Jacob W. Linder