starting from a quantum description? In classical physics one specifies position and momentum simultaneously which is forbidden by Heisenberg's uncertainity relation. Here the trick is to approximate the classical state with a suitably chosen wave packet.

The following discussion follows (G. Sundaram, Q. Niu, Phys. Rev B, 59, 14915 (1999)). Starting point is the Hamiltonian describing electrons in a periodic potential  $H_0 = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r})$ and we will need the Bloch eigenstates

$$H_0\psi_{n\boldsymbol{k}} = \varepsilon_{n\boldsymbol{k}}\psi_{n\boldsymbol{k}}, \quad \psi_{n\boldsymbol{k}} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}}(\boldsymbol{r})$$
(4.2)

We will be interested in the dynamics of the system in the presence of external fields  $A(\mathbf{r}, t), \phi(\mathbf{r}, t)$ , thus the full Hamiltonian is given by

$$H = \frac{(\boldsymbol{p} + e\boldsymbol{A}(\boldsymbol{r}, t))^2}{2m} - e\phi(\boldsymbol{r}, t) + V(\boldsymbol{r})$$

We will now construct a wave packet of width  $\Delta r$  with

$$a_{\text{lattice}} \ll \Delta r \ll a_{\text{external}}$$
 (4.3)

which is centered at position  $x_c$  and momentum  $k_c$  in a given band with index n.



Figure 4.3: A Bloch wavepacket with localized position and quasi-momentum.

As a first guess for the wavepacket (to be refined later slightly) we may use teh following object

$$\psi_{n,k_c,\boldsymbol{x}_c}(\boldsymbol{r}) \sim \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{x}_c)} u_{n\boldsymbol{k}}(\boldsymbol{r}) a_{\boldsymbol{k}}$$
(4.4)  
$$a_{\boldsymbol{k}} \sim e^{-(\boldsymbol{k}-\boldsymbol{k}_c)^2 \cdot \Delta r^2}$$

with

centered at  $\mathbf{k}_c$ . Here  $|a_{\mathbf{k}}|^2$ , the probability that the momentum is occupied is peaked at  $\mathbf{k}_c$ . The width of the peak as function of momentum is given by  $1/\Delta r$ , much smaller than a reciprocal lattice vector. After the Fourier transformation of Eq. (4.4) this gives rise to a wave packet in real space. As sketched in Fig. 4.3, the wavefunction varies rapidly due to the factor  $u_{n\mathbf{k}}(\mathbf{r})$ , but has a smooth envelope of width  $\Delta r$  and is centered around the position  $\mathbf{x}_c$ .

As a next step, we will calculate the time evolution of the wave packet with the goal to obtain an effective equation of motion for  $\boldsymbol{x}_c$  and  $\boldsymbol{k}_c$ . Thus we will study first, how the Hamiltonian acts on this wavefunction. Using that the wave function is localized in space and the inequality (4.3), we can, to leading order, evaluat the potential and vector potential at  $\boldsymbol{x}_c$ .

$$H\psi_n(\boldsymbol{r}) \approx (H_{\rm c}(\boldsymbol{x}_c,t) + \Delta H)\psi(\boldsymbol{r})$$

with

$$H_{\rm c}(\boldsymbol{x}_c,t) \approx \underbrace{\frac{(\boldsymbol{p} + e\boldsymbol{A}(\boldsymbol{x}_{\rm c},t))^2}{2m}}_{\text{shifts } \boldsymbol{p} \text{ by number}} - \underbrace{e\phi(\boldsymbol{x}_{\rm c},t)}_{\text{shifts energy by number}} + V(\boldsymbol{r}).$$

#### 4.1. Equations of Motion & Berry Curvature

The leading correction term can be obtained in a Taylor expansion and it takes the form

$$\Delta H pprox rac{1}{2} \left( (oldsymbol{r} - oldsymbol{x}_c) \cdot rac{\partial H_c}{\partial oldsymbol{x}_c} + rac{\partial H_c}{\partial oldsymbol{x}_c} \cdot (oldsymbol{r} - oldsymbol{x}_c) 
ight).$$

The eigenfunctions of  $H_c$  are simply related to the eigenfunctions of  $H_0$ , Eq. (4.2), by a shift of momentum. We therefore define a shifted momentum

$$\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{q}, \boldsymbol{x}_{\mathrm{c}}, t) = \boldsymbol{q} + e\boldsymbol{A}(\boldsymbol{x}_{\mathrm{c}}, t)$$

The function

$$\psi_{n\boldsymbol{q}}(\boldsymbol{r},t) = \langle \boldsymbol{r} | \psi_{n\boldsymbol{q}} \rangle = e^{\boldsymbol{i} \boldsymbol{k} \cdot \boldsymbol{r}} u_{n,\boldsymbol{k}}(\boldsymbol{r})$$

(note the use of k instead of q on the right-hand side) is then an eigenfunction of  $H_c$  with energy

$$\varepsilon_{\rm c}(\boldsymbol{x}_c, \boldsymbol{q}, t) = E_{\rm c}(\boldsymbol{x}_c, \boldsymbol{k}, t) = \varepsilon_{n\boldsymbol{k}} - e\phi(\boldsymbol{x}_{\rm c}, t) \Big|_{\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{x}_c, \boldsymbol{q}, t)}$$

Now let us again consider our wavepacket, which we write using a slightly modified notation.

$$|\psi\rangle = \int a(\boldsymbol{q},t) |\psi_{n\boldsymbol{q}}\rangle \,\mathrm{d}\boldsymbol{q}, \qquad a(\boldsymbol{q},t) = |a(\boldsymbol{q},t)| e^{-i\gamma(\boldsymbol{q},t)}$$

where  $|a(\mathbf{q},t)|^2$ , has a sharp peak of width  $1/\Delta r$  centered at  $\mathbf{q}_c$ . We need a theory for the average position of the wavepacket and its momentum defined by

$$oldsymbol{x}_c = \langle \psi | oldsymbol{r} | \psi 
angle, \qquad oldsymbol{q}_c = \int oldsymbol{q} |a(oldsymbol{q},t)|^2 \, \mathrm{d}oldsymbol{q}$$

where we assume the normalization condition  $\int |a(\mathbf{q},t)|^2 d\mathbf{q} = 1$ . To evaluate  $\langle \psi | \mathbf{r} | \psi \rangle$  one can use that  $\mathbf{r}e^{i\mathbf{k}\mathbf{r}} = -i\left(\frac{\partial}{\partial \mathbf{k}}e^{i\mathbf{k}\mathbf{r}}\right)$  and a partial integration or just recall that  $\mathbf{r} = i\frac{\partial}{\partial \mathbf{q}}$  is the position operator in momentum space. Using that  $\int |a(\mathbf{q},t)|\partial_{\mathbf{q}}|a(\mathbf{q},t)| d\mathbf{q} = \frac{1}{2}\int \partial_{\mathbf{q}}|a|^2 = 0$ , one obtains

$$\langle \psi | \boldsymbol{r} | \psi \rangle = \int |a(\boldsymbol{q}, t)|^2 \left( \frac{\partial \gamma}{\partial \boldsymbol{q}} + \langle u \Big| i \frac{\partial u}{\partial \boldsymbol{q}} \rangle \right) \, \mathrm{d}\boldsymbol{q}$$

Using that  $|a(q,t)|^2$  is sharply peaked at  $q_c$  this simplifies to

$$\boldsymbol{x}_{c} = \frac{\partial \gamma_{c}}{\partial \boldsymbol{q}_{c}} + \left\langle u | i \frac{\partial u}{\partial \boldsymbol{q}} \right\rangle \Big|_{\boldsymbol{q} = \boldsymbol{q}_{c}}$$
(4.5)

Now we need to obtain an equation for the time evolution of  $\boldsymbol{x}_c, \boldsymbol{q}_c$  starting from the Schrödinger equation. Here we will use a trick by reminding ourself that the Lagrange formulaism was a powerful tool when one wants to change variables. The variable change we want to do here is from the full wavefunction (i.e.,  $a(\boldsymbol{q},t)$ ) to the variables  $\boldsymbol{x}_c, \boldsymbol{q}_c$ . To make progress, we first need a Lagrange function which gives back the ordinary Schrödinger equation.

This requirement is fulfilled by (we omit factors of  $\hbar$  here)

$$L = \left\langle \psi \middle| i \frac{\mathrm{d}}{\mathrm{d}t} - H \middle| \psi \right\rangle = \int \psi^* \left( i \frac{\mathrm{d}}{\mathrm{d}t} - H \right) \psi \tag{4.6}$$

One can check that the Euler-Lagrange equation  $\frac{d}{dt} \frac{\delta L}{\delta \partial_t \psi^*(\mathbf{r})} = \frac{\delta L}{\delta \psi^*(\mathbf{r})}$  gives back the standard Schrödinger equation  $\left(i\frac{\mathrm{d}}{\mathrm{d}t} - H\right)\psi = 0$ . We now want to find an approximation where L depends only on our coordinates,

$$L \approx L(\boldsymbol{x}_{c}, \dot{\boldsymbol{x}}_{c}, \boldsymbol{q}_{c}, \dot{\boldsymbol{q}}_{c})$$

We start by computing the first term in Eq. (4.6),  $\langle \psi | i \partial_t \psi \rangle$ . Here we use exactly the same tricks we used above when calculating  $\langle \psi | i \partial_q \psi \rangle$ 

and obtain

$$\left. \left\langle \psi \left| i \frac{\mathrm{d}\psi}{\mathrm{d}t} \right\rangle \approx \frac{\partial \gamma_{\mathrm{c}}}{\partial t} + \underbrace{\int |a(\boldsymbol{q}, t)| \frac{\partial}{\partial t} |a(\boldsymbol{q}, t)| \,\mathrm{d}\boldsymbol{q}}_{=0 \text{ as } \int |a|^{2} = 1} + \left. \left\langle u \left| i \frac{\partial u}{\partial t} \right\rangle \right|_{\boldsymbol{q} = \boldsymbol{q}_{\mathrm{c}}} + \left. \dot{\boldsymbol{x}}_{\mathrm{c}} \left\langle u \left| i \frac{\partial u}{\partial \boldsymbol{x}_{\mathrm{c}}} \right\rangle \right|_{\boldsymbol{q} = \boldsymbol{q}_{\mathrm{c}}} \right.$$

$$(4.7)$$

It is, perhaps, a surprising result of this calculation that there is no contribution proportional to  $\partial_t \boldsymbol{q}_c$  from this term. The information on the value of  $\boldsymbol{q}_c$  is encoded in  $|\mathbf{a}|$ , whose time-derivative does not contribute. Below, we will, however, replace the partial derivative of  $\gamma_c$  by a total derivative using

$$\frac{d\gamma_c}{dt} = \frac{\partial\gamma_c}{\partial t} + \dot{\boldsymbol{q}}_c \cdot \frac{\partial\gamma_c}{\partial \dot{\boldsymbol{q}}_c} = \frac{\partial\gamma_c}{\partial t} + \dot{\boldsymbol{q}}_c \left(\boldsymbol{x}_c - \left\langle u \middle| i\frac{\partial u}{\partial \boldsymbol{q}} \right\rangle \right)$$
(4.8)

where the last equality was derived using Eq. (4.5). Finally, we have to compute also  $\langle \psi | H | \psi \rangle$  which we approximate by

$$E = \langle \psi \Big| H \Big| \psi \rangle \approx \langle \psi | H_{\rm c} + \Delta H | \psi \rangle \approx \varepsilon_{n \mathbf{k}_{\rm c}} - e \phi(\mathbf{r}, t) + \underbrace{\langle \psi | \Delta H | \psi \rangle}_{\substack{\text{omitted for simplicity only}\\ \text{(orbital magnetization)}}}$$
(4.9)

In the following we will not include the last term which is actually needed to describe correctly how the wavepacket responds to an external magnetic field and gives an extra contribution to the magnetization related to the orbital motion within the wave packet. We can now combine all of the results needed for Eq. (4.6) using Eq. (4.7) with (4.8) plus Eq. (4.9). This gives

$$L = \left\langle u \left| i \frac{\mathrm{d}u}{\mathrm{d}t} \right\rangle + \dot{\boldsymbol{x}}_{\mathrm{c}} \left\langle u \right| i \frac{\partial u}{\partial \boldsymbol{x}_{\mathrm{c}}} \right\rangle + \dot{\boldsymbol{q}}_{\mathrm{c}} \left\langle u \left| i \frac{\partial u}{\partial \boldsymbol{q}_{\mathrm{c}}} \right\rangle - \dot{\boldsymbol{q}}_{\mathrm{c}} \cdot \boldsymbol{x}_{\mathrm{c}} - \tilde{E} + \frac{d\gamma_{c}}{dt} \right\rangle$$

Finally, we observe that u depends on  $x_c$  and t only via the vector potential which simply shifts the momentum. Therefore we can set

$$\frac{\partial u}{\partial \boldsymbol{x}_{\rm c}} = \frac{\partial u}{\partial \boldsymbol{A}} \frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}_{\rm c}} = e \frac{\partial u}{\partial \boldsymbol{q}} \frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}_{\rm c}}, \qquad \frac{\partial u}{\partial t} = \frac{\partial u}{\partial \boldsymbol{A}} \frac{\partial \boldsymbol{A}}{\partial t} = e \frac{\partial u}{\partial \boldsymbol{q}} \frac{\partial \boldsymbol{A}}{\partial t}$$

and collect all terms proportional to  $\frac{\partial u}{\partial q}$ 

$$L(\boldsymbol{x}_{c}, \dot{\boldsymbol{x}}_{c}, \boldsymbol{q}_{c}, \dot{\boldsymbol{q}}_{c}) = \left\langle u \middle| i \frac{\partial u}{\partial \boldsymbol{q}_{c}} \right\rangle \underbrace{\left( e \frac{\partial \boldsymbol{A}}{\partial t} + e \frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}_{c}} \dot{\boldsymbol{x}}_{c} + \boldsymbol{q}_{c} \right)}_{= \dot{\boldsymbol{k}}_{c} \text{ as } \boldsymbol{k}_{c} = \boldsymbol{q}_{c} + e \boldsymbol{A}(\boldsymbol{x}_{c}, t)} - \tilde{E} + \dot{\boldsymbol{q}}_{c} \cdot \boldsymbol{x}_{c} + \underbrace{\frac{d}{dt} \left( \gamma_{c} - \boldsymbol{q}_{c} \cdot \boldsymbol{x}_{c} \right)}_{\text{total derivative}}}_{\text{can be dropped from the action}}$$

#### 4.1. Equations of Motion & Berry Curvature

Finally, we rename the variables  $x_c \to x$  and use  $k_c \to k$  as the momentum variable instead of  $q_c$ . After dropping the last total derivative term in L which (as all total derivative terms, which are just surface terms for the action) does not contribute to the Euler-Lagrange equations we arrive at our final result

$$L = -\varepsilon_{n\boldsymbol{k}} + e\phi(\boldsymbol{x}_{c}, t) + \dot{\boldsymbol{x}} \cdot (\boldsymbol{k} - e\boldsymbol{A}(\boldsymbol{x}, t)) + \dot{\boldsymbol{k}} \cdot \left\langle u \middle| i \frac{\partial u}{\partial \boldsymbol{k}} \right\rangle$$
(4.10)

The remaining task is to extract the Euler Lagrange equations from this formula. We start with the coordinate x

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}_m} = \dot{k}_m - \sum_n \frac{\partial A_m}{\partial x_n} \dot{x}_n - e\frac{\partial A}{\partial t} \stackrel{!}{=} \frac{\partial L}{\partial x_m} = e\frac{\partial \phi}{\partial x_m} - \sum_n \frac{\partial A_n}{\partial x_m} \dot{x}_n$$

We thus obtain the following equation of motion

$$\dot{k}_m = -eE_n - eF_{mn}\frac{\mathrm{d}x_n}{\mathrm{d}t} = -e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})_m \tag{4.11}$$

where we used the electric and magnetic field and the spatial components of the field-strength tensor which encode the B field

$$\boldsymbol{E} = -\boldsymbol{\nabla}\phi - \frac{\partial \boldsymbol{A}}{\partial t}, F_{mn} = \left(\frac{\partial A_n}{\partial x_m} - \frac{\partial A_m}{\partial x_n}\right) = \varepsilon_{mnp}B_p.$$

More interesting is the equation of motion for the momentum

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{k}_m} = \frac{\mathrm{d}}{\mathrm{d}t}\left\langle u \middle| i\frac{\partial u}{\partial k_m} \right\rangle \stackrel{!}{=} \frac{\partial L}{\partial k_m} = \dot{x}_m - \frac{\partial \varepsilon_{n\mathbf{k}}}{\partial k_m} + \sum_n \dot{k}_n \frac{\mathrm{d}}{\mathrm{d}k_m} \left\langle u \middle| i\frac{\partial u}{\partial k_n} \right\rangle$$

This can be brought to a form very similar to Eq. (4.11)

$$\dot{x}_m = \frac{\partial \varepsilon_{nk}}{\partial k_m} + \tilde{F}_{mn} \frac{\partial k_n}{\partial t}$$

using the following definition for the "Berry curvature"

$$\tilde{F}_{mn} = \frac{\partial \tilde{A}_n}{\partial k_m} - \frac{\partial \tilde{A}_m}{\partial k_n}$$
$$\tilde{A}_i(\mathbf{k}) = -i \left\langle u_{n\mathbf{k}} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_i} \right\rangle = \operatorname{Im} \left[ \left\langle u_{n\mathbf{k}} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_i} \right\rangle \right]$$
(4.12)

with a Berry-vector potential  $\tilde{A}$  defined in k-space. Note that the vector potential is real as can be shown from  $\partial_k \langle u | u \rangle = \partial_k 1 = 0$ . In analogy to the magnetic field, we also define

$$\Omega_{k} = \operatorname{rot}_{k} \tilde{A}(k)$$

a "dual" B field in k-space, where all derivatives are in k space. With this definitions we arrive at the following equations of motion

$$\dot{\boldsymbol{x}}_{c} = \frac{\partial \varepsilon}{\hbar \partial \boldsymbol{k}_{c}} - \dot{\boldsymbol{k}}_{c} \times \boldsymbol{\Omega}_{\boldsymbol{k}}$$
$$\hbar \dot{\boldsymbol{k}}_{c} = -e(\boldsymbol{E} + e\boldsymbol{x}_{c} \times \boldsymbol{B})$$
(4.13)

where we added the necessary factors of  $\hbar$ . As discussed above, we omitted here a B-field dependence of the energy arising both from the electronic spin (which we did not take into account in the derivation) and the orbital magnetization.

The by far most interesting term is a new contribution to the velocity of the particle, the so-called "anomalous velocity"

$$-oldsymbol{k}_{
m c} imes oldsymbol{\Omega}_{oldsymbol{k}}.$$

It takes the form quite similar to a magnetic field, but here the role of k-space and position space are interchanged.

Above we used the language of electromagnetism. On a fundamental level, electromagnetism is closely connected to the U(1) gauge invariance, i.e., to the fact that one can muliply the wavefunction by arbitrary space-dependent factors  $e^{i\phi(\mathbf{r})}$ . Similarly, the emergence of a gauge field in momentum space should therefore be associated to a gauge freedom in momentum space. Here we have to recall that normalized eigenfunctions of a Hamiltonian are only defined up to an arbitrary phase. Applying this to the eigenfunctions  $u_{n\mathbf{k}}$  of the Bloch Hamiltonian, we conclude that physical properties must be invariant under the "Gauge transformation"

$$u_{n\mathbf{k}} \to e^{i\phi_{\mathbf{k}}} u_{n\mathbf{k}}$$

Under this transformation, the Berry vector potential does change

$$ilde{m{A}}_{m{k}} 
ightarrow ilde{m{A}}_{m{k}} + \partial_{m{k}} \phi_{m{k}}$$

but the Berry curvature and  $\Omega_k$  remain invariant. Similarly, the "Berry phase"  $\int \tilde{A}(k) dk$  is gauge independent when integrated over a closed contour. Here it is useful to stress that the physics of the anomalous velocity is indeed precisely the physics of Berry phases discussed in Sec. (4.1.1). A slow change of momentum gives rise to a Berry phase of the Bloch wavefunction, which is reflected in the Berry curvature.

As  $\Omega = \operatorname{rot} \tilde{A}(\mathbf{k})$  it follows immediately that  $\operatorname{div}_{\mathbf{k}} \Omega_{\mathbf{k}} = 0$ . More precisely, this is valid up to singular points. The analysis above was done for a single, smooth band and thus does not cover band-degeneracies as they occur on Weyl points. Indeed, one can show (using, e.g., the formula for the Berry curvature discussed below) that a Weyl point at momentum  $\mathbf{k}_0$  acts as a source of Berry curvature,  $\operatorname{div}_{\mathbf{k}} \Omega_{\mathbf{k}} = \pm 2\pi \delta(\mathbf{k} - \mathbf{k}_0)$  for Weyl points with winding number  $\pm 1$ .

### 4.1.3 Anomalous Hall effect and TKNN formula

An important question is under which condition we can expect Berry curvatures to be present. From the discussion above it is clear that the presence of Weyl points enforces a finite Berry curvature. But we argued above that Weyl points do not occur in systems which have both inversion (P) and time-reversal symmetry (T). We should therefore check how these symmetries affect the Berry curvature. Time reversal implies complex conjugation and therefore an extra minus sign.

$$T: \dot{F}_{mn}(\boldsymbol{k}) \to -\dot{F}_{mn}(-\boldsymbol{k})$$
$$P: \tilde{F}_{mn}(\boldsymbol{k}) \to \tilde{F}_{mn}(-\boldsymbol{k})$$

When therefore T & P are present, we find  $\tilde{F}(\mathbf{k}) = -\tilde{F}(\mathbf{k}) = 0$ . Therefore either T or P (or both) must be broken to obtain a finite Berry curvature.

This makes ferromagnets immediately prime candidates to search for effects. Indeed, many ferromagnets show an very large "anomalous" Hall effect even when no external magnetic field is applied. This can easily be understood from our formulas. For B = 0 and a constant electric field, the anomalous velocity is simply given by

$$-\dot{oldsymbol{k}}_{
m c} imesoldsymbol{\Omega}_{oldsymbol{k}}=rac{e}{\hbar}oldsymbol{E} imesoldsymbol{\Omega}_{oldsymbol{k}}$$

Thus we get an extra velocity  $\perp E$ . The occupation of a state with momentum k is given by the Fermi function. For an electric field in x-direction, we therefore obtain a current in y direction. Each k mode contributes a current  $e\partial_t x$  and therefore we find for a d-dimensional system

$$\sigma_{xy}^{\text{anomalous}} = \frac{e^2}{\hbar} \int_{1.\text{BZ}} \frac{d^d \mathbf{k}}{(2\pi)^d} f_{\mathbf{k}} \Omega_{\mathbf{k}}^z$$
(4.14)

where  $f_{\mathbf{k}}$  is the probability that a state with momentum  $\mathbf{k}$  is occupied. This we can compute for  $\mathbf{E} = 0$  (as the anomalous velocity is already linear in the field) and is thus simply given by a Fermi function,  $f_{\mathbf{k}} = 1/(e^{\beta(\epsilon_{n\mathbf{k}}-\mu)}+1)$ .

A special case of this formula occurs in **band insulators in two dimensions**. In a band insulator at low T,  $f_{\mathbf{k}} = 1$  for all  $\mathbf{k}$ . Furthermore, other contributions to the Hall effect (discussed in the next chapter) vanish and the Hall conductivity is exactly given by the integral over the Berry curvature

$$\sigma_{xy} = \frac{e^2}{\hbar (2\pi)^2} \int_{1.\text{BZ}} d^2 \boldsymbol{k} \,\Omega_{\boldsymbol{k}}^z = \frac{ie^2}{\hbar (2\pi)^2} \int_{1.\text{BZ}} d^2 \boldsymbol{k} \left( \langle \partial_{k_x} u | \partial_{k_y} u \rangle - \langle \partial_{k_y} u | \partial_{k_x} u \rangle \right)$$
$$= \frac{e^2}{2\pi\hbar} n, \qquad n \in \mathbb{Z}$$
(4.15)

This is the famous TKNN formula, named after the four authors Thouless, Kohmoto, Nightingale, and den Nijs of a famous paper from 1982 which was (among other work) honored by the Nobel prize for Thouless in 2016. The last equality claims that when  $\Omega_{\mathbf{k}}^z$  is integrated over the full BZ, one obtains a quantized value (called a Chern number, see below). Thus it gives rise to a quantized Hall effect, more precisely, an "quantum anomalous Hall effect", which can occur even without any external magnetic field. It was first observed in 2013 on the magnetic surface of a topological insulator (Chang et al, Science 340, p 167 2013). This quantum Hall state shares all qaulitative properties with the ordinary quantum Hall effect including the presence of *n* chiral edge states, see Sec. 3.7.



Figure 4.4: To evaluate the TKKN formula for an insulator, one integrates the Berry vectorfield along the edge of the 1. BZ.

Why is the integral over the Berry curvature quantized? To see this, we first use Stoke's theorem to rewrite the integral over the 1.BZ as a line integral over the Berry vector potential along the edge of the 1.BZ, see Fig. 4.4,

$$\int_{1.\mathrm{BZ}} \Omega_{\boldsymbol{k}}^{z} = \oint_{\mathrm{edge 1.BZ}} d\boldsymbol{k} \tilde{\boldsymbol{A}}_{\boldsymbol{k}}.$$

The next step is to realize that the edges of the 1.BZ are connected by reciprocal lattice vectors  $G_i$  and can thus be identified. More precisely, the wave function at k and  $k + G_i$  are identical up to a phase

$$\left|u_{\boldsymbol{k}+\boldsymbol{G}_{i}}\right\rangle=e^{i\varphi_{\boldsymbol{k}+\boldsymbol{G}_{i}}}\left|u_{\boldsymbol{k}}\right\rangle$$

Thus, the contributions from  $\left\langle u_{n\boldsymbol{k}} \middle| \frac{\partial u_{n\boldsymbol{k}}}{\partial k_i} \right\rangle$  cancel exactly in the line integral (for example, when integrating up on the left and down on the right-side in Fig. 4.4). The only remaining contribution can arise from the phase

$$\int_{1.\mathrm{BZ}} \Omega_{\boldsymbol{k}}^{z} = \oint_{\mathrm{edge 1.BZ}} d\boldsymbol{k}(-i) e^{-i\varphi_{\boldsymbol{k}}} \partial_{\boldsymbol{k}} e^{i\varphi_{\boldsymbol{k}}} = \oint_{\mathrm{edge 1.BZ}} \partial_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}$$

where one can, e.g., choose  $\varphi_{\mathbf{k}}$  being zero along half of the integration contour and finite on the other. More importantly, one can arrange the gauge in such a way that  $\varphi_{\mathbf{k}}$  is a continuous and *periodic* function along the edge of the 1.BZ. Using that  $\int_{\lambda_1}^{\lambda_2} \partial_\lambda \varphi_\lambda = \varphi_{\lambda_2} - \varphi_{\lambda_1}$  we conclude that for periodic boundary conditions, the integral has to be a multiple of  $2\pi$  as the phase  $\varphi$ is defined modulo  $2\pi$ .

$$\int_{1.\mathrm{BZ}} \Omega_{\mathbf{k}}^{z} = \oint_{\mathrm{edge \ 1.BZ}} \partial_{\mathbf{k}} \varphi_{\mathbf{k}} = 2\pi \ n \qquad n \in \mathbb{Z}$$

which concludes our derivation of the anomalous quantum Hall effect, Eq. (4.15).

It is also instructive to calculate the Berry curvature for a two-band model where

$$H_{k} = d_{k} \cdot \sigma$$

which we also considered in Sec. 3.7.2 (in that section we assumed, however, that this Hamiltonian combines with its time-reversed partner to a 4-band model, now we assume that we just have these two bands, e.g., in a ferromagnet with broken time-reversal symmetry). By an explicit evaluation of the eigenfunctions, one obtains for the *i*th component of the Berry curvature of the lower of the two bands

$$\Omega^{i}_{\boldsymbol{k}} = \frac{1}{4} \epsilon_{inm} \hat{d}_{\boldsymbol{k}} \cdot (\partial_{k_{n}} \hat{d}_{\boldsymbol{k}} \times \partial_{k_{m}} \hat{d}_{\boldsymbol{k}})$$

which simplifies in two dimensions to

$$\Omega_{\boldsymbol{k}}^{z} = \frac{1}{2} \hat{d}_{\boldsymbol{k}} \cdot (\partial_{k_{x}} \hat{d}_{\boldsymbol{k}} \times \partial_{k_{y}} \hat{d}_{\boldsymbol{k}})$$

For the upper band a curvature of opposite sign is obtained. Using this in Eq. (4.15) to calculate the Hall conductivity we obtain

$$\sigma_{xy} = \frac{e^2}{2\pi\hbar}w$$

where w is the topological winding number of the unit vectors  $\hat{d}_k$  defined and discussed in Sec. 3.7.2.

# 4.2 Bloch Oscillations and Zener Tunneling

As a next step we will discuss a few applications of the semiclassical equations of motion. We start with the simplest case, where there is no electric field and where all Berry curvatures are absent (e.g., due to the presence of both time-reversal and inversion symmetry, B = 0,  $\Omega_k = 0$ . In this case we can simply solve the equation for the momentum

$$\dot{\boldsymbol{p}} = -e\boldsymbol{E} \qquad \Longrightarrow \qquad \boldsymbol{k} = rac{\boldsymbol{p}}{\hbar} = -rac{e}{\hbar}\boldsymbol{E}t + \boldsymbol{k}_0$$

Using that

$$\dot{oldsymbol{x}} = rac{\partialarepsilon}{\hbar\partialoldsymbol{k}} = oldsymbol{v}(oldsymbol{k})$$

we obtain

$$x(t) = \boldsymbol{x}(0) + \int_{0}^{t} \boldsymbol{v}(\boldsymbol{k}(t')) \, \mathrm{d}t'.$$

Now it is important to realize that  $\varepsilon(\mathbf{k})$  is a **periodic** function of momentum. Therfore also period  $\mathbf{v}(\mathbf{k})$  has to be periodic which implies that  $\mathbf{x}(t)$  is periodic: the electrons perform so-called **Bloch oscillations**.

For example for 
$$\varepsilon(k) = -D\cos(ka)$$
 one obtains  $x(t) = x(0) + \frac{1}{\hbar} \int_{0}^{t} aD\sin\left(\frac{e}{\hbar}Et'a + k_0a\right) dt'$ 

and thus

$$x(t) = x(0) - \frac{D}{eE} \left[ \cos\left(\frac{e}{\hbar}Et'a + k_0a\right) - \cos\left(k_0a\right) \right]$$

One way to interpret the Bloch oscillations is that the electron increases its momentum until its wavelength matches the lattice constant. Then it undergoes a Bragg reflection, changing its momentum by a reciprocal lattice vector and gets again accelerated. An experimental consequence is, e.g., a radiation with a frequency  $\omega = \frac{e}{\hbar} Ea$ .



Figure 4.5: Bloch oscillations of an electron.

Bloch oscillations are very difficult to observed in solids. The problem becomes evident when one looks at the typical amplitude of the Bloch oscillation. Above we found that  $\Delta x \approx \frac{D}{eE}$ , where D is the bandwidth. For example, if we use  $E \sim 10 \frac{V}{\text{mm}}$ ,  $D \sim 1 \text{ eV}$  we obtain  $\Delta x \approx 0.1 \text{ mm} = 10^6 \text{ Å}$ . Along this distance it is very likely that the electron meets some impurity of the crystal or is scattered by other electrons or phonons. In this case the Bloch oscillations are destroyed.

Bloch oscillations have, however, been observed in artificial periodic "superlattices" with large lattice constants a. Perhaps more interesting, Bloch oscillations of ultracold atoms moving in the standing wave of a laser have been observed. Here one can use, e.g., replace the electric field by the gravitational force. Using that frequencies can be measured with very high precision, this setup can be used to measure the gravitational field with a very high precision using that the frequency depends only on the force and the lattice constant,  $\omega = \frac{F}{\hbar}a$ .



Figure 4.6: Zener tunneling: In the presence of an electric field, electrons can tunnel from one band to the next-higher band.

Even in the absence of scattering, due to their classical nature, the semiclassical equations of motion are not exact. For example, in the presence of an electric field, it is the possibility for an electron to tunnel into the next higher band (Zener tunneling). This is best understood by studying the semiclassical energy landscape defined by

$$\varepsilon_n(k) - eEx$$

as shown in Fig. 4.6, where the dashed regions are the classical allowed energies. The gap region is energetically forbidden and thus unaccessible within a classical theory. Quantum mechanically, however, the tunneling from one band to the neighboring band is possible.

potential



Figure 4.7: Barrier tunneling schematic.

To get a rough estimate for the order of magnitude of this effect, let us estimate the tunneling rate in a very rough way. Recall the formula for the tunneling through a barriere as depicted in Fig. 4.7. In this case the tunneling rate is exponentially small in the width  $\Delta x$  of the potential,  $\Gamma_t \propto e^{-2\Delta x \sqrt{2mV_0}/\hbar}$ . In our case, the distance of tunneling is given by  $\Delta x \approx \frac{E_{gap}}{eE}$ , where  $E_{gap}$
is the gap between the two bands, see Fig. 4.6. Therefore the rate of tunneling is exponentially small for small electric fields

$$\Gamma_{Zener} \propto e^{-\frac{E_0}{E}}$$

where an order of magnitude estimate for  $E_0$  is given by the "typical" electric fields in a solid,  $E_0 \sim \frac{1 \text{ eV}}{1 \text{ Å}}$  which is much larger than any external fields which can be applied to a solid. Therefore Zener tunneling is very rare in a generic solid-state system (but may occur, e.g., in some cold-atom realizations).

## 4.3 Bloch Electrons in a Magnetic Field

The previous chapter discussed the dynamics of electrons in an electric field. Now we will consider the case when there is no electric field but instead only a constant magnetic field B. Furthermore, we assume for simplicity that there are no Berry curvature effects,  $\Omega_k = 0$ . In this case the equations of motion are simply given by

$$\dot{\boldsymbol{r}} = \boldsymbol{v}_{\boldsymbol{k}} = rac{\partial \epsilon_{n \boldsymbol{k}}}{\partial \boldsymbol{k}}, \qquad \dot{\boldsymbol{k}} = -rac{e}{\hbar c} \dot{\boldsymbol{r}} imes \boldsymbol{B}$$

Importantly, the change of momentum is always perpendicular to the velocity. This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_{n\boldsymbol{k}} = \frac{\partial\epsilon_{n\boldsymbol{k}}}{\partial\boldsymbol{k}}\cdot\dot{\boldsymbol{k}}\propto\boldsymbol{v}_{\boldsymbol{k}}\cdot(\boldsymbol{v}_{\boldsymbol{k}}\times\boldsymbol{B}) = \boldsymbol{0}$$

Therefore energy is conserved and the motion in k-space is along an equal-energy contour perpendicular to the magnetic field. The momentum parallel to the B-field,  $k_{\parallel}$ , is thereby conserved. This geometric construction, a cut of the three-dimensional Fermi surface with a plane perpendicular to the magnetic field, gives a one-dimensional curve  $\mathbf{k}(\lambda)$  along which  $\mathbf{k}(t)$  has to move. This already gives a pretty good picture of the motion of electrons in momentum space which we will use below.



Figure 4.8: Picture of a Fermi surface. Cutting the 3d Fermi surface by a plane perpendicular to the magnetic field allows to determine a contour on which the particle moves in momentum space.

If  $\mathbf{k}(t)$  is known, one can also construct the motion in real space in a plane perpendicular to the magnetic field. To do this, just consider the following product

$$\hat{\boldsymbol{B}} \times \hbar \dot{\boldsymbol{k}} = -\frac{eB}{c} \left( \dot{\boldsymbol{r}} - \hat{\boldsymbol{B}} (\hat{\boldsymbol{B}} \cdot \dot{\boldsymbol{r}}) \right) = -\frac{eB}{c} \dot{\boldsymbol{r}}_{\perp}$$
(4.16)

Here  $\hat{B}$  is the unit vector parallel to B and  $r_{\perp}$  is the projection of the coordinate  $r \perp$  to B. The equation implies that the trajectories in momentum space and real space are simply linked by a cross product

$$\boldsymbol{r}_{\perp}(t) - \boldsymbol{r}_{\perp}(0) = -\frac{\hbar c}{eB} \hat{\boldsymbol{B}} \times (\boldsymbol{k}(t) - \boldsymbol{k}(0))$$

Geometrically, the cross product can be interpreted as a 90 degree rotation.



Figure 4.9: Case 1 - closed curves in k space.

Now we should distinguish two cases. In case 1, Fig. 4.9, the curve is closed. Therefore on naturally obtains a periodic orbit in k-space and thus also in position space. Later, we will need only on quantity: the time it takes to go once around the orbit

Case 1: Closed curves in k space

$$T = \int_{\mathcal{C}\perp \mathbf{B}} \left| \frac{\mathrm{d}\mathbf{k}}{\mathbf{k}} \right| = \frac{\hbar^2 c}{eB} \oint \frac{|\mathrm{d}\mathbf{k}|}{\left| \left( \frac{\mathrm{d}E}{\mathrm{d}\mathbf{k}} \right)_{\perp} \right|} = \frac{\hbar^2 c}{eB} \frac{\partial A(E, k_{\parallel})}{\partial E}$$

Here  $A(E, k_{\parallel})$  is the area enclosed by  $\mathbf{k}_{\perp}(t)$  for a cut at a given energy E and momentum  $k_{\parallel}$ . To explain the formula above, it is useful to rewrite the integral using energy and the momentum parallel to the contour as variables

$$A = \int_{E_k \leqslant E} \mathrm{d}k_\mathrm{c} \mathrm{d}k_\perp = \int_{E_k \leqslant E} \mathrm{d}k_\mathrm{c} \left| \frac{\mathrm{d}k_\perp}{\mathrm{d}E} \right| \mathrm{d}E$$

which shows that

$$\frac{\partial A(E,k_{\parallel})}{\partial E} = \oint \frac{|\mathrm{d}\boldsymbol{k}|}{\left| \left( \frac{\mathrm{d}E}{\mathrm{d}\boldsymbol{k}} \right)_{\perp} \right|}$$

The (generalized) cyclotron frequency describing the motion on closed orbits is given by

$$\omega_{\rm c}(E,k_z) = \frac{2\pi}{T} = \frac{eB}{cm^*} \quad \text{with} \quad m^*(E,k_z) = \frac{\hbar^2}{2\pi} \frac{\partial A}{\partial E} \tag{4.17}$$

One can easily check that in the case of free electrons with mass m, where  $A(E) = \pi k_F^2 = 2\pi m E/\hbar^2$  one obtains  $m^* = m$ .



Figure 4.10: Case 2 - open curves in k space.

A second option is depicted in Fig. 4.10, the cut of the Fermisurface can also result in an open orbit, where  $k_{\perp}(t)$  is only smooth when one uses an extended zone scheme, drawing also higherorder Brillouine zones. Here  $k_{\perp}(t)$  and therefore also  $\mathbf{r}_{\perp}(t)$  grow indefinitely. The two different cases - open or closed trajectories - leads also to different physical properties. For example, it turns out that for closed orbits the resistivity saturates at a finite value for large magnetic fields while it grows larger and larger for open trajectories (transport theory is discussed in the next chapter).

There are also important quantum effects which cannot be directly described by classical equations of motion but are based on (semiclassical) quantization.



**Figure 4.11:** Oscillations of density of states in the presence of a B-field (free electrons, without disorder). Disorder will broaden the curve and smoothen the peaks but a certain periodic structure will remain as long as the disorder is not too strongd.

To understand the physics of this, it is useful to remember the physics of Landau levels. For a free electron (without a periodic potential), electrons move classically on closed orbits with cyclotron frequency  $\omega_{\rm c} = \frac{eB}{mc}$ . Quantum mechanically one can describe electrons in a constant magnetic field by a corresponding harmonic oscillator and thus the total energy is given by

$$E = \frac{(\hbar k_{\parallel})^2}{2m} + \hbar \omega_{\rm c} \left( n + \frac{1}{2} \right), \quad n \in \mathbb{N}_0$$

As a consequence, the density of state obtains a periodic structure: a sequence of  $1/\sqrt{\omega}$  peaks starting at each Landau level energy, see Fig. 4.11. Disorder in a material will smear out the sharpest features but (as long as  $\omega_c \tau \gtrsim 1$  with  $\tau$  defined in the next chapter) some oscillatory features often remain if the system if there are not too many defects in the sample.



Figure 4.12: Oscillations of magnetization as function of 1/B.

While we will not give a derivation (for large quantum numbers n one can use "semiclassical quantization" in the spirit of the Bohr-Sommerfeld quantization well known from the hydrogen atom), it is plausible that the periodic motion of electrons will give rise to similar effect with a modified cyclotron frequency given by Eq. (4.17).



**Figure 4.13:** Quantum oscillations are determined by cuts of the Fermi surface perpendicular to the magnetic field with an extremal (maximal or minimal) area as function of  $k_{\parallel}$ .

As a consequence any experimental quantities change periodically when changing B. As we will show below, this physics is best analyzed by plotting the physical quantity as function of 1/B to identify periodic structures as sketched in Fig. 4.12. The period of these oscillations is given by

$$\Delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar c} \frac{1}{A_{\rm e}} \tag{4.18}$$

where  $A_{\rm e}$  are extremal cross-sectional area of Fermi surface in planes  $\perp B$ , as sketched in Fig. 4.13. The derivation is given below.

When one observes such periodic structions in the magnetization, M(B), this is called the de Haas-van Alphen effect. Similarly, Shubhnikov-de Haas oscillations describe periodic oscillations in the resistivity. The importance of these oscillations arise from the fact that they allow to extract quantitatively the shape of Fermi surface by measuring the relevant extremal areas as a function of the orientation of the magnetic field. In combination with theoretically calculated band structures one can often obtain a very good quantitative understanding of the shape of the Fermi surface. A more detailed analysis of, e.g., the temperature dependence allows furthermore also to extract quantities like the effective mass of the electronic bands.

We will now derive (4.18) under the assumption that the cyclotron frequency is much smaller than the Fermi energy. Let us define  $E_n(k_z)$  as the energy of the *n*-th peak in density of states such that

$$E_{n+1}(k_z) - E_n(k_z) = \hbar\omega_c = \frac{2\pi}{T}, \text{ with } T(k_z) = \frac{\hbar^2 c}{eB} \frac{\partial A(E, k_z)}{\partial E}$$

as derived above. Using this formula, we find that

$$\left(E_{n+1} - E_n\right)\frac{\partial A}{\partial E} = \frac{2\pi eB}{\hbar c}$$

In the limit that the Fermi energy is much larger than the cyclotron frequency, we can interpret the term on the left-hand side as a Taylor expansion and conclude that

$$A(E_{n+1}, k_z) - A(E_n, k_z) \approx \frac{2\pi eB}{\hbar c}$$

or, equivalently, that

$$A(E_n, k_z) = (n + \text{const}) \frac{2\pi eB}{\hbar c} \quad \forall n \gg 1, \text{constant} = \text{unknown}$$

Let us now consider some physical quantity as function of B and let us denote the position of subsequent peaks (or dips) in this quantity by  $B_1, B_2$ . The area A of the Fermi surface is fixed and thus we can expect that

$$A = (n + \text{const}) \frac{2\pi eB_1}{\hbar c} = (n + 1 + \text{const}) \frac{2\pi eB_1}{\hbar c}$$

Thus,  $\frac{A}{B_2} - \frac{A}{B_1} = \frac{2\pi e}{\hbar c}$ , or equivalenty,

$$\frac{1}{B_2} - \frac{1}{B_1} = \frac{2\pi e}{\hbar c} \frac{1}{A(E_{\rm F}, k_z)}$$

The equation above has been derive for fixed  $k_z$  and it predicts oscillations in 1/B with a  $k_z$  dependent frequency. But most experimental observables will average over  $k_z$  and thus most of these oscillations will cancel with each other. Only oscillations with an extremal area,  $\partial_{k_{\parallel}}A = 0$ , will survive averaged out when averaging over  $k_z$  which concludes the derivation of Eq. (4.18).

We finish this section with a short discussion of the solution of the classical equations of motion in the presence of **crossed electric and magnetic fields**,  $E \perp B$ . In this case one has to solve

$$\dot{\boldsymbol{r}} = \boldsymbol{v}_{\boldsymbol{k}} \quad \hbar \dot{\boldsymbol{k}} = -e\boldsymbol{E} - \frac{e}{c}\boldsymbol{v} \times \boldsymbol{B}$$

As above, Eq. (4.16) we multiply by the unit vector  $\hat{B}$ 

$$\hat{m{B}} imes \hbar \dot{m{k}} = -rac{eB}{c} \dot{m{r}}_{\perp} - e \hat{m{B}} imes m{E}$$

and obtain

$$\mathbf{r}_{\perp}(t) - \mathbf{r}_{\perp}(0) = -\frac{\hbar c}{eB} \mathbf{B} \times (\mathbf{k}(t) - \mathbf{k}(0) + \underbrace{\frac{cE}{B}(\hat{\mathbf{E}} \times \hat{\mathbf{B}})t}_{\text{drift} \perp \text{ to} \mathbf{E} \text{ and } \mathbf{B}}$$

Figure 4.14: In crossed electric and magnetic fields, an oscillatory motion is combined with a drift motion perpendicular to the electric and magnetic fields.

Thus the motion of electrons in real space again tracks the motion of in momentum space but now an extra drift term  $\frac{cE}{B}(\hat{\boldsymbol{E}} \times \hat{\boldsymbol{B}})t$  has been added. One therefore obtains a superposition of a cyclotron motion and a drift motion perpendicular to the electric field and thus parallel to equipotential lines.

Sometimes it is useful to introduce a modified energy,  $\tilde{E}_{k} = \varepsilon_{k} - \hbar k \cdot w$  which allows to write

$$\hbar \dot{\boldsymbol{k}} = -\frac{e}{\hbar c} \frac{\mathrm{d}E_{\boldsymbol{k}}}{\mathrm{d}\boldsymbol{k}} \times \boldsymbol{B} \text{ as } -\frac{e}{\hbar c} \boldsymbol{w} \times \boldsymbol{B} = e\boldsymbol{E} \text{ for } \boldsymbol{E} \perp \boldsymbol{B}.$$

# TRANSPORT THEORY

# 5.1 Drude Model

Drude developed his model for electron transport more than a century ago around 1900, at a time when even the most basic insight on the physics of electrons was not available. Nevertheless, his phenomenological approach is used until today to fit experiments and we will use it also as a reference for more refined theories later.

We start by defining the conductivity using the relation

$$\boldsymbol{j} = \sigma \boldsymbol{E}$$

where j is the current density defined as current per area. Similarly, the resistivity is defined by

 $E = \rho j$ 

with  $\rho = \sigma^{-1}$ . In general  $\sigma$  and  $\rho$  are  $3 \times 3$  matrices which we will ignore in this section. The resistance of a sample with area A and length l can be obtained from  $R = \frac{U}{I} = \frac{El}{jA} = \rho \frac{l}{A}$ .



Figure 5.1: Characterizing resistivity of a wire.

A more microscopic formula for the current density is

$$j = nev$$

where *n* is the density of electrons, *e* the electron charge (negative), and *v* average "drift velocity" defined by the equation above. The velocity of electrons is determined by two factors, 1) the acceleration by the electric field,  $\dot{\boldsymbol{p}} = e\boldsymbol{E}$  and collissions. Here we assume that after a

collission the velocity of the particle is set back to zero. If we denote the typical time between collissions by  $\tau$ , we obtain the following estimate of the average velocity (a more precise version of this is given below)

$$\boldsymbol{v} = \left\langle \frac{\boldsymbol{p}}{m} \right\rangle \approx \frac{e\boldsymbol{E}}{m} \boldsymbol{\tau}$$

Using this velocity, we obtain for the dc conductivity

$$\sigma = \frac{ne^2\tau}{m}, \quad \frac{1}{\tau} = \text{collision rate}$$
(5.1)

Let us do this a bit more careful by assuming that collisions lead to an exponential decay of the momentum on a time scale set by  $\tau$  at least if we average over many electrons and collission events. Thus we assume

$$\dot{\boldsymbol{p}} = e\boldsymbol{E}(t) - \frac{\boldsymbol{p}}{\tau}$$

A Fourier transformation gives  $-i\omega \mathbf{p}(\omega) = e\mathbf{E}(\omega) - \frac{\mathbf{p}(\omega)}{\tau}$  which is solved by  $\mathbf{p}(\omega) = \frac{e\mathbf{E}(\omega)}{\frac{1}{\tau} - i\omega}$ . Defining the frequency dependent conductivity (usually called optical conductivity) using  $\mathbf{j}(\omega) = \sigma(\omega)\mathbf{E}(\omega)$  we obtain the Drude conductivity

$$\sigma(\omega) = \frac{ne^2}{m} \frac{1}{\frac{1}{\tau} - i\omega} = \frac{\sigma_0}{\frac{1}{\tau} - i\omega}, \quad \sigma_0 = \frac{ne^2\tau}{m}$$
(5.2)

The real part is given by

$$\operatorname{Re}(\sigma(\omega)) = \frac{\sigma_0}{1 + (\omega\tau)^2}$$

It is a sharp peak of width  $1/\tau$  with a height linear in  $\tau$  observed in many experiments.



Figure 5.2: Sketch of a Drude peak in the optical conductivity as observed in many metals. The width is set by the scattering rate,  $1/\tau$ .

Experimentally  $\sigma(\omega)$  is determined from measurements of the reflection/transmission of electromagnetic waves (microwaves, infrared, visible, ...). Focussing on the visible range, the Drude conductivity can also be used to understand why and under what condition metals look "metallic". To discuss this, we start from Maxwell's equation

#### 5.2. Boltzmann Equation

$$oldsymbol{
abla} oldsymbol{
abla} oldsymbol{H} imes oldsymbol{H} = rac{1}{c}rac{\partial oldsymbol{E}}{\partial t} + rac{4\pi}{c}oldsymbol{j}$$

Using a Fourier transformation and  $\boldsymbol{j}(\omega) = \sigma(\omega)\boldsymbol{E}(\omega)$  we obtain

$$\nabla \times \boldsymbol{H} = \frac{1}{c}(-i\omega)E(\omega)\left(1 + \frac{4\pi\sigma(\omega)}{-i\omega}\right)\frac{1}{c} = (-i\omega)\epsilon(\omega)E(\omega)$$

with the frequency-dependent dielectric constant

$$\epsilon(\omega) = 1 + \frac{4\pi i \sigma(\omega)}{\omega}$$

Assuming that  $\omega \tau > 1$  for optical frequencies, we can approximate  $\sigma(\omega) = \frac{ne^2}{m} \frac{1}{-i\omega}$  which leads to

$$\epsilon(\omega) \underset{\omega \tau \gg 1}{\approx} 1 - \frac{\omega_{\rm p}^2}{\omega^2}$$

where the so-called "plasma frequency"  $\omega_p$  is defined by  $\omega_p^2 = \frac{4\pi n e^2}{m}$ . Importantly, for  $\omega < \omega_p$  the dielectric constant is negative which leads to total reflection: the metal acts like a mirror (mirror), while the formula (which ignored absorption as we approximated the conductivity by a purely imaginary function) predicts a positive dielectric function and therefore transmission for larger frequencies.

The Drude model has turned out to be highly successful but it does not know anything about the Pauli principle which governs the energetics of electrons in a solid. It is thus far from clear why this can be a reasonable approximation. For example, we know that electrons in a solid at the Fermi energy move with about a 1% of the speed of light, while drift velocities even for large current densities in the range of only mm per second, many orders of magnitude smaller. Therefore we will need a more microscopic theory of transport.

## 5.2 Boltzmann Equation

We would like to derive a theory from transport valid on length scales which are large compared to the wave length of electrons. Therefore a good starting point of theoretical description is the discussion of Chapter 4, where we derived the semiclassical dynamics of wave packages centered at  $\boldsymbol{r}$  with crystalline momentum  $\boldsymbol{k}$ . Now we want to combine this with a more "statistical" approach, by keeping track of the average number of electrons in band n with momentum  $\boldsymbol{k}$  at a certain postition  $\boldsymbol{r}$ .

$$f_{n\boldsymbol{k}}(\boldsymbol{r},t)$$

In equilibrium this function is space and time-independent

$$f_{n\boldsymbol{k}}^{eq}(\boldsymbol{r},t) = \frac{1}{e^{(\varepsilon_{n\boldsymbol{k}}-\mu)/k_{\rm B}T}+1}$$
(5.3)

and given by the Fermi function. Our goal is, however, to obtain a theory out of equilibrium, e.g., due to the presence of external forces. We are also seeking a description which is compatible

with the semiclassical equations of motion. Our goal it to calculate  $\partial_t f_{nk}(\mathbf{r}, t)$ . For this, we have to consider several contributions.



**Figure 5.3:** Due to the velocity of particles, their distribution function changes giving rise to  $\frac{\partial f}{\partial t}\Big|_{\text{flow}}$ .

The first arises from the flow of particles

$$\dot{m{r}}=rac{1}{\hbar}rac{\partialarepsilon}{\partialm{k}}-\hbar\dot{m{k}} imesm{\Omega}_{m{k}}= ilde{m{v}}_{m{k}}$$

Note that the velocity includes the "anomalous velocity". The velocity changes the distribution function according to  $f_{\mathbf{k}}(\mathbf{r}, t + \Delta t) \simeq f_{\mathbf{k}}(\mathbf{r} - \mathbf{v}\Delta t, t)$  or, equivalently,  $f(\mathbf{r}, t) \approx f(\mathbf{r} - \mathbf{v}\Delta t)$ , which gives rise to

$$\left.\frac{\partial f}{\partial t}\right|_{\text{flow}} = -\tilde{\boldsymbol{v}}_{\boldsymbol{k}} \frac{\partial}{\partial \boldsymbol{r}} f_{n\boldsymbol{k}}(\boldsymbol{r},t)$$

Similarly, external forces change the momentum,

$$\hbar \dot{k} = F_k(r)$$

with  $F = e(E + v \times B)$  which leads to  $f_k(r, t + \Delta t) = f_{k-\Delta t \dot{k}}(r, t)$  or

$$\left. \frac{\partial f}{\partial t} \right|_{\text{force}} = -\frac{\boldsymbol{F}_{\boldsymbol{k}}(\boldsymbol{r})}{\hbar} \frac{\partial}{\partial \boldsymbol{k}} f_{n\boldsymbol{k}}(\boldsymbol{r},t)$$

where the force will in general depend on both position and momentum. Finally, collisions with defect, other electrons, phonons have to be taken into account. They are discussed in the next section. For now we can collect our results using  $\frac{\partial f}{\partial t} = \frac{\partial f}{\partial t} \Big|_{\text{flow}} + \frac{\partial f}{\partial t} \Big|_{\text{force}} + \frac{\partial f}{\partial t} \Big|_{\text{collisions}}$ .

This leads to the  ${\bf Boltzmann}$  equation

$$\frac{\partial f_{\boldsymbol{k}}(\boldsymbol{r},t)}{\partial t} + \tilde{\boldsymbol{v}}_{\boldsymbol{k}} \cdot \frac{\partial f_{n\boldsymbol{k}}(\boldsymbol{r},t)}{\partial \boldsymbol{r}} + \frac{\boldsymbol{F}_{\boldsymbol{k}}(\boldsymbol{r},t)}{\hbar} \cdot \frac{\partial f_{n\boldsymbol{k}}(\boldsymbol{r},t)}{\partial \boldsymbol{k}} = \left. \frac{\partial f}{\partial t} \right|_{\text{collisions}}$$
(5.4)

## 5.3 Collision terms

Missing from Eq. (5.4) are a term which describes the effect of collision. We first discuss the role of defects arising from misplaced atoms in the solid.

After averaging over the position of these defects, we assume that we can describe the scattering probabilistically. We assume that the transition from  $\mathbf{k} \to \mathbf{k}'$  occurs with the rate  $W_{\mathbf{k}\mathbf{k}'}$ 

(calculated below). Importantly, such a transition will only occur if the state k is occupied. Thus it will be proportional to  $f_k$  (we omit band indices here for simplicity). Furthermore, the Pauli principle demands that the state k' is empty. This is taken into account by a factor  $(1 - f_{k'})$ . This results in the following collision term due to disorder

$$\frac{\partial f_{\boldsymbol{k}}}{\partial t}\Big|_{\text{coll, disorder}} = -\int W_{\boldsymbol{k}\boldsymbol{k}'}f_{\boldsymbol{k}}(1-f_{\boldsymbol{k}'}) - \underbrace{W_{\boldsymbol{k}'\boldsymbol{k}}f_{\boldsymbol{k}'}(1-f_{\boldsymbol{k}})}_{\text{scatter from }\boldsymbol{k} \text{ to } \boldsymbol{k}'} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \tag{5.5}$$

Here we took into account both the scattering from  $\mathbf{k} \to \mathbf{k}'$  (first term) which reduces  $f_{\mathbf{k}}$  and the opposite process which increases  $f_{\mathbf{k}}$ . As scattering from impurities is energy-conserving, the transition rates obey this property

$$W_{kk'} = \delta(\varepsilon_k - \varepsilon_{k'}) \hat{W}_{kk'}$$

Note that the collision term is local in space describing collisions occuring on a length scale of order of wave length of electrons. Below we will argue that the nonlinear terms proportional to  $f_{\mathbf{k}}f_{\mathbf{k}'}$  cancel with each other (Pauli blocking is only of relevance in interacting systems).



Figure 5.4: *e-e* interaction with two incoming and to outgoing electrons.

Next, we investigate the collision term arising from electron-electron scattering. In this case there are two incoming electrons with quantum numbers 1 and 2 and two outgoing ones with quantum number 3 and 4. We denote the corresponding scattering rate

$$W_{34,12} = \sum_{\boldsymbol{G}} \tilde{W}_{34,12}^{\boldsymbol{G}} \delta(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3 - \boldsymbol{k}_4 - \boldsymbol{G}) \delta\left(\varepsilon_{\boldsymbol{k}_1} + \varepsilon_{\boldsymbol{k}_2} - \varepsilon_{\boldsymbol{k}_3} - \varepsilon_{\boldsymbol{k}_4}\right)$$
(5.6)

where we again suppress band indices. Electron-electron scattering conserves energy. Momentum is only conserved up to reciprocal lattice vectors G. Scattering can only occur, when the two incoming states are occupied, the outging states empty which is taken into account in the following formula

$$\frac{\partial f}{\partial t}\Big|_{\text{coll, e-e scattering}} = -\left(\sum_{\boldsymbol{k}_{2},\boldsymbol{k}_{3},\boldsymbol{k}_{4}} W_{\boldsymbol{k}_{3}\boldsymbol{k}_{4},\boldsymbol{k}\boldsymbol{k}_{2}} f_{\boldsymbol{k}} f_{\boldsymbol{k}_{2}} \left(1 - f_{\boldsymbol{k}_{3}}\right) \left(1 - f_{\boldsymbol{k}_{4}}\right) - \sum_{\boldsymbol{k}_{1},\boldsymbol{k}_{2},\boldsymbol{k}_{3}} W_{\boldsymbol{k}_{3}\boldsymbol{k},\boldsymbol{k}_{1}\boldsymbol{k}_{2}} f_{\boldsymbol{k}_{1}} f_{\boldsymbol{k}_{2}} \left(1 - f_{\boldsymbol{k}_{3}}\right) \left(1 - f_{\boldsymbol{k}}\right)\right) \qquad (5.7)$$

Note that we have not given a precise derivation of the form of the collision terms. A precise derivation (using, for example, techniques from non-equilibrium quantum field theory) is beyond the scope of this lecture. Here we would like to emphasize that even without a derivation the structure of the collision term with factors  $f_k$  for incoming occupied states and  $1 - f_k$  for outgoing states which have to be empty before the collision is highly plausible.

Similar formulas also apply to the scattering of electrons from phonons. Introducing the phonon distribution function  $n_k$ .



Figure 5.5: *e*-phonon interaction.

In this case the scattering term takes the form

$$\begin{split} \left. \frac{\partial f}{\partial t} \right|_{\text{coll, e-ph scattering}} &== -\sum_{\mathbf{k}',\mathbf{k}''} \left( W^{A}_{\mathbf{k}'\mathbf{k},\mathbf{k}''} f_{\mathbf{k}} \left(1 - f_{\mathbf{k}'}\right) n_{\mathbf{k}''} - W^{A}_{\mathbf{k}\mathbf{k}',\mathbf{k}''} f_{\mathbf{k}'} \left(1 - f_{\mathbf{k}}\right) n_{\mathbf{k}''} \right. \\ &+ W^{E}_{\mathbf{k}'\mathbf{k},\mathbf{k}''} f_{\mathbf{k}} \left(1 - f_{\mathbf{k}'}\right) \left(1 + n_{\mathbf{k}''}\right) - W^{E}_{\mathbf{k}\mathbf{k}',\mathbf{k}''} f_{\mathbf{k}'} \left(1 - f_{\mathbf{k}}\right) \left(1 + n_{\mathbf{k}''}\right) \right) \end{split}$$

where the first two terms describe the absorption of a phonon with momentum k'' while the last two term describe a phonon-emission process.

Above, we assumed that the transition rates are known. They can also be calculated by solving the corresponding scattering problem. Often one simply uses perturbation theory. Here we can apply directly Fermi's golden rule which states that for a weak perturbation  $\Delta H$  the transition rate from an initial to a final state is given by

$$W_{\boldsymbol{k}'\boldsymbol{k}} = \frac{2\pi}{\hbar} |\langle \boldsymbol{k}' | \Delta H | \boldsymbol{k} \rangle|^2 \delta(\varepsilon_{\boldsymbol{k}} - \varepsilon_{\boldsymbol{k}'})$$

Let us, for example, consider the scattering from defects located at random positions  $\mathbf{R}_i$  described by the potential term  $\Delta H = \sum_{\mathbf{R}_i} V_{imp}(\mathbf{r} - \mathbf{R}_i)$ . In this case

$$\langle \boldsymbol{k}' | V_{\mathrm{imp}}(\boldsymbol{r} - \boldsymbol{R}_i) | \boldsymbol{k} 
angle = \langle \boldsymbol{k}' | V_{\mathrm{imp}}(\boldsymbol{r}) | \boldsymbol{k} 
angle e^{i \boldsymbol{R}_i (\boldsymbol{k} - \boldsymbol{k}')}$$

with

$$\langle m{k}' | V_{
m imp}(m{r}) | m{k} 
angle = \int \psi^*_{nm{k}'}(m{r}) V_{
m imp}(m{r}) \psi^*_{nm{k}}(m{r}) \; \mathrm{d}m{r}$$

where we use a normalization condition where  $\int_{\text{cell}} \psi_{nk}^*(\mathbf{r})\psi(\mathbf{r}) \, \mathrm{d}\mathbf{r}$  gives the volume of the unit cell.

After averaging over the positions in

$$\langle \boldsymbol{k}' | \Delta H | \boldsymbol{k} \rangle |^{2} = \sum_{\boldsymbol{R}_{i} \boldsymbol{R}_{j}} \langle \boldsymbol{k} | V_{\text{imp}}(\boldsymbol{r} - \boldsymbol{R}_{j}) | \boldsymbol{k}' \rangle \langle \boldsymbol{k}' | V_{\text{imp}}(\boldsymbol{r} - \boldsymbol{R}_{i}) | \boldsymbol{k} \rangle$$
$$= \sum_{\boldsymbol{R}_{i} \boldsymbol{R}_{j}} e^{i(\boldsymbol{R}_{i} - \boldsymbol{R}_{j})(\boldsymbol{k} - \boldsymbol{k}')} \langle \boldsymbol{k} | V_{\text{imp}}(\boldsymbol{r}) | \boldsymbol{k}' \rangle \langle \boldsymbol{k}' | V_{\text{imp}}(\boldsymbol{r}) | \boldsymbol{k} \rangle$$
(5.8)

only terms with i = j survive and one finds

#### 5.4. Linearized Boltzmann equation and relaxation time approximation

$$W_{\boldsymbol{k}'\boldsymbol{k}} = \frac{2\pi}{\hbar} n_i |\langle \boldsymbol{k}' | V_{\rm imp} | \boldsymbol{k} \rangle|^2 \delta(\varepsilon_{\boldsymbol{k}} - \varepsilon_{\boldsymbol{k}'})$$

with  $n_i$  is the density of impurities per volume. A similar equation can also be derived for the electron-electron interactions, where one should use the matrix elements of the *screened* Coulomb interaction, a concept discussed later. Most often one uses the matrix elements as phenomenological parameters, ignoring, e.g., their k dependence.

A very important property of the scattering rate above turns out to be valid even beyond perturbation theory

$$W_{kk'} = W_{k'k}$$
 and similarly  $W_{34,12} = W_{12,34}$  (5.9)

It follows from the unitarity of time evolution. Using this equation one finds for example that

$$W_{kk'}f_{k}(1-f_{k'}) - W_{k'k}f_{k'}(1-f_{k}) = W_{kk'}(f_{k}-f_{k'})$$

as mentioned above.

Combining (5.9) with the energy conservation encoded in Eq. (5.6) one can show (see Eq. 5.10 below) that the collision term vanishes if and only if when one inserts a Fermi function, Eq. (5.3). An important consequence is that electron-electron collisions actually drive electrons towards a thermal state described by Fermi function. One can even view the collision term as a way to *derive* the Fermi function, which naturally arises as the steady state of a Boltzmann equation.

# 5.4 Linearized Boltzmann equation and relaxation time approximation

## 5.4.1 Perturbative solution of the Boltzmann equation

As discussed above, one of the most important properties of the collission terms in the Boltzmann equation is that they vanish for the Fermi function

$$f_{\boldsymbol{k}} = f_{\varepsilon_{\boldsymbol{k}}}^{0} = \frac{1}{\exp\left(\frac{\varepsilon_{\boldsymbol{k}} - \mu}{k_{\mathrm{B}}T}\right) + 1}$$

Indeed, one can check that

$$\left(f^{0}_{\varepsilon_{\boldsymbol{k}}}f^{0}_{\varepsilon_{\boldsymbol{k}'}}\left(1-f^{0}_{\varepsilon_{\boldsymbol{k}''}}\right)\left(1-f^{0}_{\varepsilon_{\boldsymbol{k}'''}}\right)-f^{0}_{\varepsilon_{\boldsymbol{k}''}}f^{0}_{\varepsilon_{\boldsymbol{k}'''}}\left(1-f^{0}_{\varepsilon_{\boldsymbol{k}}}\right)\left(1-f^{0}_{\varepsilon_{\boldsymbol{k}'}}\right)\right)\delta\left(\varepsilon_{\boldsymbol{k}}+\varepsilon_{\boldsymbol{k}'}-\varepsilon_{\boldsymbol{k}''}-\varepsilon_{\boldsymbol{k}'''}\right)=0$$
(5.10)

Thus the Boltzmann equation describes the relaxation towards thermal equilibrium. If the system is now perturbed by weak external fields, e.g., a weak electric field, one can use perturbation theory using the ansatz

$$f_{\boldsymbol{k}} \approx f_{\varepsilon_{\boldsymbol{k}}}^0 + \delta f_{\boldsymbol{k}} \tag{5.11}$$

linearizing the equation in the external field under the assumption that  $\delta f_k$  is linear in E. This is the strategy we will follow below.

This strategy fails when the system is inhomogeneous on large scales with large variations of density or temperature. If locally, the system can still be approximated by an equilibrium state, one can describe it by **hydrodynamic equations**, as, for example, the **Navier-Stokes equation**. We will not follow this route here but simply point out the necessary strategy: in this case one linearizes the Boltzmann equation around a *local* equilibrium, which depends on space and time

$$f_{\mathbf{k}}^{\text{local}}(\mathbf{r},t) = \frac{1}{\exp\left(\frac{\varepsilon_{\mathbf{k}} - \mu(\mathbf{r},t) - \mathbf{k} \cdot \mathbf{u}(\mathbf{r},t)}{k_{\text{B}}T(\mathbf{r},t)}\right) + 1}, \quad f_{\mathbf{k}} \approx f_{\mathbf{k}}^{\text{local}}(\mathbf{r},t) + \delta f_{\mathbf{k}}(\mathbf{r},t).$$
(5.12)

This is justified because for such a local equilibrium, the collision term vanishes. As collisions dominate in the hydrodynamic regime, one can use this as the start of the perturbative expansion. Next one determines  $\mu(\mathbf{r},t), T(\mathbf{r},t)$  and velocity  $\mathbf{u}(\mathbf{r},t)$  from the condition that the local density, the local energy and the local momentum are described by  $f_{\mathbf{k}}^{\text{local}}(\mathbf{r},t)$ , e.g.,  $n(\mathbf{r},t) = \sum_{k} f_{\mathbf{k}}^{\text{local}}(\mathbf{r},t)$  or, equivalently,  $\sum_{k} \delta f_{\mathbf{k}}(\mathbf{r},t) = 0$ . Here a finite  $\mathbf{u}(\mathbf{r},t)$  is only used systems with (approximate) momentum conservation, like a liquid. With this approach one can obtain equations for  $\mu(\mathbf{r},t), T(\mathbf{r},t)$  and  $\mathbf{u}(\mathbf{r},t)$  which are then translated to equations for particle, energy and momentum density.

We will, however, study a simpler case where we Taylor expand around a homogeneous solution,  $T(\mathbf{r},t) = T, \mu(\mathbf{r},t) = \mu = \text{const}, \mathbf{u} = \mathbf{0}$ . For example, we will consider the case of a small constant electric field. In this case we approximate

$$e\boldsymbol{E}\cdot\frac{\partial}{\partial\boldsymbol{k}}f_{\boldsymbol{k}}\approx e\boldsymbol{E}\cdot\frac{\partial}{\partial\boldsymbol{k}}f^{0}_{\varepsilon_{\boldsymbol{k}}}=e\boldsymbol{E}\cdot\boldsymbol{v}_{\boldsymbol{k}}\frac{\partial f^{0}}{\partial\varepsilon_{\boldsymbol{k}}}$$

and linearize the rest of the Boltzmann equation in  $\delta f_k$ . For the stationary state,  $\partial_t \delta f_k = 0$ , we obtain the the **linearized Boltzmann equation** 

$$e\boldsymbol{E} \cdot v_{\boldsymbol{k}} \frac{\partial}{\partial \varepsilon_{\boldsymbol{k}}} f^{0}_{\varepsilon_{\boldsymbol{k}}} = \frac{\partial f}{\partial t} \bigg|_{\text{coll}} = -\underbrace{\int M_{\boldsymbol{k}\boldsymbol{k}'} \delta f_{\boldsymbol{k}'}}_{\text{linearized collision term}} \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^{3}}, \qquad (5.13)$$

which is a linear integral equation. Formally (but often not in practice, see next section) it can be solved by matrix inversion

$$\delta f_{\boldsymbol{k}} = (M^{-1})_{\boldsymbol{k}\boldsymbol{k}'} e^{\boldsymbol{E}} \cdot \boldsymbol{v}_{\boldsymbol{k}'} \frac{\partial f^0_{\varepsilon_{\boldsymbol{k}'}}}{\partial \varepsilon_{\boldsymbol{k}'}}.$$
(5.14)

## 5.4.2 Scattering kernel

Let us consider, for example, impurity scattering, where  $\left. \frac{\partial f}{\partial t} \right|_{\text{coll}} = -\int W_{\boldsymbol{k}\boldsymbol{k}'} \left( f_{\boldsymbol{k}} - f_{\boldsymbol{k}'} \right) \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$ . In this case we obtain for the scattering kernel

$$M_{\boldsymbol{k}\boldsymbol{k}'} = \delta(\boldsymbol{k} - \boldsymbol{k}') \int W_{\boldsymbol{k}\boldsymbol{k}''} \frac{\mathrm{d}\boldsymbol{k}''}{(2\pi)^3} - W_{\boldsymbol{k}\boldsymbol{k}'}$$

Next, we determine  $M_{kk'}$  for electron electron interactions. It turns out to be useful (both for analytical and numerical calculatoins) to use the parameterization

$$\delta f_{\boldsymbol{k}} = \frac{\partial f^0}{\partial \varepsilon_{\boldsymbol{k}}} \phi_{\boldsymbol{k}}$$

#### 5.4. Linearized Boltzmann equation and relaxation time approximation

with

$$\frac{\partial f^0}{\partial \varepsilon_{\mathbf{k}}} = -\beta f^0 (1 - f^0)$$

Using this formula, one can show (using Eq. (5.10) twice) that to linear order in  $\phi_k$ 

$$-W_{kk_{2},k_{3}k_{4}}\left(f_{k}f_{k_{2}}\left(1-f_{k_{3}}\right)\left(1-f_{k_{4}}\right)-f_{k_{3}}f_{k_{4}}\left(1-f_{k}\right)\left(1-f_{k_{2}}\right)\right) \\ \approx \underbrace{W_{kk_{2},k_{3}k_{4}}f_{\varepsilon_{k}}^{0}f_{\varepsilon_{k_{2}}}^{0}\left(1-f_{\varepsilon_{k_{3}}}^{0}\right)\left(1-f_{\varepsilon_{k_{4}}}^{0}\right)\beta}_{\tilde{W}_{kk_{2},k_{3}k_{4}}}\left(\phi_{k}+\phi_{k_{1}}-\phi_{k_{3}}-\phi_{k_{4}}\right)$$

Therefore the scattering matrix can be written as the sum of four terms

$$M_{\boldsymbol{k}\boldsymbol{k}'} = \left(\delta(\boldsymbol{k}-\boldsymbol{k}')\sum_{\boldsymbol{k}_{2},\boldsymbol{k}_{3},\boldsymbol{k}_{4}}\tilde{W}_{\boldsymbol{k}\boldsymbol{k}_{2},\boldsymbol{k}_{3}\boldsymbol{k}_{4}} + \sum_{\boldsymbol{k}_{3},\boldsymbol{k}_{4}}\tilde{W}_{\boldsymbol{k}\boldsymbol{k}',\boldsymbol{k}_{3}\boldsymbol{k}_{4}} - \sum_{\boldsymbol{k}_{2},\boldsymbol{k}_{4}}\tilde{W}_{\boldsymbol{k}\boldsymbol{k}_{2},\boldsymbol{k}'\boldsymbol{k}_{4}} - \sum_{\boldsymbol{k}_{2},\boldsymbol{k}_{3}}\tilde{W}_{\boldsymbol{k}\boldsymbol{k}_{2},\boldsymbol{k}_{3}\boldsymbol{k}'}\right) \left(\frac{\partial f_{\varepsilon_{\boldsymbol{k}'}}^{0}}{\partial\varepsilon_{\boldsymbol{k}'}}\right)^{-1}.$$

$$(5.15)$$

## 5.4.3 Conservation laws

Before we proceed to calculate physical quantities like the conductivity in the next chapter, we will now investigate the consequence of **conservation laws** for the linearized Boltzmann equation. First, we consider particle number conservation  $\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k} f_{k} = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{k} \delta f_{k} = 0$ . This implies that

$$\int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \int \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} \ M_{\boldsymbol{k}\boldsymbol{k}'} \delta f_{\boldsymbol{k}'} = 0 \qquad \forall \ \delta f$$

Therefore,

$$\int M_{\boldsymbol{k}\boldsymbol{k}'} \; \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} = 0 \qquad \forall \; \boldsymbol{k}'$$

In other words, the vector (1, 1, 1, 1, 1, ...) is a left-eigenvector of  $M_{kk'}$  with eigenvalue 0. Besides the physical importance, the fact that there are vanishing eigenvalues is also important for the matrix inversion needed for Eq. (5.14).

Similarly, energy conservation:  $\frac{\mathrm{d}}{\mathrm{d}t} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \delta f_{\mathbf{k}} = 0$  implies that  $\int \varepsilon_{\mathbf{k}} M_{\mathbf{k}\mathbf{k}'} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3} = 0.$ 

To show that this follows from Eq. (5.15) one has to use that  $W_{12,34} \propto \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4)$ .

Momentum conservation is usually *not* valid for a solid due to Umklapp scattering processes where  $\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4 = \mathbf{G}_0$ . For a small Fermi surface, however, such processes are absent close to the Fermi energy. If one ignores Umklapp processes, implying that  $W_{12,34} \propto \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4)$ , then

$$\int \boldsymbol{k} M_{\boldsymbol{k}\boldsymbol{k}'} \; \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} = 0$$

and one obtains another eigenvector with eigenvalue zero. If there is a little bit of Umklapp scattering, one obtains a very small eigenvalue.

### 5.4.4 Relaxation time approximation

An important approximation, which we will also use in the following is the relaxation time approximation, where one approximates

$$\left.\frac{\partial f_{\boldsymbol{k}}}{\partial t}\right|_{\rm coll}\approx-\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}}$$

where  $\tau_{\mathbf{k}} \geq 0$  is called a relaxation time, while  $\frac{1}{\tau_{\mathbf{k}}}$  is the so-called relaxation rate. Often the  $\mathbf{k}$  dependence can be (or simply is) ignored. In this case one replaces  $\tau_{\mathbf{k}}$  by the constant  $\tau$ . Note the minus sign above which arises because the system relaxes towards equilibrium in the absence of external forces,

$$\delta f_{\mathbf{k}}(t) \propto e^{-t/\tau_{\mathbf{k}}}.$$

Within the relaxation-time approximation the solution of the Boltzmann equation is trivial. From

$$e\boldsymbol{E}\cdot\boldsymbol{v}_{\boldsymbol{k}}\frac{\partial}{\partial\varepsilon_{\boldsymbol{k}}}f^{0}_{\varepsilon_{\boldsymbol{k}}}=-\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}}$$

one obtains

$$\delta f_{\boldsymbol{k}} = -e\boldsymbol{E}\cdot\boldsymbol{v}_{\boldsymbol{k}}\frac{\partial}{\partial\varepsilon_{\boldsymbol{k}}}f^0_{\varepsilon_{\boldsymbol{k}}}\cdot\tau_{\boldsymbol{k}}$$

What is  $\frac{1}{\tau_k}$ ? Ideally it should obey the equation

$$-\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}} = -\int M_{\boldsymbol{k}\boldsymbol{k}'} \delta f_{\boldsymbol{k}'} \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$
(5.16)

This implies that the best approximation for  $\frac{1}{\tau_k}$  depends on the form of  $\delta f_k$ . Therfore for different physical properties, one should use different formulas for the relaxation time. Therefore a "single-particle relaxation time" (defined below) differes from a "transport relaxation time" (next section), and transport relaxation times are different when one considers charge or energy transport.

Before we continue our analysis of the relaxation time, it is important to stress that the relaxation time approximation (at least in its most commonly used form which we also used above) notoriously **violates conservation laws** as, e.g.,  $\sum_{k} \frac{1}{\tau_{k}} \neq 0$ . This can, however, sometimes be ignored if one makes sure that  $\sum_{k} \delta f_{k} = 0$ .

Let us now discuss the single particle relaxation rate, which is used to describe how a particle added to the system at a certain momentum  $k_0$  decays. Therefore one uses

$$\delta f_{\boldsymbol{k}} = \alpha \delta(\boldsymbol{k} - \boldsymbol{k}_0) \tag{5.17}$$

which is used in Eq. (5.16)

$$\frac{\alpha\delta(\boldsymbol{k}-\boldsymbol{k}_0)}{\tau_{\boldsymbol{k}}} = \int M_{\boldsymbol{k}\boldsymbol{k}'}\alpha\delta(\boldsymbol{k}'-\boldsymbol{k}_0) \; \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$

We evaluate this for impurity scattering, where we argued that  $M_{kk'} = \delta(k-k') \int W_{kk''} \frac{dk''}{(2\pi)^3} - W_{kk'}$ .

Close to  $k_0$  only the  $\delta$ -function contributes and we find

$$\frac{1}{\tau_{\boldsymbol{k}}^{\text{imp}}} = \int W_{\boldsymbol{k}\boldsymbol{k}^{\prime\prime}} \, \frac{\mathrm{d}\boldsymbol{k}^{\prime\prime}}{(2\pi)^3}$$

Due to the integral over k'', usually most of the momentum dependence is gone and one can simply approximate

$$\frac{1}{\tau_{\boldsymbol{k}}^{\text{imp}}} \approx \frac{1}{\tau^{\text{imp}}} = \int W_{\boldsymbol{k}_F \boldsymbol{k}^{\prime\prime}} \frac{\mathrm{d}\boldsymbol{k}^{\prime\prime}}{(2\pi)^3}$$

using one point  $k_F$  on the Fermi surface.

The same type of argument can also be used for electron-electron interactions where only the  $\delta$ -function contribution in Eq. (5.15) survives. Thus one obtains as a single-particle relaxation rate

$$\frac{1}{\tau_{k}^{\text{int}}} = \iiint W_{kk_{2},k_{3}k_{4}} f^{0}_{\varepsilon_{k_{2}}} \left(1 - f^{0}_{\varepsilon_{k_{3}}}\right) \left(1 - f^{0}_{\varepsilon_{k_{4}}}\right) \frac{\mathrm{d}k_{2}}{(2\pi)^{3}} \frac{\mathrm{d}k_{3}}{(2\pi)^{3}} \frac{\mathrm{d}k_{4}}{(2\pi)^{3}}.$$
(5.18)

Here, we only consider terms where an electron starting at momentum k is scattered somewhere else. Due to our ansatz (5.17) we do not take into account the reverse processes.



Figure 5.6: Scattering due to electron-electron interactions in presence of a Fermi sea.

The single particle relaxation time due to interactions depends strongly on the energy of the incoming particle due to the Fermi functions and due to energy conservation

$$W \propto \delta \left( \varepsilon_{\boldsymbol{k}_1} + \varepsilon_{\boldsymbol{k}_2} - \varepsilon_{\boldsymbol{k}_3} - \varepsilon_{\boldsymbol{k}_4} \right) \delta(\boldsymbol{k}_1 + \boldsymbol{k}_2 - \boldsymbol{k}_3 - \boldsymbol{k}_4 + \boldsymbol{G}_n)$$

which in combination with the Fermi functions  $f^0$  strongly restricts the possible scattering processes. Let us consider the case of zero temperature,

T = 0 where

$$f_{\epsilon}^{0} = \begin{cases} 1 & \varepsilon < \mu \\ 0 & \varepsilon > \mu \end{cases}$$

We consider the situation depicted in Fig. (5.6) studying the relaxation of a particle initially at energy  $\varepsilon_{\mathbf{k}} > \mu$ . It scatters from a second electron with momentum  $\mathbf{k}_2$  to states with momenta  $\mathbf{k}_3$  and  $\mathbf{k}_4$ . It follows from the Pauli principle – or simply from the Fermi functions in Eq. (5.18) which encodes this physics – that

$$\varepsilon_{k_2} < \mu, \qquad \varepsilon_{k_3}, \varepsilon_{k_4} > \mu$$

At the same time, the energy conservation encoded in W enforces that

$$\underbrace{\varepsilon_{\boldsymbol{k}}}_{>\mu} + \underbrace{\varepsilon_{\boldsymbol{k}_2}}_{<\mu} = \underbrace{\varepsilon_{\boldsymbol{k}_3} + \varepsilon_{\boldsymbol{k}_4}}_{>\mu}$$

This enforces that

$$\mu < \varepsilon_{\boldsymbol{k}_3} < \varepsilon_{\boldsymbol{k}_1}$$

while

$$0 < \mu - \varepsilon_{k_2} < \varepsilon_k - \varepsilon_{k_3}$$

Our goal is to find out how the integral on the right-hand side of Eq. (5.18) depends on the energy  $\epsilon_{\mathbf{k}}$  of the incoming particle. For this we ignore for a moment the detailed  $\mathbf{k}$ -dependence in the integrant and replace the momentum integrals by energy integrals taking into account the inequalities derived above. This suggests that the scattering rate is proportional to

$$\int\limits_{\mu}^{\varepsilon_{\boldsymbol{k}}} \mathrm{d}\varepsilon_{\boldsymbol{k}_{3}} \int\limits_{\mu+\varepsilon_{\boldsymbol{k}_{3}}-\varepsilon_{\boldsymbol{k}}}^{\mu} \mathrm{d}\varepsilon_{\boldsymbol{k}_{2}} = \int\limits_{\mu}^{\varepsilon_{\boldsymbol{k}}} (\varepsilon_{\boldsymbol{k}}-\varepsilon_{\boldsymbol{k}_{3}}) \mathrm{d}\varepsilon_{\boldsymbol{k}_{3}} = \frac{1}{2} (\varepsilon_{\boldsymbol{k}}-\mu)^{2}$$

This suggests that the scattering rate drops rapidly when the energy of the particle approaches the Fermi energy,  $\frac{1}{\tau_k^{\text{int}}} \propto (\varepsilon_k - \mu)^2$ . The formula above did, however, ignore all the detailed momentum dependencies, and, most notably, the constraints arising from energy conservation. At least for a three-dimensional situation it turns out that only energy conservation is important for the single-particle relaxation rate, no extra constraints arise from momentum conservation. Thus our result holds and only the precise prefactor depends on details of the momentum dependence of W.

The analysis above was done for T = 0. At finite temperature one has to take into account that

both  $f^0$  and  $1 - f^0$  are finite in a window of width T around the Fermi energy. Repeating the analysis given above, one concludes that

$$\frac{1}{\tau_{\boldsymbol{k}}^{\text{int}}} \propto \max\left((\varepsilon_{\boldsymbol{k}} - \mu)^2, T^2\right) \tag{5.19}$$

This result tells us that at least at sufficient low temperatures and sufficiently close to the Fermi energy the relaxation rate is *much* smaller than the energy of the particles  $\epsilon_{\mathbf{k}} - \mu \sim T$  of the particles measured from the Fermi energy. Even at room temperature a typical metal with a Fermi energy of 10000 K is expected to have a scattering rate which is more than an order of magnitude smaller than the typical energy of thermal excitations. The suppression of scattering due to the Pauli principle is thus the main reason why non-interacting electrons are such a good starting point to understand the property of metals. The calculation sketched above is therefore *key to almost all of our understanding of the properties of metals*, including the validity of ab-initio approximation schemes discussed in Sec. 2.3. Also the derivation of the Boltzmann equation which we used above builds strongly on the concept that there is little scattering of the low-energy quasiparticles which is thus justified in hindsight. More refined theoretical tools (like a renormalization group analysis within quantum field theory) fully confirm the results for ordinary three-dimensional metals.

The full theory of interacting electrons weakly scattering from each other is called "**Fermi liquid theory**". Fermi liquid theory is not included in this course. Compared to the Boltzmann

equation it includes one more ingredient: the energy (and thus the velocity) of a (quasi-) particle with momentum k depends on the occupation of the other electrons

$$\epsilon_{\boldsymbol{k}} = \epsilon^0_{\boldsymbol{k}} + \sum_{\boldsymbol{k}'} F_{\boldsymbol{k}\boldsymbol{k}'} \delta f_{\boldsymbol{k}'}$$

This small addition to the Boltzmann equation makes it an asymptocially exact theory for low energies.

Only few cases are know where Fermi liquid theory does not hold. For example, close to a phase transition scattering from fluctuations can become so singluar that Fermi liquid theory breaks down. Another example are one-dimensional systems where the interplay of energy and momentum conservation leads to a strong enhancement of scattering. Luckily in this case it turns out that charge- and spin density waves remain weakly interacting and the concept of a Fermi liquid is replaced by a so-called Luttinger liquid, also a topic beyond the scope of this lecture.

## 5.5 Conductivity and transport scattering time

## 5.5.1 Conductivity in relaxation time approximation

We now come back to our original goal, the calculation of the conductivity. We will first do the calculation within a relaxation time approximation, postponing the question of how to compute the relevant scattering time. For a time dependent electric field  $\boldsymbol{E}(t) = \boldsymbol{E}_0 \cos \omega t$ , we should use the linearized time-dependent Boltzmann equation

$$\frac{\partial}{\partial t} \delta f_{\boldsymbol{k}} - e \boldsymbol{E}(t) \cdot \boldsymbol{v}_{\boldsymbol{k}} \frac{\partial}{\partial \varepsilon_{\boldsymbol{k}}} f^0_{\varepsilon_{\boldsymbol{k}}} = -\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}}$$

We assume that  $\omega$  is much smaller than the Fermi energy, otherwise the equation above would not be valid.

After Fourier transformation,  $\delta f_{\mathbf{k}}(\omega) = \int \delta f_{\mathbf{k}}(\omega) e^{i\omega t} dt$ , one finds

$$-i\omega\delta f_{\boldsymbol{k}}(\omega) - e\boldsymbol{E}(\omega)\cdot\boldsymbol{v}_{\boldsymbol{k}}\frac{\partial f^{0}}{\partial\varepsilon_{\boldsymbol{k}}} = -\frac{\delta f_{\boldsymbol{k}}(\omega)}{\tau_{\boldsymbol{k}}}$$

which is solved by

$$\delta f_{\boldsymbol{k}}(\omega) = \frac{e\boldsymbol{v}_{\boldsymbol{k}}\cdot\boldsymbol{E}(\omega)\frac{\partial f^0_{\varepsilon_{\boldsymbol{k}}}}{\partial\varepsilon_{\boldsymbol{k}}}}{\frac{1}{\tau_{\boldsymbol{k}}} - i\omega}$$

The current density is given by

$$j^{i} = -\underbrace{2}_{\text{spin}} \int e v_{k}^{i} f_{k} \frac{\mathrm{d}k}{(2\pi)^{3}}$$

Using the definition of the conductivity,  $j^i(\omega) = \sigma^{ij}(\omega)E^j(\omega)$ , we arrive at the result

$$\sigma^{ij}(\omega) = -2\int \frac{e^2 v_{\mathbf{k}}^i v_{\mathbf{k}}^j \frac{\partial f_{\varepsilon_{\mathbf{k}}}^0}{\partial \varepsilon_{\mathbf{k}}} \cdot \tau_{\mathbf{k}}}{1 - i\omega \tau_{\mathbf{k}}} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}$$
(5.20)



Figure 5.7: The derivative of the Fermi function is sharply peaked. When the temperature is much smaller than the Fermi energy only states close to the Fermi surface contribute.

Only states close to Fermi energy contribute when the temperture is much smaller than the Fermi energy, see figure, with

$$-\frac{\partial f^0}{\partial \varepsilon}\approx \delta(\varepsilon-\mu)$$

At first glance, this appears to be different from the Drude formula

$$\sigma = \frac{ne^2\tau/m}{1-i\omega\tau}$$

where all electrons contribute as n is the total electron density. But the formulas are actually quite close to each other. To see that let us consider for simplicity a k-independent relaxation rate (e.g. rotationally invariant system) within the relevant energy window. Using  $\frac{\partial}{\partial k^i} f^0 = \frac{\partial \varepsilon}{\partial k_i} \frac{\partial}{\partial \varepsilon} f^0$ , we obtain

$$\begin{split} \sigma^{ii}(\omega) &= -\frac{e^2\tau}{1-i\omega\tau} 2 \int \frac{\partial\varepsilon}{\partial k_i} \frac{\partial}{\partial k^i} f^0_{\varepsilon_k} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3} \\ &= \frac{e^2\tau}{1-i\omega\tau} 2 \int \frac{\partial^2\varepsilon}{\partial k_i^2} f^0_{\varepsilon_k} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3} \\ &= \frac{ne^2\tau}{1-i\omega\tau} \left\langle \frac{1}{m} \right\rangle \end{split}$$

where we define an averaged effective mass by the following formula

$$\left\langle \frac{1}{m} \right\rangle = \frac{2 \int \frac{\partial^2 \varepsilon}{\partial k_i^2} f_{\varepsilon_{\mathbf{k}}}^0 \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}}{2 \int f_{\varepsilon_{\mathbf{k}}}^0 \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}}$$

With this definition of the mass term and for a momentum-independent scattering rate, we actually recover precisely the Drude formula!

### 5.5.2 Transport scattering rate

What is "transport" scattering rate  $\frac{1}{\tau_k}$  that enters the Drude formula?

As discussed in the previous chapter, ideally we would like to define the scattering rate such that it mimics the collision-integral formula. For impurity scattering, for example, one would like to have

$$\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}} = \int W_{\boldsymbol{k}\boldsymbol{k}'} \left(\delta f_{\boldsymbol{k}} - \delta f_{\boldsymbol{k}'}\right) \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$

Remarkably, a full analytic solution can be found for an isotropic (rotationally invariant) system where  $\varepsilon_{\mathbf{k}} = \varepsilon(|\mathbf{k}|)$  and where

the scattering  $W_{kk'}$  depends only on angle between k and k'. Due to the rotational invariance the scattering rate is then k independent for fixed energy,

 $\frac{1}{\tau_{\mathbf{k}}} = \frac{1}{\tau} = \text{const.}$  This motivates a simple ansatz for the distribution function

$$\delta f_{\boldsymbol{k}} \propto \boldsymbol{v}_{\boldsymbol{k}} \cdot \boldsymbol{E} = \hat{\boldsymbol{k}} \cdot \boldsymbol{a}(\varepsilon_{\boldsymbol{k}}).$$

With this ansatz we get

$$\int W_{\boldsymbol{k}\boldsymbol{k}'} \left(\delta f_{\boldsymbol{k}} - \delta f_{\boldsymbol{k}'}\right) \ \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} = \boldsymbol{a}(\varepsilon_{\boldsymbol{k}}) \int W_{\boldsymbol{k}\boldsymbol{k}'} \left(\hat{\boldsymbol{k}} - \hat{\boldsymbol{k}}'\right) \ \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$

By symmetry,  $\int W_{\boldsymbol{k}\boldsymbol{k}'}\hat{\boldsymbol{k}}' \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} \parallel \hat{\boldsymbol{k}}$ , thus  $\int W_{\boldsymbol{k}\boldsymbol{k}'}\hat{\boldsymbol{k}}' \mathrm{d}\boldsymbol{k}' = \hat{\boldsymbol{k}}\int W_{\boldsymbol{k}\boldsymbol{k}'}\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{k}} \mathrm{d}\boldsymbol{k}'$  and we obtain finally

$$\frac{\boldsymbol{a}(\varepsilon_{\boldsymbol{k}})}{\tau_{\boldsymbol{k}}}\hat{\boldsymbol{k}} = \boldsymbol{a}(\varepsilon_{\boldsymbol{k}})\hat{\boldsymbol{k}}\int W_{\boldsymbol{k}\boldsymbol{k}'}\left(1-\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{k}}\right) \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$

Therefore, the transport relaxation rate for impurity scattering is given by

$$\frac{1}{\tau_{\boldsymbol{k}}} = \int W_{\boldsymbol{k}\boldsymbol{k}'} \left( 1 - \hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{k}} \right) \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3} = \int W_{\boldsymbol{k}\boldsymbol{k}'} \left( 1 - \cos(\theta_{\boldsymbol{k},\boldsymbol{k}'}) \right) \frac{\mathrm{d}\boldsymbol{k}'}{(2\pi)^3}$$
(5.21)

where  $\theta_{\mathbf{k},\mathbf{k}'}$  is the angle between  $\mathbf{k}$  and  $\mathbf{k}'$ . The physics of the  $1 - \cos\theta$  term is discussed below. If one has impurities with the differential cross section  $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}$  and the density  $n_i$  one can is rewrite this equation (no derivation is given here) in the form

$$\frac{1}{\tau} = n_i v_{\rm F} \int (1 - \cos\theta) \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi \tag{5.22}$$

where  $v_F$  is the Ferme velocity. An intuitive interpretation of the formula is that an electron traveling the time  $\tau$  explores the volume  $v_F \tau \sigma$ . Thus it encounters on average one impurity if  $n_i \sigma v_F \tau = 1$  and therefore one expects that  $\frac{1}{\tau} = n_i \sigma v_F$  consistent with the formula above, if one identifies  $\sigma$  where  $\sigma$  with the "weighted" cross section  $\int (1 - \cos \theta) \frac{d\sigma}{d\Omega} \sin \theta \, d\theta \, d\phi$ .



**Figure 5.8:** Volume sampled by an  $e^-$  moving with  $v_{\rm F}$  in time  $\tau$ .

The factor  $1 - \cos(\theta)$  in the formulas above takes into account that one scattering processes which change the velocity give a sizable contribution to the conductivity, see Fig. 5.9. For small  $\theta$ ,  $1 - \cos \theta \approx \theta^2/2$  is small and the scattering process which contributes little to current relaxation also contributes very little to the transport scattering rate.



Figure 5.9: When the scattering angle  $\theta$  is small, the velocity changes only very little and thus the contribution of this scattering process to the scattering rate is suppressed by the factor  $1 - \cos(\theta)$ .



Figure 5.10: Change in k due to phonon scattering.

This is, for example, important for the scattering of electrons from phonos at low temperature. For acoustic phonons with the dispersion  $\omega = v_{\rm s}q$ , the typical frequency is given by  $\hbar\omega \sim k_{\rm B}T$ , thus the typical momentum transfer arising from the absorption or emission of photons is given by  $q \sim \frac{k_{\rm B}T}{\hbar v_{\rm s}}$  where  $v_{\rm s}$  is the sound velocity. For the scattering of electrons, the typical scattering angle, see Fig. 5.10, is  $\frac{q}{k_{\rm F}} \sim \frac{k_{\rm B}T}{\hbar v_{\rm s}k_{\rm F}}$ . Thus a low temperature

$$(1 - \cos \theta) \propto T^2$$

While the single-particle relaxation rate due to phonons turns out to be proportional to  $T^3$ , the transport relaxation rate obtains the extra factor  $T^2$  and is thus proportional to  $T^5$ . Thus the resistivity is also proportional to  $T^5$ . In practice this is rarely observed because  $T^5$  is so small at low T that other scattering processes (from disorder or electron-electron interactions) dominate.

Another example of how transport relaxation rates and single-particle relaxation rates can be very different concerns electron-electron interactions and the role of momentum conservation. In this case the transition rate for a scattering process

$$W_{\boldsymbol{k}_{3}\boldsymbol{k}_{4},\boldsymbol{k}_{1}\boldsymbol{k}_{2}} = \sum \delta\left(\varepsilon_{\boldsymbol{k}_{1}} + \varepsilon_{\boldsymbol{k}_{2}} - \varepsilon_{\boldsymbol{k}_{3}} - \varepsilon_{\boldsymbol{k}_{4}}\right) \delta\left(\boldsymbol{k}_{1} + \boldsymbol{k}_{2} - \boldsymbol{k}_{3} - \boldsymbol{k}_{4} - \boldsymbol{G}_{n}\right) \Gamma_{\boldsymbol{k}_{1}\boldsymbol{k}_{2}\boldsymbol{k}_{3}\boldsymbol{k}_{4}}^{n}$$

conserves momentum only modulo reciprocal lattice vectors  $G_n$ . Here it is useful to distinguish normal processes with  $G_n = 0$  from so-called "Umklapp processes" with  $G_n \neq 0$ .

#### 5.5. Conductivity and transport scattering time

Let us assume for a moment that only normal processes are relevant. In this case momentum is conserved. As a state with a finite momentum typically carries also a finite current, this implies that the conductivity (and therefore also the transport scattering time) is infinite in the absence of both disorder and Umklapp scattering. In a system with a small Fermi surface adding only momenta close to the Fermi surface, one cannot reach a reciprocal lattice vector,  $k_1 + k_2 - k_3 - k_4 \ll G$ . The probability that large momenta (which contribute to Umklapp scattering) are occupied is exonentially small, therefore also the transport scattering rate is exponentially small in this case (if 2-particle scattering is considered) and one finds

$$\frac{1}{\tau} \propto \exp\left(-\frac{\Delta}{k_{\rm B}T}\right)$$

while the single-particle scattering rate is proportional to  $T^2$  in the same system.

For a sufficiently large Fermi surface, in contrast, processes with  $k_1 + k_2 - k_3 - k_4 = G_n$  exist and one obtains

$$\frac{1}{\tau} \propto T^2$$

as in the case of the single-particle relaxation rate. Thus the resistivity,  $\rho = \frac{1}{\sigma} \propto T^2$ , is proportional to  $T^2$ .



**Figure 5.11:** Large Fermi surface allows for two-particle scattering process where the total momentum is changed by a reciprocal lattice vector.

### 5.5.3 Adding scattering processes – Matthiessen's rule

In a real material electrons scatter from disorder, other electrons and phonons. Within the relaxation time approximation, we can easily study the interplay of these scattering processes as scattering rates are simply additive within this approximation

$$\frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}^1} + \frac{\delta f_{\boldsymbol{k}}}{\tau_{\boldsymbol{k}}^2} = \left(\frac{1}{\tau_{\boldsymbol{k}}^1} + \frac{1}{\tau_{\boldsymbol{k}}^2}\right) \delta f_{\boldsymbol{k}}$$

This is clearly not exact. For example, it does not describe that electron-electron interactions are locally modified close to defects. But at least in a clean system one can hope that this gives rise only to small corrections as the elastic scattering typically will dominate over inelastic scattering close to impurities (as the latter is suppressed by Pauli blocking). The situation is, for example, different in strongly disordered systems where electrons are localized and can move from defect-state to defect state only with the help of phonons.

Coming back to weakly interacting systems, we simply add scattering rates

$$\frac{1}{\tau_{\boldsymbol{k}}} = \frac{1}{\tau_{\boldsymbol{k}}^{\text{disorder}}} + \frac{1}{\tau_{\boldsymbol{k}}^{\text{interactions}}} + \frac{1}{\tau_{\boldsymbol{k}}^{\text{phonons}}}$$

to calculate the total conductivity

$$\sigma = -\int e^2 v_{\boldsymbol{k}}^2 \tau_{\boldsymbol{k}} \frac{\partial f_{\varepsilon_{\boldsymbol{k}}}^0}{\partial \varepsilon_{\boldsymbol{k}}} \stackrel{\text{spin}}{\frown} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3}.$$

The formula simplified substantially, in cases where one can nect lect the k dependence of the scattering times. In this case the resistivity is directly proportional to  $1/\tau$  and thus the different contributions to the resistivity are additive

Matchiessen's rule: 
$$\rho \approx \rho^{\text{disorder}} + \rho^{\text{interactions}} + \rho^{\text{phonons}}$$
 (5.23)

This semiheuristic rule is often used to analyze experiments but one should keep in mind that - even within the relaxation time approximation - this is not exactly true as for k dependent scattering

$$\sigma \propto \left\langle \tau_{\boldsymbol{k}} \right\rangle = \left\langle \left( \frac{1}{\tau_{\boldsymbol{k}}^{1}} + \frac{1}{\tau_{\boldsymbol{k}}^{2}} \right)^{-1} \right\rangle \neq \left( \left\langle \frac{1}{\tau_{\boldsymbol{k}}^{1}} \right\rangle + \left\langle \frac{1}{\tau_{\boldsymbol{k}}^{2}} \right\rangle \right)^{-1}$$

In typical metals (assuming a sufficiently large Fermi surface, the different contributions to the resistivity are given by

$$\rho^{\text{disorder}} = \text{const}, \ \rho^{\text{interactions}} = AT^2, \ \rho^{\text{phonons}} \propto \begin{cases} T^5 & \text{low } T \\ T & \text{high } T \end{cases}$$

At room temperatures in a typical metal often phonons dominate giving rise to a resistivity linear in T while for  $T \to 0$  the scattering from defects is most important. At low but finite temperatures,  $\rho(T) \approx \rho_0 + A T^2$  is often a good fit.

# LATTICE VIBRATIONS

Up to now, we have mainly focussed on electrons in solids, neglecting the ions. One reason for this is simply that ions are much heavier,  $M_{\rm ion} \gg m_{\rm electron}$ , an not mobile. Nevertheless, excitations of the ionic lattice is responsible for many important effects, including the propagation of sound in a solid, thermal expansion and melting, and many of the optical properties. Furthermore, in most superconductors, phonons play the key role in providing an effective attractive interaction, see Sec. 7.3,

## 6.1 Classical Lattice Dynamics

The starting point of our discussion is the adiabatic approximation discussed in Sec. 2.1. Using  $M_{\rm ion} \gg m_{\rm electron}$ , we were approximating in that section the total wavefunction of ions and atoms as a product

$$\Phi(\underbrace{\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N}_{\text{electrons}},\underbrace{\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N}_{\text{atom positions}}) \approx \psi_{\boldsymbol{R}_1,\boldsymbol{R}_2,\ldots,\boldsymbol{R}_N}(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)\phi(\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N)$$

and we argued, that the energy of electrons for fixed ion position,  $E^{el}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$  provides an effective potential which holds the solid together. The Schrödinger equation for the ions thus reads

$$(H_{\text{ion}} + E^{\text{el}}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N))\phi(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N) = E\phi(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)$$

In the following we will combine the repulsive ion-ion potential and the attractive interaction arising from  $E^{\text{el}}(\mathbf{R}_1, \ldots, \mathbf{R}_N)$  to obtain  $U_{\text{eff}}$ 

$$H_{\text{ion}} + E^{\text{el}} = \sum_{i=1}^{N} \frac{\boldsymbol{P}_i^2}{2M} + U_{\text{eff}}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_N)$$

For the following we need only to know one important property of  $U_{\text{eff}}$ : in any crystal, it is minimized by periodic arrangement of ions  $\mathbf{R}_i = \mathbf{R}_i^0$  (otherwise no crystal would form). We set

$$oldsymbol{R}_i = oldsymbol{R}_i^0 + oldsymbol{u}(oldsymbol{R}_i^0)$$

Due to the large mass of ions, we expect that the deviation of the position of ions from the minimal value remains small,  $|\boldsymbol{u}| \ll |\boldsymbol{R}_i^0 - \boldsymbol{R}_j^0|$  (more precisely, it is sufficient to demand that  $|\boldsymbol{u}(\boldsymbol{R}_i^0) - \boldsymbol{u}(\boldsymbol{R}_j^0)| \ll |\boldsymbol{R}_i^0 - \boldsymbol{R}_j^0|$ ). Thus we can Taylor expand the effective potential in  $\boldsymbol{u}$ 

$$U_{\text{eff}} \approx U_0 + \frac{1}{2} \sum_{\mathbf{R}_i^0, \mathbf{R}_j^0} \sum_{\alpha=1}^3 \sum_{\alpha'=1}^3 u_\alpha(\mathbf{R}_i^0) D_{\alpha\alpha'}(\mathbf{R}_i^0 - \mathbf{R}_j^0) u_{\alpha'}(\mathbf{R}_j^0) + \mathcal{O}(|\mathbf{u}|^3)$$

with

$$D_{\alpha\alpha'}(\boldsymbol{R}_{i}^{0},\boldsymbol{R}_{j}^{0}) = \frac{\partial^{2}U_{\text{eff}}}{\partial R_{i\alpha}\partial R_{j\alpha}} \bigg|_{\boldsymbol{R}_{i}=\boldsymbol{R}_{i}^{0},\boldsymbol{R}_{j}=\boldsymbol{R}_{j}^{0}} \stackrel{\text{translational}}{\underset{\text{invariance}}{\overset{\text{translational}}{\overset{translational}}{\overset{tran$$

and  $D_{\alpha\alpha'}(\mathbf{R}_i^0 - \mathbf{R}_j^0) = D_{\alpha'\alpha}(\mathbf{R}_j^0 - \mathbf{R}_i^0)$  (symmetry of 2nd order partial derivative). For thermore, for inversion symmetric lattices we have  $D_{\alpha\alpha'}(\mathbf{R}_i^0 - \mathbf{R}_j^0) = D_{\alpha'\alpha}(\mathbf{R}_i^0 - \mathbf{R}_j^0)$ .

Let us first discuss the simplest case, namely a system with a single atom per unit cell. Starting from Newton's equation

$$M\frac{\partial^2}{\partial t^2}\boldsymbol{u}(\boldsymbol{R}_i^0) = -\sum_{\boldsymbol{R}_j^0} \overleftarrow{D}(\boldsymbol{R}_i^0 - \boldsymbol{R}_j^0)\boldsymbol{u}(\boldsymbol{R}_j^0)$$

we can solve this simply by a Fourier ansatz

$$\boldsymbol{u}(\boldsymbol{R},t) = \hat{\boldsymbol{e}} e^{i(\boldsymbol{k}\cdot\boldsymbol{R}-\omega t)}$$

with  $\mathbf{R} \in$  Bravais lattice,  $\mathbf{k} \in 1$ . BZ. This leads to the 3x3 matrix equation

$$\overleftrightarrow{D}(\boldsymbol{k})\hat{\boldsymbol{e}} = M\omega^{2}\hat{\boldsymbol{e}} \quad ext{with} \quad \overleftrightarrow{D}(\boldsymbol{k}) = \sum_{\boldsymbol{R}_{i}}\overleftrightarrow{D}(\boldsymbol{R}_{i}^{0})e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_{i}}$$

Diagonalization gives three eigenvalues  $\omega_{\alpha}(\mathbf{k})$ ,  $\alpha = 1, 2, 3$  (real as Bravais lattice inversion symmetric) together with the eigenvectors  $\hat{\boldsymbol{e}}_{\alpha}(\mathbf{k})$  describing the polarization of the system.

Importantly, for k 
ightarrow 0

$$\overleftrightarrow{D}(\boldsymbol{k}=0) = \sum_{\boldsymbol{R}_i} \overleftrightarrow{D}(\boldsymbol{R}_i^0) = 0$$

This follows from the fact that a uniform translation does not cost any energy,  $U_{\text{eff}}(\mathbf{R}_1, \ldots, \mathbf{R}_N) = U_{\text{eff}}(\mathbf{R}_1 + \Delta \mathbf{u}, \ldots, \mathbf{R}_N + \Delta \mathbf{u})$ . For small momenta, a Taylor expansion gives

$$\overleftarrow{D}(oldsymbol{k} 
ightarrow oldsymbol{0}) \propto oldsymbol{k}^2$$

and thus

 $\omega_{\alpha}(\mathbf{k}) = c_{\alpha}(\hat{\mathbf{k}})|\mathbf{k}|, \text{ for } |\mathbf{k}| \to 0, \text{ linear dispersion of acoustic phonons}$ (6.1)

where  $c_{\alpha}$  is the sound velocity of the so-called acoustic phonons. One usually distinguishes one branch of longitudinal *acoustic* phonon where  $\hat{\boldsymbol{e}} \parallel \boldsymbol{k}$  (approximately) from two branches of transversal acoustic phonon:  $\hat{\boldsymbol{e}} \perp \boldsymbol{k}$  (approximately) also called transversal sound. For momenta pointing in some high-symmetry direction (e.g. a rotation axis) there will be a mode with  $\hat{\boldsymbol{e}} \parallel \boldsymbol{k}$  exactly, but this is typically not the case for a randomly chosen orientation.



Figure 6.1: Longitudinal phonons.



Figure 6.2: Transverse phonons.

Note that in liquids or gases one has only longitudinal sound (due to an interplay of momentum and mass conservation), thus one can view transverse sound as the new ingredient characterizing a solid.

We also briefly the case of a system with p ions per unit cell. This implies that there are 3p degrees of freedom per unit cell and thus we obtain 3p phonon modes, 3 of them acoustic phonons with  $\omega_{\alpha}(|\mathbf{k}| \to 0) = 0$ ,  $\alpha = 1, 2, 3$ , which describe that

for  $|\mathbf{k}| \to 0$  the parallel movement of all atoms. The remaining 3(p-1) modes are called *optical* phonons and in this case one should expect that  $\omega_{\alpha}(|\mathbf{k}| \to 0) > 0$ ,  $\alpha = 4, \ldots, 3p$  (with exceptions only at structural phase transitions). In this case the ions change their mutual distance within the unit cell. Often this is associated to an oscillating dipole moment which makes them pptically active: they can absorb (and also emit) light. To compute the eigenmodes we start again from Newton's equation

$$M_i rac{\partial^2}{\partial t^2} oldsymbol{u}(oldsymbol{R}^0_i) = - \sum_{oldsymbol{R}^0_j} \overleftarrow{D} (oldsymbol{R}^0_i - oldsymbol{R}^0_j) oldsymbol{u}(oldsymbol{R}^0_j)$$

Now we have to take into account that the masses are differnt. This is done most easily by using the ansatz

$$\boldsymbol{u}(\boldsymbol{R}_{i}^{0}) = \underbrace{\frac{1}{\sqrt{M_{\mu_{i}}}}}_{\substack{\text{to get Hermitian} \\ \text{eigenvalue problem}}} \hat{\boldsymbol{e}}_{\mu_{i}} e^{i(\boldsymbol{k}\cdot\boldsymbol{R}_{i}-\omega t)}$$

with  $M_{\mu_i}$  = mass of ion  $\mu_i, \mu_i = 1, 2, \dots p$ . With this ansatz one obtains the eigenvalue equation

$$\omega^{2} \hat{\boldsymbol{e}}_{\mu} = \sum \frac{1}{\sqrt{M_{\mu}}} \overleftarrow{D}^{\mu\mu'}(\boldsymbol{k}) \frac{1}{\sqrt{M_{\mu'}}} \hat{\boldsymbol{e}}_{\mu'}$$

By solving the Hermitian  $3p \times 3p$  eigenvalue problem to obtain the eigenfrequencies  $\omega_{\alpha}(\mathbf{k})$ ,  $\alpha = 1, 2, \ldots, 3p$  and the corresponding eigenvectors  $\hat{\mathbf{e}}_{\mu\alpha}(\mathbf{k})$ .

In a tpycial 3D lattice with 2 atoms per unit cell a typical spectrum is sketched in Fig. 6.3.



Figure 6.3: Sketch of typical phonon modes in a 3d lattice with 2 atoms per unit cell.

As a simple example, let us consider a one-dimensional lattice with 2 atoms per unit-cell, considering only the motion of atoms in the x direction. Thus Newtons equation take the form



Figure 6.4: 1d lattice with 2 atoms per unit cell.

$$M_1 \frac{\partial^2}{\partial t^2} u_i^{(1)} = -\alpha \left( 2u_i^{(1)} - u_i^{(2)} - u_{i-1}^{(2)} \right) \qquad M_2 \frac{\partial^2}{\partial t^2} u_i^{(2)} = -\alpha \left( 2u_i^{(2)} - u_i^{(1)} - u_{i-1}^{(1)} \right)$$

With the ansatz  $u_j^{(1,2)} = u_{1,2}e^{i(kja-\omega t)}$  we obtain (here we are solving it with a slightly differnt ansatz than mentioned above)

$$-\omega^2 M_1 u_1 = -2\alpha u_1 + \alpha u_2 (1 + e^{-ika})$$
$$-\omega^2 M_2 u_2 = -2\alpha u_2 + \alpha u_1 (1 + e^{ika})$$

or written as a matrix equation

$$\begin{pmatrix} M_1\omega^2 - 2\alpha & \alpha(1+e^{ika})\\ \alpha(1+e^{-ika}) & M_2\omega^2 - 2\alpha \end{pmatrix} \begin{pmatrix} u_1\\ u_2 \end{pmatrix} = 0$$

We obtain the eigenmodes from det  $\begin{pmatrix} M_1\omega^2 - 2\alpha & \alpha(1+e^{ika})\\ \alpha(1+e^{-ika}) & M_2\omega^2 - 2\alpha \end{pmatrix} = 0$  which leads to

$$(M_1\omega^2 - 2\alpha)(M_2\omega^2 - 2\alpha) - \alpha^2(2 + 2\cos ka) = 0$$

and we find

$$\omega_{\pm}^{2} = \alpha \left(\frac{1}{M_{1}} + \frac{1}{M_{2}}\right) \pm \alpha \sqrt{\left(\frac{1}{M_{1}} + \frac{1}{M_{2}}\right)^{2} - \frac{4(1 - \cos^{2} ka/2)}{M_{1}M_{2}}}$$

As shown in Fig., we obtain one optical and one acoustic branch in this case.

#### 6.2. Quantum Theory of Lattice Vibration: Phonons



Figure 6.5: Phonon dispersion for a one-dimensional system with two atoms per unit cell,  $M_2 = 2M_1$ . There is one acoustic and one optical branch.

## 6.2 Quantum Theory of Lattice Vibration: Phonons

Above, we discussed a purely classical theory of oscillations of atoms but it is straightforward to boost this to a quantum theory as we understand the quantum theory of oscillators very well.

We will use exactly the techniques used to solve the 1D harmonic oscillator with Hamiltonian  $H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$ . An algebraic solution is obtained using the operators

$$a = \sqrt{\frac{m\omega}{2\hbar}}x + \sqrt{\frac{1}{2\hbar m\omega}}p, \qquad a^{\dagger} = \sqrt{\frac{m\omega}{2\hbar}}x - \sqrt{\frac{1}{2\hbar m\omega}}p$$

chosen such that one obtains the commutation relations

$$[a, a^{\dagger}] = 1, [a, a] = 0, [a^{\dagger}, a^{\dagger}] = 0$$

and that the Hamiltonian takes the simple form

$$H = \hbar\omega \left( a^{\dagger}a + \frac{1}{2} \right)$$

Here the number operatro  $\hat{n} = a^{\dagger}a$  has eigenvalues  $0, 1, 2, \ldots$ , the ground state obeys  $a|0\rangle = 0$ and and excited state can be written as  $|n\rangle = \frac{1}{\sqrt{n!}}(a^{\dagger})^{n}|0\rangle$ .

In the section above we described how we can diagonalize the classical system, effectively obtaining for each fixed  $\mathbf{k}$  3p harmonic oscillators with frequencies  $\omega = \omega_{\alpha}(\mathbf{k}), \alpha = 1, \ldots, 3p$ . Thus we simply repeat the harmonic oscillator construction sketched above. For each  $\mathbf{k}, \alpha$  we define phonon creation operator  $a_{\mathbf{k}\alpha}^{\dagger}$  with

$$[a_{\boldsymbol{k}\alpha}, a^{\dagger}_{\boldsymbol{k}'\alpha'}] = \delta_{\alpha\alpha'}\delta_{\boldsymbol{k}\boldsymbol{k}'}, \quad [a_{\boldsymbol{k}\alpha}, a_{\boldsymbol{k}'\alpha'}] = 0, \quad [a^{\dagger}_{\boldsymbol{k}\alpha}, a^{\dagger}_{\boldsymbol{k}'\alpha'}] = 0$$

using

$$\begin{cases} a_{\boldsymbol{k}\alpha} \\ a_{\boldsymbol{k}\alpha}^{\dagger} \end{cases} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}\in 1. \text{ BZ}} e^{\mp i \boldsymbol{k} \cdot \boldsymbol{R}_{i}^{0}} \left[ \sqrt{\frac{M\omega_{\alpha}(\boldsymbol{k})}{2\hbar}} \boldsymbol{u}(\boldsymbol{R}_{i}^{0}) \pm \sqrt{\frac{1}{2\hbar M\omega_{\alpha}(\boldsymbol{k})}} \boldsymbol{P}(\boldsymbol{R}_{i}^{0}) \right] \hat{\boldsymbol{e}}(\boldsymbol{k}) \end{cases}$$

or, equivalently,

Chapter 6. Lattice Vibrations

$$\boldsymbol{u}(\boldsymbol{R}_{i}^{0}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}\in1.\,\mathrm{BZ}} \sqrt{\frac{\hbar}{2M\omega_{\alpha}(\boldsymbol{k})}} \left(a_{\boldsymbol{k},-\alpha} + a_{\boldsymbol{k}\alpha}^{\dagger}\right) e^{i\boldsymbol{k}\cdot\boldsymbol{R}_{i}^{0}} \,\hat{\boldsymbol{e}}(\boldsymbol{k})$$
(6.2)

which leads by construction to the Hamiltonian

$$H_{\text{phonon}} = \sum_{\boldsymbol{k} \in 1. \text{ BZ}} \hbar \omega_{\alpha}(\boldsymbol{k}) \left( a_{\boldsymbol{k}\alpha}^{\dagger} a_{\boldsymbol{k}\alpha} + \frac{1}{2} \right)$$
(6.3)

Missing here are terms of order  $\mathcal{O}(|u|^3)$ . These non-linearities describe interactions of phonons, e.g.  $a^{\dagger}aa$ , which are necessary to describe, e.g., thermal expansion or possibly sound attenuation.



Figure 6.6: An anharmonic interaction term can lead, e.g., to a process where two phonons are converted in a single phonon.

Above, we described how to technically perform a quantization of lattice vibrations. At least equally important is the interpretation of these equations. The key is to interpret the lattice vibrations as bosonic particles, the phonons, with the energy-momentum relation  $\hbar\omega_{\alpha}(\mathbf{k})$ . The operator  $a_{\mathbf{k}\alpha}^{\dagger}$  thus creates a phonon with quantum number  $\alpha$  and momentum  $\hbar \mathbf{k}$ . Note that for the theory of lattice fluctuations it was completely irrelevant whether the underlying ions are fermionic or bosonic (as long as the atoms roughly stay in their positions as we assumed in our theory). The phonons are an example of an "emergent" particle, low-energy excitations with quantum numbers quite unrelated to those of the underlying microscopic theory.

The calculation of the elastic properties of solids and the resulting phonon modes is rather straightforward. It is so straightforward that one can easily overlook some remarkable features of the theory. To given an example, our discussion has focussed on translations but one may wonder what happens to rotations. Why is there not an extra mode with vanishing energy for  $k \to 0$  corresponding to rotations of the crystal? If such a mode would exist, it would have shown up in our calculations. So why it is not there? The answer is that it is not possible to rotate a crystal locally without breaking it apart, only a uniform rotation of the crystal is possible. This example shows that via the stiffness of the crystal some long-ranged forces can be transmitted through the system. Another example is that you cannot locally (let us say on the left side of a macroscopic crystal) change the lattice constant without changing it also on the right side at least not without making a crack in the crystal. These long-ranged forces can also have a profound impact on the nature of phase transitions. If there is a phase transition in the electronic system whose order parameter (see discussion in Sec. 7.1.2) couples directly to the lattice displacements (e.g., the endpoint of some 1st order phase transition, or a "nematic" transition where the electronic system spontaneously breaks a rotation symmetry of a crystal), these forces completely change the nature of the phase transition by strongly suppressing critical fluctuations. A good example of a fluctuation-free phase transition is one where the only the lattice constant changes which can only happen globally but not locally in a large region of length  $\xi$  if  $\xi$  becomes very large, see discussion above.

# 6.3 Thermodynamics of phonons

If interactions can be neglected, the occupation of each phonon mode is simply given by a Bose function,

$$\langle n_{\boldsymbol{k}\alpha} \rangle = \langle a_{\boldsymbol{k}\alpha}^{\dagger} a_{\boldsymbol{k}\alpha} \rangle = \frac{1}{\exp\left(\frac{\hbar\omega_{\alpha}(\boldsymbol{k})}{k_{\mathrm{B}}T}\right) - 1} = n_{\mathrm{B}}(\hbar\omega_{\alpha}(\boldsymbol{k}))$$

Using that we can calculate the expectation value for the energy

$$E(T) = \langle H \rangle = \sum_{\substack{\mathbf{k} \in 1. \text{ BZ} \\ \alpha}} \hbar \omega_{\alpha}(\mathbf{k}) \left( n_{\text{B}}(\hbar \omega_{\alpha}(\mathbf{k})) + \frac{1}{2} \right)$$

Defining a bosonic density of states

$$\mathcal{D}(\omega) = \sum_{\alpha} \int_{\boldsymbol{k} \in 1. \text{ BZ}} \delta(\omega - \omega_{\alpha}(\boldsymbol{k})) \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3}$$

the energy density e(T) is written as

$$\frac{E}{V} = e(T) = \int \hbar\omega \left( n_{\rm B}(\hbar\omega) + \frac{1}{2} \right) \mathcal{D}(\omega) \, \mathrm{d}\omega \tag{6.4}$$

For  $\omega \to 0$  only acoustic modes with  $\omega_{\alpha}(\mathbf{k}) = c_{\alpha}(\hat{\mathbf{k}})|\mathbf{k}|$  contribute which leads to a density of state quadratic in energy

$$D(\omega \to 0) = \frac{1}{(2\pi)^3} \sum_{\alpha=1}^3 \int_{\mathbf{k}\in 1. BZ} \delta(\omega - c_\alpha(\hat{\mathbf{k}})|\mathbf{k}|) |\mathbf{k}|^2 d|\mathbf{k}| d\Omega_{\mathbf{k}}$$
$$= \omega^2 \frac{1}{(2\pi)^3} \sum_{\alpha=1}^3 \int_{\mathbf{k}\in 1. BZ} \left(\frac{1}{c_\alpha(\hat{\mathbf{k}})}\right)^3 d\Omega_{\mathbf{k}} = \frac{3\omega^2}{2\pi^2} \left\langle \frac{1}{c^3} \right\rangle$$

where  $\left\langle \frac{1}{c^3} \right\rangle$  is simply he average over all polarizations and spatial directions of the cubic inverse of the sound velocities. In D dimensions ond finds instead  $\mathcal{D}(\omega \to 0) \propto \omega^{D-1}$  for phonons (and also photons).

At higher frequencies, the density of states is more complicated and – similar to the electronic case – characterized by a number of van-Hove singularities of the form s  $\pm \sqrt{\omega - \omega_0}$  in D = 3.



**Figure 6.7:** Typical sketch of density of states  $\mathcal{D}(\omega)$ .

Next, we discuss the specific heat, the most frequently measured thermodynamic quantity

$$c_{v} = \frac{1}{V} \frac{\partial E}{\partial T} = \int_{0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}T} n_{\mathrm{B}}(\hbar\omega) \mathcal{D}(\omega) \,\mathrm{d}\omega = \frac{1}{V} \sum_{\substack{\boldsymbol{k} \in 1. \mathrm{BZ} \\ \alpha}} \frac{\partial}{\partial T} \left( \frac{\hbar\omega_{\alpha}(\boldsymbol{k})}{\exp\left(\frac{\hbar\omega_{\alpha}(\boldsymbol{k})}{k_{\mathrm{B}}T}\right) - 1} \right)$$

For high temperatures,  $k_{\rm B}T \gg \max(\hbar\omega_{\alpha}(\mathbf{k}))$  we can use that  $\beta\omega_{\alpha}(\mathbf{k}) \ll 1$  and Taylor expand the Bose function  $\frac{1}{e^x - 1} \approx \frac{1}{x} \left( 1 - \frac{x}{2} + \frac{x^2}{12} - \right)$  which leads to

$$c_{v} \approx \frac{1}{V} \sum_{\substack{\mathbf{k} \in 1. \text{ BZ} \\ \alpha}} \frac{\partial}{\partial T} \left( k_{\text{B}}T - \frac{1}{2} \hbar \omega_{\alpha}(\mathbf{k}) + \frac{1}{12} \frac{(\hbar \omega_{\alpha}(\mathbf{k}))^{2}}{k_{\text{B}}T} - \dots \right)$$

$$= 3 \frac{N}{V} k_{\text{B}} \left[ \underbrace{\underbrace{1}_{\substack{\text{Dulong Petit} \\ (\text{classical result})}}_{\text{quantum correction with } \langle \omega_{\alpha}^{2} \rangle = \frac{1}{3N} \sum_{\substack{\mathbf{k}, \alpha}} \omega_{\alpha}(\mathbf{k})^{2}} + \dots \right]$$

where N is the number of atoms in the solid (p times the number of unit cells). At high temperatures, one obtains the classical result, with quantum corrections arising when the temparture is lowered.

At low temperatures, we can use that high-energy modes are exponentially suppressed,  $\frac{1}{\exp\left(\frac{\hbar\omega}{k_{\rm B}T}\right) - 1} \approx$ 

 $e^{-\frac{\hbar\omega}{k_{\rm B}T}} \ll 1$  for  $\frac{\hbar\omega}{k_{\rm B}T} \gg 1$ , and only modes with  $\frac{\hbar\omega}{k_{\rm B}T} \lesssim 1$  contribute. Thus we can use  $\mathcal{D}(\omega \to 0) = \frac{3\omega^2}{2\pi^2} \left\langle \frac{1}{c^3} \right\rangle$  as derived above. Using  $\int_{0}^{\infty} \frac{x^3}{e^x - 1} \, \mathrm{d}x = \frac{\pi^4}{15}$  one obtains for the

specific heat at low temperatures the **Debye law** 

$$c_v = \frac{2\pi^2}{5} \frac{k_{\rm B}^4}{\hbar^3} \left\langle \frac{1}{c^3} \right\rangle T^3 \tag{6.5}$$

Finally, we briefly discuss two simplified models for phonons, which are often used for fitting. For acoustic phonons this is the **Debye model** where one simply assumes that  $\omega_{\alpha}(\mathbf{k}) = c|\mathbf{k}|$  for all 3 acoustic branches. Furthermore, instead of considering sums over the full 1.BZ, one replaces it by a sphere, i.e., one allows for momenta with  $0 \leq |\mathbf{k}| \leq k_0$  and choose  $k_0$  such that:  $\frac{4\pi}{3}k_0^3 =$  volume of 1. BZ =  $\frac{(2\pi)^3}{V/N}$ , where N is the number of unit cells in the system.



Figure 6.8: Debye approximation for density of states by approximating 1. BZ with a sphere.

#### 6.3. Thermodynamics of phonons

Therefore,  $k_0 = \left(\frac{6\pi^2 N}{V}\right)^{1/3}$  which is used to definie the Debye frequency

$$\omega_D = ck_0$$

and the corresponding Debye temperature  $\Theta_{\rm D} = \frac{\hbar\omega_{\rm D}}{k_{\rm B}}$ . Thus the Debye model has the density of states

$$\mathcal{D}(\omega) = \begin{cases} \frac{3\omega^2}{2\pi^2} \frac{1}{c^3} & \omega \leqslant \omega_{\rm D} \\ 0 & \omega > \omega_{\rm D} \end{cases}$$

This allows to calculate the specific heat



Figure 6.9: Left: Plot of f(y). Right: Specific heat in the Debye approximation scheme.

Also for optical phonons there is a widely used simplified model, the **Einstein model**, where one simply approximates the frequencies by a constant:  $\omega_{\alpha}(\mathbf{k}) = \omega_0 = \text{const.}$ , which leads to



Figure 6.10: Specific heat predicted by the Einstein model.

In typical solids the Debye temperature  $\Theta_{\rm D}$  and the typical temperature related to optical phonons  $\frac{\hbar\omega_0}{k_{\rm B}}$  are in the range of 100 - 500K. In diamond values almost up to 2000 K can be reached.

## 6.4 Electron-Phonon Scattering

Also electrons couple to phonons. Physically, this simply arises because a change of the position of the ions also changes the potential experienced by the electrons. To lowest order this potential change is proportional to the relative displacement  $u(\mathbf{R}_i) - u(\mathbf{R}_j)$ . Expressed in terms of phonon creation and annihiliation operators thus the electronic density couples to terms like  $(a_{\alpha,k} + a^{\dagger}_{\alpha,-k})$ . Physically, this describes that electrons may both absorb or emit phonons.



Figure 6.11: Phonon absorption and emission processes.

Within the Botzmann equation, phonons are described by their own distribution function  $n_k(r)$  (omitting here for notational simplicity the quantum number  $\alpha$ ) and a typical collision term in the Boltzmann equation for electrons then reads

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\sum_{\mathbf{k}',\mathbf{k}''} \left( f_{\mathbf{k}} \left( 1 - f_{\mathbf{k}'} \right) n_{\mathbf{k}''} W^{A}_{\mathbf{k}\mathbf{k}',\mathbf{k}''} - f_{\mathbf{k}'} \left( 1 - f_{\mathbf{k}} \right) n_{\mathbf{k}''} W^{A}_{\mathbf{k}'\mathbf{k},\mathbf{k}''} \\ &+ f_{\mathbf{k}} \left( 1 - f_{\mathbf{k}'} \right) \left( 1 + n_{\mathbf{k}''} \right) W^{E}_{\mathbf{k}\mathbf{k}',\mathbf{k}''} - f_{\mathbf{k}'} \left( 1 - f_{\mathbf{k}} \right) \left( 1 + n_{\mathbf{k}''} \right) W^{E}_{\mathbf{k}'\mathbf{k},\mathbf{k}''} \end{aligned}$$

where the first two terms describe phonon absorption, the last two phonon emission. Similar terms (can you guess which?) arise in the Boltzmann equations for the phonons. Using that for  $T \gg \Theta_{\rm D}$ ,  $n_{\bf k}({\bf r}) \approx \frac{k_{\rm B}T}{\hbar\omega_{\bf k}}$  one finds that the resistivity of metals at high temperatures is linear in temperture as observed in many metals. In contrast, at low tempertures,  $T \ll \Theta_{\rm D}$ , there are only few phonons as  $\mathcal{D}(\omega) \propto \omega^2$ . An explicit calculation (not shown here) then allows to calculate a single-particle relaxation rate  $\propto T^3$ , while (as discussed in Sec. 5.5.2) the transport scattering rate and thus the resistivity correction to phonons scales as  $T^5$ .

# INTERACTING PHASES OF MATTER

## 7.1 Spontaneous Symmetry Breaking

One of the central questions in physics is how one can classify states of matter. Here the two most important concepts are symmetries and topology.

Here the classification according to topology starts from the question which phases of matter can "smoothly" be deformed into each other without triggering something non-analytic like the closing of a energy gap or a phase transition. By now the most famous example for this (see Sec. 3.7 and 4.1.3) are topological insulators, which can be characterized by a certain quantized winding number, have topological surface states and cannot be smoothly be transformed into a trivial band insulator if certain symmetries are obeyed.

In this section, the main topic are, however, symmetries and spontaneous symmetry breaking.

## 7.1.1 Symmetry breaking

Often the theoretical analysis of a system starts from two questions

- 1. What are the symmetries of the Hamiltonian?
- 2. What is symmetry of the ground state or of the density matrix  $\rho$ ?

Mathematically, a symmetry operation is described by a unitary operator U (or, if time reversal is involved, antiunitary operators, see Sec. 3.6.1), with

$$H = U^{\dagger} H U$$
 and/or  $\rho = U^{\dagger} \rho U$ 

Continuous symmetries like translations can be written as

$$U = e^{i\alpha Q}$$

where  $\alpha \in \mathbb{R}$  and Q is a hermition operator, the generator of the symmetry with [H, Q] = 0 as  $H = U^{\dagger}HU$  as can be shown by Taylor expanding in  $\alpha$ . For translations along the x-direction, for example, Q is identified with  $P_x$ , the momentum in x-direction and  $\alpha$  is simply the distance by which the system is translated. The equation

$$[H,Q_i] = 0$$

can be read in two different ways. The first one is that  $Q_i$  generates a symmetry as it does not change the Hamilitonian,  $H = U^{\dagger}HU$  for  $U = e^{i\alpha Q}$ . The second one is that  $Q_i$  is a conserved quantity. Therefore each continuous symmetry is directly associated to a conservation law (Noether's theorem).

Remarkably, we encounter in nature frequently situation where the symmetries of the laws of nature (i.e. of the Hamiltonian) are different from the symmetries of matter described by the density matrix  $\rho$ .

$$H = U^{\dagger} H U$$
 but  $\rho \neq U^{\dagger} \rho U$  (7.1)

This is called **spontaneous symmetry breaking**. The equivalent statement for continuous symmetries is

$$[H, Q_i] = 0 \quad \text{but} \quad [\rho, Q_i] \neq 0 \tag{7.2}$$

For a mathematically precise definition of spontaneous symmetry breaking (which we we will not try here) one has to be careful in the precise definition of the relevant density matrix  $\rho$ . Before discussing this issue, we give a few examples

A prime examples are solid, where translational symmetry is spontaneously broken. As we discussed already in Sec. 1.1, the continuous translational symmetry is spontaneously broken and only a discrete translational invariance by lattice vectors remains. This has far reaching consequences. First, one can use solids for dissipationless momentum transport: they can sustain forces without any energy consumption. Another important consequence is the presence of three types of acoustic phonons with vanishing energy,  $\omega(\mathbf{k}) \to 0$  for  $\mathbf{k} \to 0$ . Physically, this arises because a shift of the crystal doesn't cost energy and is discussed on a more formal level below in Sec. 7.1.3.

A second example, discussed in more detail in Sec. 7.2 are ferromagnets. One of the symmetries spontaneously broken by ferromagnets is time reversal invariance. Furthermore, in most cases various rotational symmetries of a crystal are spontaneously broken by a ferromagnet. If we ignore spin-orbit coupling completely, then a continous symmetry of the Hamiltonian describing such a ferromagnet is spin-rotation invariance generated by the components of the total spin of the system  $\mathbf{S} = \sum_i \mathbf{S}_i$ . A ferromagnetic state where all spins point up,  $\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow$ , breaks the spin-rotation invariance. In this case  $\rho$  does not commute with  $S_x$ . Consequences of the spontaneous symmetry breaking is that dissipationless spin-transport is possible using global rotations of the spin,  $\uparrow\uparrow\uparrow\uparrow\uparrow\to \uparrow\uparrow\uparrow\uparrow\uparrow$ . Furthermore, there are spin waves (magnons) with  $E(k) \to 0$  for  $k \to 0$  ( $E(k) \sim k^2$ ). The gapless nature of the spin-orbit interaction is neglected. If one takes spin-orbit interactions into account, the orientation of the ferromagnet is fixed to certain (high-symmetry) directions, spin-waves are gapped and dissipationless spin transport is not possible any more.

The third example, superconductors and superfluids, are discussed in Sec. 7.3. Here the spontaneously broken symmetry is related to the phase of the quantum mechanical wave function

$$\Psi \to e^{i\phi} \Psi$$

The symmetry is generated by the total charge of the system. Here the consequences (discussed in Sec. 7.3) are dissipationless charge transport and for neutral superfluids (He, ultracold atoms) one also obtains a gapless excitation. Remarkably, this is not true in the presence of long-ranged Coulomb interactions (or, equivalently, if the presence of a gauge invariance). This is related to the famous Anderson-Higgs mechanism, discussed in Sec. 7.3.2 below.
After discussing this examples, we come back to the question of how to define the density matrix  $\rho$  in the defining formula for spontaneous symmetry breaking, Eq.(7.1) or Eq. (7.2). If one takes for a *finite* system the limit  $t \to \infty$  where the system will obtain a thermal equilibrium state, with  $\rho \sim e^{-\beta H}$  then obviously  $U^{\dagger}\rho U = \rho$ . If we calculate, e.g., the expectation value of the magnetization of a ferromagnetc in such a setup we would obtain zero. The setup is, however, unrealistic as the time scale on which the magnetization changes orientation becomes larger and larger when the system size is increased: **the thermodynamic limit and the limit**  $t \to \infty$  **do not commute**. During the duration of the experiment, the magnetization does not change its orientation and one should therefore consider a density matrix with a fixed orientation of the average magnetization (in computations one can use a tiny external field to select such a state). Spontaneous symmetry breaking can therefore be thought of as arising from a **breakdown of ergodicity** which allows sufficiently large systems to select a single symmetry-broken state.

## 7.1.2 Order parameter

A powerful concept to describe symmetry broken states of matter and phase transitions is the "order parameter". An order parameter is an operator  $\phi_i(\mathbf{r})$  which "detects" locally spontaneous symmetry breaking. This is most easily seen by considering an example. To detect ferromagnetism, the obvious order parameter is the **local magnetization**  $\phi_i(\mathbf{r})$ . (in second quantization one defines  $\phi_i(\mathbf{r}) = \psi^{\dagger}_{\alpha}(\mathbf{r})\sigma^i_{\alpha\beta}\psi_{\beta}(\mathbf{r})$ ). For an s-wave superconductor, see Sec. 7.3.2, in contrast, one chooses  $\phi_i(\mathbf{r}) = \psi^{\dagger}_{\uparrow}(\mathbf{r})\psi^{\dagger}_{\downarrow}(\mathbf{r})$ . We want to use the order parameter to detect spontaneous symmetry breaking

$$\langle \phi_i(\mathbf{r}) \rangle \begin{cases} = 0 & \text{symmetry exists} \\ \neq 0 & \text{symmetry spontaneoulsy broken} \end{cases}$$

and to formulate a theory for symmetry breaking.

Here a somewhat tricky issue concerns the precise definition of  $\langle \dots \rangle$  as it depends on the precise definition of the density matrix, see discussion above. One can avoid this issue completely by studying in the thermodynamic limit the correlation function

symmetry not broken:  $\langle \phi_i^{\dagger}(\boldsymbol{r})\phi_i(\boldsymbol{r}')\rangle \to 0$  for  $|\boldsymbol{r}-\boldsymbol{r}'| \to \infty$ 

symmetry broken:  $\langle \phi_i^{\dagger}({m r}) \phi_i({m r}') 
angle 
ightarrow c$  for  $|{m r}-{m r}'| 
ightarrow \infty$ 

This condition even works both when one uses  $\rho \sim e^{-\beta H}$  where  $\langle \phi_i(\mathbf{r}) \rangle = 0$  or a density matrix where one of the symmetry broken ground states has been chosen such that  $\langle \phi_i(\mathbf{r}) \rangle \neq 0$ . In the latter case one simply obtains  $c = |\langle \phi_i(\mathbf{r}) \rangle|^2$ .



**Figure 7.1:** Landau free energy for  $T > T_c$  where  $\langle \phi \rangle = 0$ .



Figure 7.2: Landau free energy for  $T < T_c$  where the system chooses one minimum giving rise to a finite  $\phi$ .

Without going into details, let me mention a powerful theoretical approach to describe the physics of symmetry breaking. Within the **Landau theory** one expresses the free energy in terms of the order parameter. In its simplest version one considers a spatially constant order parameter (the more refined version, the Ginzburg-Landau theory, also allows for space-dependent order parameters). Close to a phase transition at the temperature  $T_c$  we can try to guess the form of the free energy as function of  $\phi$  (which is treated here as a real field, not an operator). Importantly, the Landau free energy is assumed to be an analytic function of  $\phi$  with

$$f(\phi) = \alpha (T - T_{\rm c})\phi^2 + \frac{U}{2}\phi^4$$

where  $\alpha, U > 0$  and f is the free energy per volume. Here we assumed a real order parameter and the presence of a symmetry  $\phi \to -\phi$ . For  $T > T_c$  the free energy is minimized by  $\phi = 0$ , Fig. 7.1. For  $T < T_c$  in contrast, the system chooses one of two minima, Fig. 7.2. As function of temperature the order parameter rises in a non-analytic way as depicted in Fig. 7.3.



Figure 7.3: The order parameter becomes finite below the transition temperature.

## 7.1.3 Goldstone Theorem

The Goldstone theorem states that whenever a continuous symmetry is spontaneously broken (with an exception for long-ranged interactions, see below), there exist a gapless collective mode with  $\omega \to 0$  for  $k \to 0$ . This can already be understood from our definition, Eq. (7.1). Consider a spontaneously broken *continuous* symmetry,

$$U = e^{i\alpha L}$$

with generator D which commutes with the Hamiltonian

$$[D,H] = 0$$

According to Eq. (7.1) this implies that

$$e^{i\alpha D}\rho e^{-i\alpha D} \neq \rho$$
 but  $\underbrace{\operatorname{Tr}\left(He^{i\alpha D}\rho e^{-i\alpha D}\right) = \operatorname{Tr}\left(H\rho\right)}_{\text{different states with same energy}}$ 

This implies that the system has to different states with exactly the same energy.

Let us try to formulate this a bit more precisely (and closer to the experiment). Let us assume that we can describe the system with an multicomponent order parameter  $\phi^{\alpha}$ ,  $\alpha = 1, \ldots M$  (a possible example is the magnetization of a magnet). As the order parameter has to be sensitive to the broken symmetry, it has to transform under D

$$[D, \phi^{\alpha}] = \sum_{\beta} C_{\alpha\beta} \phi^{\beta}, \ \det(C) \neq 0$$

If  $\det(C) = 0$  there are parts of the order parameter, which are not sensitive to the symmetry transformation, and one can simply construct a different order parameter with fewer components (by eliminating eigenvectors with eigenvalue 0) such that  $\det(C) \neq 0$ .

To proceed, we will need a local version of the symmetry generator and the order parameter

$$D = \int D(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, D_{\boldsymbol{q}} = \int D(\boldsymbol{x}) e^{iqx} \, \mathrm{d}\boldsymbol{x} \quad \text{with } D_{\boldsymbol{q}=\boldsymbol{0}} = D$$
$$\phi^{\alpha} = \int \phi^{\alpha}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \phi_{\boldsymbol{q}} = \int \phi^{\alpha}(\boldsymbol{x}) e^{iqx} \, \mathrm{d}\boldsymbol{x} \quad \text{with } \phi = \phi^{\alpha}$$

On a lattice one can replace the integrals by sums. The presence of a Goldstone mode is visible in the following correlation function

$$L_{\alpha}(\boldsymbol{k},\omega) = \int e^{-i\omega t} \langle [D_{\boldsymbol{k}}(t),\phi^{\alpha}_{-\boldsymbol{k}}] \rangle \frac{\mathrm{d}\omega}{2\pi} = \int e^{-i\omega t} \mathrm{Tr} \left( \rho \left[ D_{\boldsymbol{k}}(t),\phi^{\alpha}_{-\boldsymbol{k}} \right] \right) \frac{\mathrm{d}\omega}{2\pi}$$

Here we define the operators in the Heisenberg picture,  $D_{k}(t) = e^{iHt}D_{k}e^{-iHt}$ . As D commutes with the H, we obtain  $D_{k=0}(t) = D_{k=0} = D$  and thus we obtain

$$L_{\alpha}(\boldsymbol{k}=\boldsymbol{0},\omega)=C_{\alpha\beta}\delta(\omega)\mathrm{Tr}\left(\rho\phi^{\beta}
ight)$$

This allows us to formulate (with some caveats explained below) the Goldstone theorem

Goldstone theorem: If a continuous symmetry is spontaneous broken and the symmetry broken state is described by the finite order parameter  $\langle \phi \rangle$  and if interactions are short-ranged then a "Goldstone mode" with vanishing energy and diverging lifetime exists described by

$$L_{\alpha}(\boldsymbol{k} \to 0, \omega) = C_{\alpha\beta}\delta(\omega)\langle \phi^{\beta} \rangle \tag{7.3}$$

Here the condition of short-ranged interactions has been introduced to ensure that  $\lim_{k\to 0} L^{\alpha}(\mathbf{k}, \omega) = L^{\alpha}(\mathbf{k} = \mathbf{0}, \omega)$ . As we will discuss below, this is a non-trivial requirement, not fulfilled when long-ranged Coulomb interactions (or, a Gauge theory) are involved.



Figure 7.4: The function  $L_{\alpha}(\mathbf{k},\omega)$  exhibits for small momenta a sharp peak at an energy which vanishes for  $\mathbf{k} \to 0$ .

In many cases  $L_{\alpha}(\mathbf{k}, \omega)$  can be measured directly (using a scattering experiment like neutron scattering in cases where the scattered particle couples to  $\phi$ ). For fixed and small  $\mathbf{k}$ ,  $L_{\alpha}(\mathbf{k}, \omega)$  will generically exhibit a peak of width  $\Gamma_{\mathbf{k}}$  as function of  $\omega$  located at the position  $E_{\mathbf{k}}$ . The Goldstone theorem then guarantees that

$$\lim_{k \to 0} \Gamma_{\mathbf{k}} = 0, \qquad \lim_{k \to 0} E_{\mathbf{k}} = 0.$$

Often, but not always, one furthermore has the property that

$$\frac{\Gamma_{\boldsymbol{k}}}{E_{\boldsymbol{k}}} \to 0 \quad \text{for } \boldsymbol{k} \to 0$$

In this case one can identify the Goldstone mode with a **particle**, i.e., an object with a welldefined energy-momentum relation. Famous examples are the spin waves in a ferromagnet with  $E_{\mathbf{k}} \propto k^2$ , the spin-waves in an antiferromagnet with  $E_{\mathbf{k}} \propto k$  or acoustic phonons in a solid with  $E_{\mathbf{k}} \propto k$ .

In a system with local interactions one can expect that  $\lim_{k\to 0} L(k) = L(k = 0)$  for all correlation functions. This is, however, not the case in the presence of long-ranged interactions as they arise due to dipole-dipole or Coulomb interactions. Below, we give a semi-heuristic argument by studying a case where the Goldstone mode is "charged", i.e. moves charge around. Considering a finite system, we expect that in this case charge accumulates on surface as depicted in Fig. 7.5.



Figure 7.5: If a Goldstone mode couples directly to charge, it induces during an oscillation surface charges.

The surface charges cost a substantial amount of energy due to the long-ranged Coulomb interactions which we can compute using the formulas familiar from the plate capacitor. We assume a bulk charge density  $\rho_c$  which is moved by the distance  $\Delta x$ . Thus, for a surface area A, the total surface charge is given by

$$Q_{\rm surf} = A\rho_{\rm c}\Delta x$$

and the total mass moved is  $M = \rho_c \frac{m}{e} AL$ . Using the formula for the capacity (we use here SI units),  $C = \epsilon_0 \epsilon \frac{A}{L}$ , the total energy stored in the electric fields is given by

$$E = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} \frac{A^2 \rho_{\rm c}^2}{\epsilon_0 \epsilon A} L(\Delta x)^2 \label{eq:E}$$

which is a simple harmonic oscillator potential. Note that it is only the long-ranged nature of the Coulomb potential which makes this interaction energy linear in the distance L. Using the equation of motion  $M\Delta \ddot{x} = V \frac{\rho_c^2}{\epsilon} \Delta x$ , we obtain a harmonic oscillator oscillating with the "plasma frequency"

$$\omega = \omega_{\rm pl} = \sqrt{\frac{V\rho_{\rm c}^2}{\epsilon_0\epsilon\rho_{\rm c}\frac{m}{e}V}} = \sqrt{\frac{e\rho_{\rm c}}{m\epsilon_0\epsilon}}$$
(7.4)

Due to long-ranged Coulomb interaction the corresponding Goldstone mode is thus shifted to the plasma frequency in our three-dimensional example. A physical example where this occurs is a superconductor where  $\rho_c$  has to be identified with the superfluid density. While the calculation above was strongly simplified, a detailed analysis shows that it actually gives the exact result for the plasma frequency with no corrections from quantum effects or from the screeing of Coulomb interactions.

# 7.2 Magnetism

## 7.2.1 Itinerant Magnetism

As a first example, we investigat the emergence of ferromagnetism in a metal. The basic physical mechanism is that by forming a ferromagnetic state, one can avoid interaction energies. Here one makes use of the Pauli principle: two electrons with the same spin are never at the same place. For electrons with opposite spins there is no such restrictions. By polarizing spins, i.e., by having more spins of one type, one therefore reduces the energy cost of repulsive interactions. Note that Hund's rule, used to determine the spin-state of atoms, is based on the same principle.

The physics is captured in a simple mean-field descriptions. For our problem, we approximate the electronic wavefunction of non-interacting electrons. Technically, we consider the following problem

minimize  $\langle \psi | H | \psi \rangle$  with  $| \psi \rangle$  wave function for non-interacting electrons (slater determinant)

When one optimized over all possible wavefunctions this is equivalent to a (generalized) Hartree-Fock approximation discussed in Sec. 2.2. The only difference to the previous discussion is that in the non-interacting reference Hamiltian one now allows also for the presence of (ficticious) local magnetic fields coupling to the spin of the electrons. As discussed in Sec. 2.2, the full Hartree-Fock approximation is very difficult to evaluate if one tries to apply it to the full problem which involves a periodic potential and Coulomb interactions. Here, we will use a substantially simplified version, obtained by restricting the wavefunctions  $|\psi\rangle$  to a much smaller subset. We consider a model with known dispersion  $\epsilon_k$  and corresponding density of states per spin orientation.

$$\mathcal{N}(\omega) = \sum_{k} \delta(\omega - \epsilon_{k})$$

Our variational approach (here we discuss the zero-temperature version) is simply to occupy the single particle states for up and down spins differently, up to the energy

$$\epsilon_{F\uparrow/\downarrow} = \mu \pm \frac{\Delta\epsilon}{2}$$

The the density of up and down spins is given by

$$n_{\sigma} = \int_{-\infty}^{\varepsilon_{\sigma}^{\mathrm{F}}} \mathcal{N}(\omega) \,\mathrm{d}\omega, \quad \text{with} \qquad n = n_{\uparrow} + n_{\downarrow}$$

and the kinetic energy reads

$$e_{\rm kin} = \int_{-\infty}^{\varepsilon_{\downarrow}^{\rm F}} \omega \mathcal{N}(\omega) \, \mathrm{d}\omega + \int_{-\infty}^{\varepsilon_{\uparrow}^{\rm F}} \omega \mathcal{N}(\omega) \, \mathrm{d}\omega.$$

For the interactions, we use for simplicity a model with only local interactions, which we parametrized by

$$e_{int} = gn_{\uparrow}n_{\downarrow}$$



Figure 7.6: Mean-field result for the magnetization  $\langle n_{\uparrow} - n_{\downarrow} \rangle$  as function of the interaction g for the density of states  $\mathcal{N}(\omega) = \sqrt{\omega}$ .

How the parameter g is precisely related to the Coulomb interaction in a metal is a bit tricky as one has to include the effect of screening of the bare Coulomb ineraction by the presence of other electrons (a topic beyond the scope of this lecture). For a weakly interacting Hubbard model, Eq. 7.6 below, one obtains simply g = U st for small U.

Now the mean-field approach is simply defined by

minimize 
$$e_{kin} + e_{pot}$$
 for fixed  $n = n_{\uparrow} + n_{\downarrow}$ 

As a result of this minimization, one can determine the parameters  $\mu, \Delta e$  and from that the magnetization  $\langle n_{\uparrow} - n_{\downarrow} \rangle$  shown in Fig. (7.6).

Above a critical value of the interaction,  $g_c$ , the system gains energy by forming a ferromagnetic state for  $g > g_c$ . In the following we want to get some analytical insight into this result. Our goal is to calculate the "critical" coupling  $g_c$  analytically. At  $g_c$ , the effective magnetic field  $\Delta e$ , which splits up and down electrons energetically, vanishes. Therefore we will perform a Taylor  $u^{\pm \Delta e/2}$ 

expansion in 
$$\Delta e = \varepsilon_{\uparrow}^{\mathrm{F}} - \varepsilon_{\uparrow}^{\mathrm{F}}$$
. For example,  $n_{\uparrow} = \int_{-\infty}^{\mu+2\varepsilon/2} \mathcal{N}(\omega) \,\mathrm{d}\omega \approx \int_{-\infty}^{\mu} \mathcal{N}(\omega) \,\mathrm{d}\omega + \mathcal{N}(\mu) \frac{\Delta e}{2}$  and

thus with the density of states at the Fermi energy (of the non-magnetic system),  $\mathcal{N}_{\rm F} = \mathcal{N}(\mu)$ 

$$n_{\uparrow/\downarrow} = \frac{n}{2} \pm \mathcal{N}_F \frac{\Delta \varepsilon}{2},$$

The change of kinetic energy turns is given by

$$\Delta(e_{\rm kin} - \mu n) \approx \int_{\mu}^{\mu + \frac{\Delta\varepsilon}{2}} (\omega - \mu) \mathcal{N}(\omega) \, \mathrm{d}\omega + \int_{\mu}^{\mu - \frac{\Delta\varepsilon}{2}} (\omega - \mu) \mathcal{N}(\omega) \, \mathrm{d}\omega \approx \mathcal{N}_F \left(\frac{\Delta\varepsilon}{2}\right)^2$$

(A minor technical remark: subtracting the chemical potential in the equation above is a trick by which simplifies the calculation a little bit. If one calculates the change of  $e_{kin}$  only, one obtains the same result but one has to take into account a shift of  $\mu$  proportional to  $(\Delta \epsilon)^2$ which drops in the equation above).wh avoid having to calculate the shift of  $\mu$  with  $\Delta e$ . where we Interaction energy: Needed Hartee & Fock combination (see chapter 2)). Importantly, the change of kinetic energy is positive: it *costs energy* to create a ferromagnet.

In contrast, we gain energy from the interaction term when

$$\Delta e_{int} \approx g\left(\frac{n}{2} + \mathcal{N}_F \frac{\Delta \varepsilon}{2}\right) \left(\frac{n}{2} - \mathcal{N}_F \frac{\Delta \varepsilon}{2}\right) - g\left(\frac{n}{2}\right)^2 = -g\mathcal{N}_F^2 \frac{(\Delta \varepsilon)^2}{4}$$

Adding kinetic and interaction energy, the total change of energy per volume is thus

$$e_{tot} = \mathcal{N}_F \left(\frac{\Delta \varepsilon}{2}\right)^2 \left(1 - g\mathcal{N}_F\right).$$

Thus in total one *gains* energy whenever

$$g\mathcal{N}_F > 1$$
 Stoner criterion (7.5)

Thus the critical coupling is given by  $g_c = 1/\mathcal{N}_F$ . The Stoner criterium shows under what condition the formation of a ferromagnetic state will occur at zero temperature. When one uses a Taylor expansion to higher order,  $(\Delta e)^4$ , one finds that the magnetization rises with  $\sqrt{g-g_c}$  close to the transition, see Fig. 7.6. The approach discussed above can also be generalized to finite temperature. In this case, the energy has to be replace by the free energy,

where one can use that to linear order in g the free energy of the interacting system can be approximated by  $F_0 + \langle H_{int} \rangle$ , where  $F_0$  is the free energy of the noninteracting reference system. The minimization of energy is also fully equivalent to a meanfield approximation. The mean-field approximation (also discussed in the next section) can be obtained in the following way

$$H_{int} = \int g n_{\uparrow} n_{\downarrow} \approx \int g(\langle n_{\uparrow} \rangle n_{\downarrow} + n_{\uparrow} \langle n_{\downarrow} \rangle) + const. = -\int \delta \mu (n_{\downarrow} + n_{\uparrow}) + B(n_{\uparrow} - n_{\downarrow})$$

where  $\delta \mu = -g n/2$  can be absorbed in the chemical potential. The formula

$$B = g \left( \langle n_{\downarrow} \rangle - \langle n_{\uparrow} \rangle \right) / 2$$

is the mean-field equation: a finite occupation in up and down generates an effective magnetic (Zeeman) field which stabilitzes the occupation. Note that B can be identified with  $\Delta \epsilon/2$  introduced above.

## 7.2.2 Local Magnetism

Above, we discussed the origin of ferromagnetism in a metal. Another common form of magnetism is antiferromagnetism in insulators governed by a rather different mechanism.

We will look at systems with one electron per unit cell. In the absence of interactions, such a system is a good non-magnetic metal with on average 1/2 electrons per spin orientation per unit cell. If only a single band contributes, the band is only half filled. Surprisingly, these materials are, however, insulators, often with a very large gap towards charge excitations.

Being an insulater thus has to be an effect of (strong) interactions, these systems are so-called "Mott insulators". The simplest model describing this type of physics is the **Hubbard model** 

$$H = -t \sum_{\substack{\langle i,j \rangle \\ \sigma = \uparrow/\downarrow}} c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow}$$
(7.6)

The first term (written in the language of second quantization) describes the hopping of electron on nearest-neighbor sites of a lattice, the second one, with  $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$ , encodes that having both an up-electron and a down-electron on a given site costs an interaction energy U.



Figure 7.7: Sketch of a mott insulator where strong interactions prohibit the hopping of electrons which are thus localized.

For U = 0, one obtains a simple metal with a dispersion of  $\varepsilon_{\mathbf{k}} = -2t \left(\cos\left(k_x a\right) + \cos\left(k_y a\right) + \cos\left(k_z a\right)\right)$ on a simple cubic lattice. We will, however, concentrate on the opposite limit of  $U \gg t$  assuming one electron per unit cell. In this case, optimization of the interaction energy leads to a state with one electron per site as sketched in the figure. Importantly, any hopping process costs the large energy U and is thus energetically forbidden. More precisely, hopping occurs only in virtual processes discussed next.

We start our discussion by considering two neighboring sites, both carrying a spin pointing in the up direction. In this case hopping from one site to the next is not possible as the Pauli principle forbids two electrons with the same spin to occupy the same site. The situation is, however, different if the spins are initially aligned antiparallel. Denoting on a single site an empty states by  $|0\rangle$ , a single occupied state by  $|\uparrow\rangle$  and  $|\downarrow\rangle$ , and a doubly occulied state by  $|\uparrow\downarrow\rangle$ , we find that for two neighboring sites the following process is possible

$$|\uparrow,\downarrow\rangle \quad \rightarrow \begin{cases} |0,\uparrow\downarrow\rangle \\ |\uparrow\downarrow,0\rangle \end{cases} \quad \rightarrow \begin{cases} |\uparrow,\downarrow\rangle \\ |\downarrow,\uparrow\rangle \end{cases}$$

The intermediate, doubly occupied state has a very high energy of U but quantum-mechancially it can be admixed to the wave function. Second-order perturbation theory predicts that admixing this term to the wave function changes the energy by

$$\Delta E \approx -\sum_{n} \frac{|\langle n | \Delta H | 0 \rangle|^2}{E_n - E_0} = -2\frac{t^2}{U}$$

where we identified the initial state by  $|0\rangle = |\uparrow,\downarrow\rangle$ , the perturbation by the hopping term  $\Delta H = -t \sum_{\sigma=\uparrow/\downarrow} c_{i\sigma}^{\dagger} c_{j\downarrow} + c_{j\sigma}^{\dagger} c_{i\downarrow}$  for the two sites *i* and *j*, and the admixed state  $|n\rangle$  is either  $|0,\uparrow\downarrow\rangle$  or  $|\uparrow\downarrow,0\rangle$  with  $E_n - E_0 = U$ .

Importantly, we have found an energy gain for the antiparallel configuration. The system can at least do virtual hopping processes thus gaining a little bit of kinetic energy when two neighboring sites are antiparallel. We can now "guess" the correct Hamiltonian, describing our system at low energies. The localized spins have their spin orientation as the only degree of freedom. Therefore we can describe the low-energy properties using on each site *i* a spin-1/2 operators  $S_i^{\alpha}$  with  $\alpha = x, y, z, S_i^2 = \frac{3}{4}$  and the usual commutation relations for spin operators,  $[S_i^{\alpha}, S_j^{\beta}] = \delta_{ij} \epsilon^{\alpha\beta\gamma} S_i^{\gamma}$ . The resulting Hamiltonian has to be spin-rotation invariant and to lowest order in the hopping only nearest neighbors interact within our model. This leads us to the conclusion that the system is described by the **Heisenberg model** 

$$H = J \sum_{\langle i,j \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j \text{ with } J = \frac{4t^2}{U} > 0 \text{ antiferromagnetic superexchange}$$
(7.7)

The strength of the interaction, called antiferromagnetic superexchange J is obtained from the following consideration. Consider as above two neighboring spins. According to the Heisenberg model the energy difference between a  $|\uparrow,\uparrow\rangle$  and a  $|\uparrow,\downarrow\rangle$  configuration is given by

$$\Delta E = J\frac{1}{2} \cdot \frac{1}{2} - J\frac{1}{2} \cdot \left(-\frac{1}{2}\right) = \frac{J}{2}$$

Our 2nd order perturbative calculation for the Hubbard model showed that the same energy difference is  $2\frac{t^2}{U}$ . From this we can conclude that  $J = \frac{4t^2}{U}$  to leading order in 1/U.

What is the physics of the Heisenberg model? This depends quite a bit on the underlying lattice and its spatial dimension. We will focus on a cubic lattice in three dimension, or -

slightly more general - on a so-called bipartite lattice in three dimensions. A bipartite lattice has the properties that one can give labels A and B to sites in such a way that all neighbors of A sites are B sites and all neighbors of B sites are A sites. A cubic lattice in three dimension or a square lattice in two dimension have this property (a counter example is a triangular lattice where this is not possible). This is important for our discussion, as in such a system, one can easily sustain an antiferromagnetic order, where spins on neighboring lattice sites point in opposite directions.

To obtain a more quantitative understanding of the physics of the Heisenberg model, we can use again mean-field theory which amounts to replace the spins neighboring a given site i by their expectation value

$$H = J \sum_{\langle i,j \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j \approx -\sum_i \boldsymbol{S}_i \cdot \boldsymbol{B}_i + const$$

with the mean-field equation

$$\boldsymbol{B}_{\mathrm{i}} = -J \sum_{\mathrm{neighbors}} \left\langle \boldsymbol{S}_{j} \right\rangle = J \, z \left\langle \boldsymbol{S}_{i} \right\rangle$$

where z is the number of nearest neighbors (z = 6 for a cubic lattice in three dimensions) and we in the last equation that  $\langle S_j \rangle = -\langle S_i \rangle$  for nearest neighbors. The approximation to replace the operators by their expectation value gets better and better the more neighbors one has as we will discuss more extensively below. An equivant formulation of this approximation is that we have set

$$oldsymbol{S}_i \cdot oldsymbol{S}_j pprox \langle oldsymbol{S}_i 
angle \cdot oldsymbol{S}_j + oldsymbol{S}_i \cdot \langle oldsymbol{S}_j 
angle - \langle oldsymbol{S}_i 
angle \cdot \langle oldsymbol{S}_j 
angle.$$

As the Hamiltonian is purely local within the mean field approximation, one can easily solve the single-site problem  $H_i = -\mathbf{S}_i \cdot \mathbf{B}_i$  exactly using that  $\mathbf{S}_i$  describes a spin-1/2 particle. The eigenenergies are given by  $E_{\uparrow/\downarrow} = \pm \frac{1}{2}B_i$ . Therefore the eigenstates are occupied with probability  $e^{-E_{\uparrow/\downarrow}/k_{\rm B}T}/Z$  with  $Z = e^{B_i/2k_{\rm B}T} + e^{-B_i/2k_{\rm B}T}$  and one thus obtains

$$\left\langle \boldsymbol{S}_{i} \right\rangle = \frac{\frac{1}{2} \left( e^{B_{i}/2k_{\mathrm{B}}T} - e^{-B_{i}/2k_{\mathrm{B}}T} \right)}{e^{B_{i}/2k_{\mathrm{B}}T} + e^{-B_{i}/2k_{\mathrm{B}}T}} = \frac{1}{2} \tanh\left(\frac{B_{\mathrm{i}}}{2k_{\mathrm{B}}T}\right)$$

which results in the mean-field equation



Figure 7.8: For high temperatures, the only solution of the mean-field equation is  $B_{\text{eff}} = 0$ .

#### 7.2. Magnetism

As sketched in the figures, at high temperatures the effective magnetic field vanishes but a finite value of  $B_i$  and therefore of  $\langle S_i \rangle$  is obtained for  $T < T_N$ , where  $T_N$  is the so-called Neél temperature of the antiferromagnet. Note that the mean-field equation does not fix the sign of  $B_i$  but the sign has to be opposite for the A and B sublattices as we used this property in the derivation of the mean-field equations.



**Figure 7.9:** At low T, solutions with  $B_{\text{eff}} \neq 0$  and hence  $\langle S_i \rangle \neq 0$  appears in the mean-field approximation.

One can obtain an analytical formula for the transition temperature  $T_N$  using that  $B_{\text{eff}} \to 0$  for  $T \to T_N$ . Thus one can linearize the mean-field equations and finds



Figure 7.10: Order parameter plot within the mean-field approximation.

It is an important and interesting question to ask, in what sense the result obtained in this section are valid. First, the mapping of the Hubbard model to the Heisenberg model at a filling of one electron per unit cell is "exact" in the limit  $U \gg t$ . Thus the formula for J is also asymptotically exact for  $U \to \infty$ . Corrections of order  $\frac{t^3}{U^2}$  lead to longer-range spin-spin interactions. The mean-field approximation for the Heisenberg model also becomes formally exact in certain limit: this is the limit of high spatial dimension,  $d \to \infty$ , when the number of nearest neighbors z grows,  $z \to \infty$ . This just follows from the fact that in a sum of many terms, one can replace a sum by an expectation value. In three dimensions one has 8 nearest neighbors for a cubic lattice and the mean-field approximation gives reasonable results. Numerical results show that in three dimensions  $T_N \approx 0.946J$  (Sandvik 1998) to be compared to the mean-field result  $T_N = 1.5J$ . While mean-field theory notoriously overestimates the

transition temperature, it provides often a very useful estimate of qualitative features of the phase diagram. Mean-field theory fails, however, completely in low dimensions. To see that, one has to consider fluctuations around the mean-field state which are dominated by the Goldstone modes of the system. In dimensions of two and lower the fluctuations due to the Goldstone modes turn out to be so violent that the long-ranged order is destroyed at any finite temperature in two dimensions and even at zero temperature in one dimension (Mermin-Wagner theorem). In two dimension the correlation length, the size of locally ordered regions does, however, grow exponentially,  $\sim e^{cJ/T}$ , at low temperature so that for practically purposes one can often use a description based on symmetry broken states.

# 7.3 Superconductivity

Superfluidity and superconductivity, the ability to flow without any resistance, are among the most fascinating properties of matter. The superfluidity of bosonic matter arises in interacting Bose Einstein condensates, first observed in He<sup>4</sup> later in ultracold bosonic atoms. We will, however, focus here on the analog phenomenon the superconductivity of electrons (or fermions, in general). The two phenomena are related: superconductivity arises in the presence of attractive interactions and it can be viewed as a condensation of pairs of fermions. As a bound state of two fermions is a boson, it is plausible that pairs of fermions can show the same type of condensation phenomena as bosons. More precisely, it turns out that the formation of a true bound state is not required and that the "pairing" and condensation typically affects only fermions close to the Fermi surface.



**Figure 7.11:** Lattice distortion by  $e^-$  results in an effective attractive force between  $e^-$ . Importantly, the a second electron is attracted to a place previously occupied by the first electron even when it is gone. Due to this retarded nature of the attraction, it can be effective even in the presence of strong Coulomb repulsion.

Above we stated that attractive interaction are required to obtain superconductivity but Coulomb interaction is always repulsive. For most superconductors, the interaction with lattice distortions (phonons) provides the attractive interactions. Roughly, an electron distorts the surrounding lattice and the lattice distortion in turn attracts another electron. An important aspect of this attraction is that it is **retarded** (="delayed"): as ions are much heavier than electrons, they move slower and parts of the distortion remain even when the first electron is gone. This allows to generate an effective attraction with is non-local in time. Using the retarded nature of the attraction one can also circumvent the problem that the electron-electron interaction is repulsive and typically much larger than the phonons-mediated attraction. The Coulomb interaction, however is strongly reduced, when the first electron is gone while the attractive interactions remain much longer active.

Also other mechanisms can lead to an effective attractive interaction. For example, the repulsive interactions may lead to the formation of an antiferromagnets. Usually antiferromagnets do not show superconductivity, the two phases are in competition with each other. If, however, antiferromagnetism is then suppressed, then close to the "quantum phase transition" of the antiferromagnet, the point where the magnetism vanishes at zero temperature, see Fig. 7.12,

the quantum fluctuations of the antiferromagnetic order parameter can - similar to the phonons - mediate an effective attractive interaction. The transition temperature of such fluctuationmediated superconductors is often higher than that of phonon-mediated superconductors.



Figure 7.12: Phase diagram observed in many systems with relatively large superconducting transition temperatures: superconductivity can arise from attractive interactions generated by antiferromagnetic spin fluctuations.

## 7.3.1 BCS Theory

The BCS theory, named after Bardeen, Cooper, Schrieffer (1957, Nobel prize 1972), is a meanfield theory which can explain how superconductivity emerge. For simplicity, we will use in our discussion a simplified model with only attractive interactions

$$H = -\sum_{\langle i,j \rangle} c^{\dagger}_{i\sigma} c_{j\sigma} - g \sum_{i} c^{\dagger}_{i\uparrow} c^{\dagger}_{i\downarrow} c_{i\downarrow} c_{i\uparrow}$$

This is exactly the Hubbard model discussed above with U = -g. The BCS theory is based on the following mean-field approximation

$$\sum_{i} c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} c_{i\downarrow} c_{i\uparrow} \approx \sum_{i} \langle c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} \rangle c_{i\downarrow} c_{i\uparrow} + \sum_{i} c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} \langle c_{i\downarrow} c_{i\uparrow} \rangle - \sum_{i} \langle c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} \rangle \langle c_{i\downarrow} c_{i\uparrow} \rangle$$
(7.8)

In spirit, this is very similar to the mean-field theory, which we used for the Heisenberg model above, where we approximated

$$oldsymbol{S}_i \cdot oldsymbol{S}_j pprox \langle oldsymbol{S}_i 
angle \cdot oldsymbol{S}_j + oldsymbol{S}_i \cdot \langle oldsymbol{S}_j 
angle - \langle oldsymbol{S}_i 
angle \cdot \langle oldsymbol{S}_j 
angle$$

The main difference is that we now claim that the operator  $c_{i\uparrow}^{\dagger}c_{i\downarrow}^{\dagger}$ , which describes the creation of a pair of fermions, can obtain a finite expectation value. Both mathematically and physically  $\langle c_{i\uparrow}^{\dagger}c_{i\downarrow}^{\dagger}\rangle$  and  $\langle \mathbf{S}_i\rangle$  are quite different objects. For example, a quantum state  $|\psi\rangle$  with a fixed number of particles N can never have a finite expectation value  $\langle \psi | c_{i\uparrow}^{\dagger}c_{i\downarrow}^{\dagger} | \psi \rangle$  as  $c_{i\uparrow}^{\dagger}c_{i\downarrow}^{\dagger} | \psi \rangle$  and  $|\psi\rangle$  as the overlap of a state with N+2 particles and N particles vanishes.\*. A finite expectation value is only possible if the symmetry

$$c_{i,\sigma}^{\dagger} \rightarrow c_{i,\sigma}^{\dagger} e^{i\phi}$$

is spontaneously broken as can be seen by the fact that  $\langle c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} \rangle \neq \langle c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} \rangle e^{2i\phi}$ . Here one has to take into account that the total phase of a wave function is - according to the principles of

<sup>\*</sup>Thus the BCS theory and superconductivity cannot easily be understood using the language of first quantization, which builds on wavefunctions with a fixed particle number. To understand this chapter, you will need a basic knowledge in the language of second quantization on the level of an advanced quantum mechanics course.

quantum mechanics - not an observable and, furthermore, that for particles with charge like the electron there exists even a local gauge invariance  $c_{i,\sigma}^{\dagger} \rightarrow c_{i,\sigma}^{\dagger} e^{i\phi_i(t)}$ . We will ignore these (quite tricky) issues for the moment but we will come back to the coupling to gauge fields in Sec. 7.3.2.

Using the mean-field approximation, Eq. (7.8), we approximate the Hamiltonian by

$$H \simeq H_{\rm BCS} = \sum_{\substack{\langle i,j \rangle \\ \sigma}} c_{i\sigma}^{\dagger} c_{j\sigma} - \Delta \sum_{i} c_{i\uparrow}^{\dagger} c_{i\downarrow}^{\dagger} - \Delta^* \sum_{i} c_{i\uparrow} c_{i\downarrow} + const.$$
(7.9)

together with the mean-field equation

$$\Delta = g \langle c_{i\uparrow} c_{i\downarrow} \rangle_{H_{\rm BCS}} \tag{7.10}$$

Importantly, the expectation value is computed using the BCS Hamiltonian  $H_{BCS}$ . Introducing creation operators in momentum space,

$$c_{i,\sigma}^{\dagger} = \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\boldsymbol{R}_{i}} c_{\boldsymbol{k},\sigma}^{\dagger}$$

the BCS Hamiltonian can be rewritten in the form

$$H_{\rm BCS} = \sum_{\boldsymbol{k},\sigma} \left(\varepsilon_{\boldsymbol{k}} - \mu\right) c_{\boldsymbol{k}\sigma}^{\dagger} c_{\boldsymbol{k}\sigma} - \sum_{\boldsymbol{k}} \Delta_{\boldsymbol{k}} c_{\boldsymbol{k}\uparrow}^{\dagger} c_{\boldsymbol{k}\uparrow}^{\dagger} - \sum_{\boldsymbol{k}} \Delta_{\boldsymbol{k}}^{*} c_{\boldsymbol{k}\downarrow}^{\dagger} c_{\boldsymbol{k}\downarrow}^{\dagger}$$
$$= \sum_{\boldsymbol{k}} \Psi_{\boldsymbol{k}}^{*} \underbrace{\begin{pmatrix} \varepsilon_{\boldsymbol{k}} - \mu & -\Delta_{\boldsymbol{k}} \\ -\Delta_{\boldsymbol{k}}^{*} & -(\varepsilon_{\boldsymbol{k}} - \mu) \end{pmatrix}}_{M_{\boldsymbol{k}}} \Psi_{\boldsymbol{k}} + \sum_{\boldsymbol{k}} \left(\varepsilon_{\boldsymbol{k}} - \mu\right)$$
(7.11)

where in the last line, we introduced the Nambu operators

$$\Psi_{\boldsymbol{k}} = \begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^{\dagger}_{-\boldsymbol{k}\downarrow} \end{pmatrix}.$$

Our next goal is to diagonalize Eq. (7.11) using the so-called **Bogoliubov transformation**. The trick is to introduce new creation and annhibition operators which are linear combinations of  $c_{\mathbf{k}\uparrow}$  and  $c^{\dagger}_{-\mathbf{k}\downarrow}$ .

$$\begin{pmatrix} \alpha_{\boldsymbol{k}\uparrow} \\ \alpha^{\dagger}_{-\boldsymbol{k}\downarrow} \end{pmatrix} = \mathbb{U} \begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^{\dagger}_{-\boldsymbol{k}\downarrow} \end{pmatrix}$$

using the  $2\times 2$  matrix  $\mathbb U.$  We demand that these operators fulfill the usual fermionic anticommutation relations

$$\left\{ \begin{pmatrix} \alpha_{\boldsymbol{k}\uparrow} \\ \alpha^{\dagger}_{-\boldsymbol{k}\downarrow} \end{pmatrix}, \begin{pmatrix} \alpha^{\dagger}_{\boldsymbol{k}\uparrow} & \alpha_{-\boldsymbol{k}\downarrow} \end{pmatrix} \right\} \stackrel{!}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \left\{ \begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^{\dagger}_{-\boldsymbol{k}\downarrow} \end{pmatrix}, \begin{pmatrix} c^{\dagger}_{\boldsymbol{k}\uparrow} & c_{-\boldsymbol{k}\downarrow} \end{pmatrix} \right\}$$

This is fulfilled if  $\mathbb{U}_{ij}\mathbb{U}_{i'j'}^*\delta_{jj'} = (\mathbb{U}\mathbb{U}^{\dagger})_{ij} = \delta_{ij}$ . Therefore,  $\mathbb{U}$  has to be a unitary matrix,  $\mathbb{U}\mathbb{U}^{\dagger} = \mathbb{U}^{\dagger}\mathbb{U} = \mathbb{I}_{2\times 2}$  (this is only true for fermions, though, and not for the analog bosonic problem).

We choose the matrix  $\mathbb{U}$  in such a way that  $M_k$  in Eq. (7.11) is diagonal

$$\mathbb{U}M_{\boldsymbol{k}}\mathbb{U}^{\dagger} = \begin{pmatrix} E_{\boldsymbol{k}} & 0\\ 0 & -E_{-k} \end{pmatrix}$$

with the eigenvalue

$$E_{k} = E_{-k} = \sqrt{|\Delta|^{2} + (\varepsilon_{k} - \mu)^{2}}$$

Using that  $\alpha_{-\boldsymbol{k}\downarrow}\alpha^{\dagger}_{-\boldsymbol{k}\downarrow} = -\alpha^{\dagger}_{-\boldsymbol{k}\downarrow}\alpha_{-\boldsymbol{k}\downarrow} + 1$ , the resulting BCS Hamiltonian takes the form

$$H_{\rm BCS} = \sum_{\boldsymbol{k},\sigma} E_{\boldsymbol{k}} \, \alpha^{\dagger}_{\boldsymbol{k}\sigma} \alpha_{\boldsymbol{k}\sigma} + const.$$
(7.12)

Thus, excitations of the superconductor have the energy  $E_{\mathbf{k}}$ . The s-wave superconductor, which we are describing here has an energy gap  $\Delta$  as  $E_{\mathbf{k}} \geq \Delta$  for all  $\mathbf{k}$ . (Some unconventional superconductors have more complicated  $\Delta_{\mathbf{k}}$  where the gap can vanish at certain points or lines on the Fermi surface).

We still have to fix the value of  $\Delta$  using the mean field equation (7.10). Here we can use that (7.12) describes non-interacting fermions. Thus we know that

$$\left\langle \alpha_{\boldsymbol{k}\sigma}^{\dagger}\alpha_{\boldsymbol{k}'\sigma}\right\rangle = \delta_{\boldsymbol{k}\boldsymbol{k}'}\frac{1}{e^{\beta E_{\boldsymbol{k}}}+1}, \qquad \left\langle \alpha_{\boldsymbol{k}\sigma}^{\dagger}\alpha_{\boldsymbol{k}'\sigma}^{\dagger}\right\rangle = \left\langle \alpha_{\boldsymbol{k}\sigma}\alpha_{\boldsymbol{k}'\sigma}\right\rangle = 0$$

Thus, we only need to express  $\left\langle c_{i\downarrow}^{\dagger}c_{i\downarrow}^{\dagger}\right\rangle$  in turns of these variables using that

$$\left\langle c_{\boldsymbol{k}\uparrow}^{\dagger}c_{\boldsymbol{k}'\downarrow}^{\dagger}\right\rangle = \delta_{\boldsymbol{k},-\boldsymbol{k}'}\left\langle \left(\mathbb{U}_{11}^{\dagger}\alpha_{\boldsymbol{k}\uparrow}^{\dagger} + \mathbb{U}_{12}^{\dagger}\alpha_{-\boldsymbol{k}\downarrow}\right)\left(\mathbb{U}_{21}^{\dagger}\alpha_{\boldsymbol{k}\uparrow} + \mathbb{U}_{22}^{\dagger}\alpha_{-\boldsymbol{k}\downarrow}^{\dagger}\right)\right\rangle$$

After a little bit of extra algebra, one arrives at the famous BCS equation

$$\frac{\Delta}{g} = \Delta \int \frac{\tanh\left(\frac{\beta E_{k}}{2}\right)}{2E_{k}} \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^{3}} \text{ with } \Delta \neq 0 \text{ for } T < T_{\mathrm{c}}$$
(7.13)

For attractive interactions, g > 0, this equation *always* has a solution with  $\Delta \neq 0$  at sufficiently low temperatures,  $T < T_c$ . To see this and to compute the critical temperature  $T_c$ , we use that  $\Delta \to 0$  for  $T \to T_c$ . In this limit  $E_{\mathbf{k}} = |\varepsilon_{\mathbf{k}} - \mu|$  and thus we obtain for  $T = T_c$ 

$$\frac{1}{g} = \int \frac{\tanh\left(\frac{|\varepsilon_{\mathbf{k}}-\mu|}{2k_{\mathrm{B}}T_{\mathrm{c}}}\right)}{2|\varepsilon_{\mathbf{k}}-\mu|} \frac{\mathrm{d}\mathbf{k}}{(2\pi)^{3}} = \int \mathcal{N}(\omega+\mu) \frac{\tanh\left(\frac{\omega}{2k_{\mathrm{B}}T_{\mathrm{c}}}\right)}{2\omega} \,\mathrm{d}\omega$$

Due to the  $1/\omega$  term, the integral is dominated by small frequencies. Thus we approximate:  $\mathcal{N}(\omega + \mu) \approx \mathcal{N}(\mu) = N_{\rm F}$  for  $|\omega - \mu| < \omega_{\rm c}$ , where  $\omega_c$  is set in our approximation by the chemical potential (in more realistic models involving phonons, it is typically set by the Debye frequency of phonons). Furthermore, to make some analytic progress, we use  $\tanh x \approx \begin{cases} 0 & |x| < 1 \\ 1 & |x| > 1 \end{cases}$  and obtain

$$\frac{1}{g} \approx N_{\rm F} \int_{k_{\rm B}T_{\rm c}}^{\omega_{\rm c}} \frac{1}{\omega} \, \mathrm{d}\omega = N_{\rm F} \ln \frac{\omega_{\rm c}}{k_{\rm B}T_{\rm c}}$$

Thus we find for the transition temperature

$$T_{\rm c} = {\rm const} \cdot \frac{\omega_{\rm c}}{k_{\rm B}} e^{-\frac{1}{gN_{\rm F}}} \tag{7.14}$$

Importantly, for an arbitrary strength of the attractive interaction, always a finite transition temperature is obtained due to the logarithmic divergence of the integral in the low-T limit. This is very different from, e.g., our theory for ferromagnetism, where according to the Stoner criterium a sufficiently strong (repulsive) interaction was needed. In contrast, superconductors are stabilized by arbitrary weak attractive interactions but the transition temperature can be very low due to its exponential suppression for small g. Most superconductors have transition temperatures in the single-digit Kelvin range with a few remarkable exceptions. In 2020 for the first time a room-temperature superconductor was reported (Snider et al, Nature 586, 373 (2020)) in a C, S and H mixture at very high pressures. Most likely, the small mass of H and resulting large phonon frequencies is a decisive factor for the large transition temperature.

The origin of the logarithmic divergence can be traced back to the "perfect" pairing of spin up and down electrons sitting on opposite sites of the Fermi surface at  $\mathbf{k}$  and  $-\mathbf{k}$  in materials with both time-reversal and inversion symmetry. Remarkably, even in the presence of disorder the logarithmic divergence prevails. In the presence of disorder  $\mathbf{k}$  is not a good quantum number any more but the pairing of  $\mathbf{k} \uparrow$  and  $-\mathbf{k} \downarrow$  is replace by the pairing of time-reversed states of matter in the disordered material.

### 7.3.2 Supercurrents and Anderson-Higgs Mechanism

Above, we showed how one obtains from a simple mean-field equation a finite expectation value of  $\langle \psi_{\uparrow}\psi_{\downarrow}\rangle$ . In this section, we will explore how this is related to superconductivity and how superconductors affect electromagnetism. The order parameter which we want to use to describe the properties of the superconductor is the complex field

$$\phi(\boldsymbol{r}_i) = \left\langle c_{i,\uparrow} c_{i,\downarrow} \right\rangle \in \mathbb{C}$$

For our discussion, we need the free energy as a function of  $\phi(\mathbf{r})$  where we allow for  $\phi(\mathbf{r})$  to vary slowly in space on a length scale much larger than the lattice constant. In principle, one can compute this quantity using an appropriate mean-field ansatz but instead we will simply guess its form. For  $T < T_c$  the system prefers a uniform order parameter with a finite amplitude. Furthermore, by symmetry the free energy has to be invariant under  $\phi \to \phi e^{i\varphi}$ . This leads to the following guess for the Ginzburg-Landau free energy

$$F \approx \int \frac{\hbar^2}{2\tilde{m}} \left( \boldsymbol{\nabla} \phi^* \right) \left( \boldsymbol{\nabla} \phi \right) + c_1 |\phi|^2 + c_2 |\phi|^4 + \dots \, \mathrm{d} \boldsymbol{r}$$

The parameter  $\tilde{m} > 0$  has the dimension of the mass but is not directly related to the electron mass. It parameterize the "stiffness" of the order parameter, the energy costs for a spacedependent  $\phi(\mathbf{r})$ . One can view this as the first term in a Taylor expansion, where we can ignore higher order terms like  $(\nabla^2 \phi^*) (\nabla^2 \phi)$  for slowly varying  $\phi(\mathbf{r})$ . As for  $T < T_c$  a finite value for  $\phi$  minimizes the free energy, we conclude that  $c_1 < 0$  for  $T < T_c$  while typically  $c_2 > 0$ (leading to a 2nd order phase transition). Close to the phase transition,  $c_1 \propto T - T_c$  but for the following we will just assume that it is negative such that F is minimized by

$$\phi(\mathbf{r}) = \phi_0 e^{i\alpha} = const.$$

Note that the phase  $\alpha$  is undetermined.

As a next step, we want to take into account that electrons carry a charge and that the coupling of electrons to the electromagnetic field is described by a gauge theory. A gauge transformation is a simultaneous change of the phase of the electronic wave function and thus of the creation operators together with a change of the vector potential

$$c_{i\sigma} 
ightarrow c_{i\sigma} e^{-ie\lambda(\boldsymbol{r}_i,t)}$$
  
 $A_{\mu}(\boldsymbol{r},t) 
ightarrow A_{\mu}(\boldsymbol{r},t) - \partial_{\mu}\lambda(\boldsymbol{r},t)$ 

where we use in this section a convention where the charge of the electron is q = e with e < 0. As we defined  $\phi(\mathbf{r}_i) = \langle c_{i,\uparrow} c_{i,\downarrow} \rangle$ , our order parameter will transform with

$$\phi(\mathbf{r},t) \to \phi(\mathbf{r},t)e^{-i2e\,\lambda(\mathbf{r},t)}$$

where the factor 2e implies that the charge of a cooper pair described by  $\phi$  is 2e. The gauge invariance enfoces us to modify our formula for the free energy. While for a single electron with wave function (or annihilation operator)  $\psi$  the combination  $(-i\nabla - eA) \psi$  is gauge invariant, now the  $(-i\nabla - 2eA) \phi$  is the relevant gauge-invariant derivative. This implies that in the presence of a vector potential the Ginzburg-Landau free energy must have the following gaugeinvariant form

$$F \approx \int \frac{\hbar^2}{2\tilde{m}} \left( \left( \boldsymbol{\nabla} - i2e\boldsymbol{A} \right) \phi^* \right) \left( \left( \boldsymbol{\nabla} + 2ie\boldsymbol{A} \right) \phi \right) + c_1 |\phi|^2 + c_2 |\phi|^4 + \dots \, \mathrm{d}\boldsymbol{r}$$
(7.15)

One advantage of this free energy is that we can immediately read off the electric current. Even if we do not know precisely how the quantities entering in F have been obtained from the microscopic degrees of freedom, we know that for the exact theory written in terms of electron, the expectation value of the current density can be obtained from  $\frac{\delta F}{\delta A}$ . Therefore this formula has also to hold for our effective free energy and thus

$$\boldsymbol{j} = \frac{\delta F}{\delta \boldsymbol{A}} \bigg|_{\boldsymbol{A} = \boldsymbol{0}} = \frac{2e}{\tilde{m}} i \left( \phi^* (\boldsymbol{\nabla} + i2e\boldsymbol{A})\phi - \phi(\boldsymbol{\nabla} - i2e\boldsymbol{A})\phi^* \right)$$

In a situation without magnetic field and in a gauge where A vanishes we can drop A in the formula above. In the ground state the amplitude of  $\phi$  is fixed but not the phase. This motivates us to consider configurations of  $\phi$  with fixed amplitude,

$$|\phi| = |\phi_0| e^{-i\alpha(\boldsymbol{r})}.$$

In this case the current evaluates to

$$\boldsymbol{j} = \frac{2e}{\tilde{m}} |\phi_0|^2 \boldsymbol{\nabla} \alpha \tag{7.16}$$

The current is proportional to the gradient of the phase of the order parameter! This is the "**supercurrent**" which leads to dissipationless charge transport in superconductors.



Figure 7.13: Superconductor in a ring geometry with a circulating current.

To check whether this is true, we have to show that supercurrents do not easily decay. This can be seen most easily by considering a ring made from a superconductor through which a circular current is flowing, see Fig. 7.13. A state with a constant circular current is described by a phase which depends linearly on the angle  $\vartheta$  depicted in the figure.

$$\alpha(\vartheta) = n\vartheta, \quad n \in \mathbb{Z}$$

Importantly, the condition that the order parameter and current are continuous enforces that  $\exp[i\alpha(0)] = \exp[i\alpha(2\pi)]$  and thus the slope n has to be an **integer**. Importantly, this quantized slope **cannot** decay continuously and thus also the current  $I \propto |\Phi_0|^2 n$  cannot decay! This has to be contrasted with a normal conductor with a finite resistivity R. In such a system the presence of a current I leads to dissipation with the power  $RI^2$  and in a our circular wire therefore any finite current would rapidly decay as soon as R is finite. The fact that the current does not decay in a superconductor thus implies that its resistivity vanishes. The only way by which the current can decay is by a tunneling process or a thermally activated process which changes  $n \to n-1$ . This is in principle possible, e.g., by using an intermediate state where  $|\phi|$  vanishes along a cut through the wire (a phase slip event) or by moving a so-called vortex state (a line where  $|\phi|$  vanishes around which the phase winds) across the wire. For any macroscopic wire the corresponding energy barrier (which can be computed from F) is so large that both tunneling processes and thermal activation processes are efficiently exponentially suppressed that the nominal lifetime of the current carrying state becomes easily larger than the age of the universe.

The presence of a superconductor has also a number of other interesting consequences. The most famous goes under the name Anderson-Higgs mechanism or, when its physical manifestation is discussed, the Meissner effect. It is based on the simple observation that for a superconductor with  $\phi(\mathbf{r}) = \Phi_0 = const.$ , the first term in Eq.(7.15) gives a contribution to the free energy of the form

$$\Delta F = \frac{\hbar^2 (2e)^2 |\Phi_0|^2}{2\tilde{m}} \int |\mathbf{A}|^2 \,\mathrm{d}\mathbf{r}$$
(7.17)

The presence of this extra contribution to the energy immediately changes Maxwell's equation (which can be derived from Euler-Lagrange equation). Written for the vector potential the Maxwell equations now take the form

$$\left(\partial^{\mu}\partial_{\mu} + \frac{1}{\lambda^2}\right)A^{\nu} = 0, \qquad (7.18)$$

where the new term

$$\frac{1}{\lambda^2} = \frac{(2e)^2 |\Phi_0|^2}{c^2 \epsilon_0 \tilde{m}}$$

arises from Eq. (7.17) and occurs only in a superconductor.  $\lambda$  is called the **London pene**tration depth, see discussion of the Meissner effect below. In the language of high-energy physics, we find that the photon is not any more a massless particle but obtains a finite mass  $m_{ph}$  where we can identify using

$$m_{ph}c^2 = \frac{\hbar}{\lambda}c = \hbar \sqrt{\frac{(2e)^2 |\Phi_0|^2}{\epsilon_0 \tilde{m}}}$$

This is precisely the physics which leads to the mass of the Z and W bosons in the standard model. In the standard model the Higgs-field has a finite expectation value and plays exactly the role of  $\phi$  in the discussion above. The mechanism that a Gauge particle obtains a mass in the presence of a condensate and that there is no Goldstone mode, as discussed below, is called the **Anderson-Higgs mechanism** (often referred to as the Higgs mechanism). Anderson (a solid-state theorist) understood this physics first in the context of superconductivity but pointed out in 1963 also the implication for high-energy physics. A year later in 1964 Higgs was the first to write down a fully relativistic theory of this effect and also discussed the resulting Higgs particle (which does not have a direct analog in most superconductors).

The most direct experimental manifestation of the Anderson-Higgs physics is the Meissner effect. For static **B**-fields one obtains from Eq. (7.18) using  $\mathbf{B} = \nabla \times \mathbf{A}$  directly the equation

$$abla^2 oldsymbol{B} = rac{1}{\lambda^2} oldsymbol{B}$$

solved by  $B(\mathbf{r}) = B_0 e^{-x/\lambda}$ . The interpretation is that a small *B* field decays exponentially in a superconductor at its boundary. The London penetration depth  $\lambda$  thus describes to what extent a magnetic field can penetrate into a superconductor. The fact that a superconductor expels a weak magnetic field from its bulk is called the Meissner effect (or Meißner-Ochsenfeld effect).

We can now come back to a discussion of the fate of the **Goldstone mode** in a superconductor. In the absence of electromagnetic interactions the spontaneous symmetry breaking would naturally lead to a Goldstone mode describing small long-wavelength fluctuations of the phase  $\alpha$  of the complex field  $\phi = |\phi| e^{i\alpha(\mathbf{r},t)}$  with an energy linear in momentum. In a gauge theory, this is, however, not true. This can be already be seen from the fact that one can simply use a gauge transformation to get rid of  $\alpha(\mathbf{r},t)$ ,

$$|\phi| e^{i\alpha(\mathbf{r},t)} \to |\phi_0| e^{i\alpha(\mathbf{r},t) - i2e\lambda(\mathbf{r},t)} = |\phi_0|$$

by simply chosing  $\lambda(\mathbf{r},t) = \alpha(\mathbf{r},t)/(2e)$ . This implies that the degree of freedom which has previously been encoded in the phase of the order parameter has been moved to the vector potential. Indeed, an analysis of the Maxwell equation with a mass term shows that it now describes not two degrees of freedom as in the massless case (x and y-polarized light for propagation in the z-direction), but three degrees of freedom as the axes of polarization can now also point parallel to the momentum. All three of the photon modes (one of them encoding the phase fluctuations) have, as shown above, a mass. Therefore also the Goldstone mode is massive. This fits to a consideration we have made before in Sec. 7.1.3, where we discussed that long-ranged Coulomb interaction may shift an oscillating sharp mode to the plasma frequency. Indeed, we can identify the phonon mass directly with the appropriate Plasma frequency  $\omega_{pl}$ 

$$m_{ph}c^2 = \hbar\omega_{pl}$$

if we identify  $2e|\phi_0|^2$  with the charge density, 2e with the charge and  $\tilde{m}$  with the mass in Eq. (7.4).

# 7.4 Outlook: Phases of matter

In solids a remarkable broad set of different phases of matter can be realized - a short overview and a few examples is given below.

#### **Band insulators**

The most boring state of matter is a "trivial" band insulator, where electronic degrees of freedom are gapped, only phonons contribute. Band insulators arise when all electronic bands are filled, which is only possible if there is an even number of electrons per untit cell. A large part of modern technology is based on the manipulation of band-insulators like Si or GaAs (called semiconductors, because their band gap is only moderately largel). In **topological insulators**, in contrast, the topology of the bulk wave function enforces the existence of gapless surface states, typically described by Dirac equations (see Sec. 3.7). A prominent example of this class are systems showing the **anomalous quantum Hall effects** (Sec. 4.1.3).

#### **Correlated Insulators**

The insulators described above all have an even number of electrons per unit cell such. Strong interactions can, however, also induce insulating behavior for an odd number of electrons per unit cell. This can happen because of a doubling of the size of the unit cell by spontaneous symmetry breaking (e.g., by antiferromagnetic order) driven by strong interactions. In a **Mott insulator**, however, this happens even in the absence of ordering, see Sec. (7.2.2). Mott insulators have local degrees of freedom which can induce magnetism, various types of "exotic order" or quantum liquid behavior, see below.

#### Anderson Insulators and Many-Body Localization

A completely different mechanism to obtain an insulator arises from disorder. Scattering from defects can lead to interference and bound states and, ultimately, to the localization of wave functions. This mechanism is most effective in one spatial dimension, where any amount of disorder generically leads to localization, while in three dimension localization occurs only for very strong disorder. In two dimension, for simple potential scattering and without spin-orbit interactions, also all states are localized but the localization length is exponentially large for weak disorder. Localization by disorder is called **Anderson-localization** and in its plain-vanilla form it occurs only in non-interaction systems without electron-electron or electron-phonon interactions. In a generic system with interactions and phonons, however, localization of this type happens only at zero temperature, while at finite temperature the resistivity is finite as electrons may, e.g., move by phonon-assisted hopping processes. In these cases the resistivity is often exponentially large at low temperatures. Remarkably, for certain 1-dimensional interacting models with short-ranged interactions and without phonons it was shown that exact localization may exist even at finite temperature. In this case, called many-body localization, the many-particle wave functions (rather than the single-particle wave functions) are localized and, remarkably, exhibit quantum coherence even in high-energy eigenstates. Furthermore, such systems never equilibrate.

#### Glasses

Glasses are also states of matter which do not equilibrate, at least not on experimentally relevant time scales. In structural glasses, the strongly disordered atoms do not manange to arrange on regular lattices but get stuck in some local minimum of the (free) energy. In spinglasses some magnetic structure "freezes" into some local minima. Many experimental systems show some type of glassy behavior at very low temperature.

#### Metals

Metals are characterized by a Fermi surface. Pauli principle strongly suppresses scattering at low temperatures and energies. Thus a simple description in terms of weakly interacting

#### 7.4. Outlook: Phases of matter

fermionic quasiparticles becomes possible, the **Fermi liquid theory**. Only in few cases, this paradigm cannot be used: in one-dimensional systems the Fermi liquid description breaks down and is replaced by the paradigm of a **Luttinger liquid**, where the physics turns out to be governed by weakly interacting charge-density and spin-density waves. A breakdown of Fermi liquid theory may also occur close to quantum phase transitions, i.e., at the onset of some zero-temperature phase transition.

#### Semimetals

In some cases, the Fermi surface may shrinl to a single point, in systems which are described by the Dirac- or the Weyl equation. Such systems allow to realize experimentally some of the phenomena originally predicted in high-energy physics (anomalies, Klein tunnelling, ....). For Weyl metals, topology enforces the presence special surface states, so called Fermi arcs, see Sec. 3.7.3

#### **Superconductors**

Superconductors arise due to attractive interactions. Superconductors conduct charge without dissipation & create massive photons via the Anderson Higgs mechanism. A general form of the order parameter is

$$\Delta_{\boldsymbol{k}}^{\alpha\beta} = \langle c_{\boldsymbol{k},\alpha} c_{-\boldsymbol{k},\beta} \rangle$$

In Sec. 7.3 we have concentrated on a simple spin-singlet s-wave superconductor with no k dependence as those are realized in almost all phonon-driven superconductors. There are, however, other types of superconductors, so-called **unconventional superconductors**, mostly driven by other mechanisms like spin-fluctuations. For example in d-wave superconductors realized in the CuO planes of high-temperature superconductors one obtains a d-wave superconductor with a different k-dependence,  $\Delta_k^{\alpha\beta} \sim k_x^2 - k_y^2$ . The order parameter vanishes for  $|k_x| = |k_y|$  which leads to gapless quasiparticles. More recently, a new class of superconductors, so-called **topological superconductors** have been studied intensively. They are gapped in the bulk but have gapless surface states described as superpositions of electrons and holes (Majorana quasiparticles). Topological superconductors may become important for future quantum computers. The surface mode of a one-dimensional topological superconductor or the vortex core of a two-dimensional superconductor hosts a zero-energy Majorana mode (described by an operator  $c^{\dagger} \pm c$ ) which can be used to store quantum information in a non-local way and protected from dephasing.

#### Magnets

The most common magnets are ferromagnets (mostly metallic) and antiferromagnets (often insulating). But also more complicated magnetization pattern can be stabilized. They also can form "topological magnetic textures" characterized by quantized winding numbers.

#### **Exotic Order**

Besides magnetism and superconductivity, a zoo of different ordering phenomena may occur in solids. For example, the orbital degree of freedom (known from atomic orbitals) can order, there can be combinations of charge-, spin-, orbital- and superconducting order. Also current patterns may, for example, form. A large, partially unexplored research field concerns the combination of topology and order. A simple example is an antiferromagnet which is also a topological insulator. Also defects in ordered phases may have special topological properties (like a superconducting vortex hosting Majorana states as mentioned above).

#### Quantum Liquids

Upon cooling matter, typically some sort of order sets in. Liquids become solids, metals may form superconductors. Also most Mott insulator with local degree of freedom (e.g., localized spins) show some sort of ordering phenomena. Quantum liquids are states of matter where this simple type of ordering, captured by local order parameters does not happen. For example, two spins can form a non-magnetic singlet states. Such singlets can form a periodic structure, spontaneously breaking symmetries, or, alternatively, form a liquid instead. Such quantum liquids can have different types of properties, e.g., the ground state can have a gap separating it from excited states or can be gapless. A common property is, however, that the quantum number of the low-energy excitations of quantum liquids are typically differ from the quantum numbers characterizing the constituents from which the quantum liquid has formed. While there are many possible quantum liquids and many theories of quantum liquids, there are only a few fully confirmed experimental systems forming such quantum liquids. The most famous example is arguably the fractional quantum Hall effect in two-dimensional systems. In these systems the Hall conductivity is a fraction n/m of the conductance quantum. For example for n/m = 1/3 the excitations of the quantum Hall state turn out have a charge of 1/3 of an electron charge (and also unusual statistical properties). The emergence of new quantum numbers is theoretically best described by gauge theories, which naturally arise when new degrees of freedom with with new quantum numbers are introduced theoretically.

#### Beyond thermal equilibrium

Most of the phases of matter discussed above (with the notable exception of glasses and manybody localized states) occur in thermal equilibrium. There is, however, a much less explored world of phenomena beyond the paradigm of thermal equilibrium. A laser or a simple solar cell connected to a circuit are an example of simple driven systems out of thermal equilibrium.

# 7.5 Concluding remarks

Most of the rich variety of physical phenomena observed in solids "emerges" from the interplay of a giant number of quantum particles. Taking into account how difficult it is to solve the Schrödinger equation already for a handful of interacting particles, it is amazing, that theoretical physics can manage to understand and describe many different phases of matter. Remarkably, in many cases it is possible to find a simple description of the relevant low-energy physics. The success of band-structure theory, which essentially describes complex interacting systems in terms of non-interacting fermions is just one such example. Obtaining simple descriptions for complex, emergent phenomena is a major goal of solid state theory.

There are many reasons why solid state theory is an exciting field of research. One aspect is certainly the application potential. The technology of the world around us is a proof for this – and many challenges lie ahead on the quest to build, e.g., a quantum computer, or to develop environmentally more friendly technologies. But also from the viewpoint of fundamental physics, solid state physics is a field with a high potential to discover something completely new both in experiment and theory.

# ACKNOWLEDGEMENT

This script on solid state theory was written in the winter semester 2020/2021 where the Corona pandemic made real-life teaching impossible. The lecture was therefore given in a video format and accompanied by the script, problem sheets and tutorials. The course was running over 14 weeks. There have been 21 lectures with a duration of 75-95 min (on average 1.5 lectures per week) split into more than 80 videos. Alternating there have been tutorials (discussing a bi-weekly problem sheet) and discussion sessions.

I would like to apologize for many mistakes and omissions in the script. No proof-reading was done on the manuscript. I would like to thank Vivek Lohani who made a first version of the manuscript based on my hand-written lecture notes, by transfering the formulas and a large fraction of the figures to LaTeX using LyX. Thanks also to Jinhong Park and Anne Matthies who worked on the problem sets and the tutorials for the lecture, an indispensible part of the online teaching.

Cologne, 30.1.2021

Achim Rosch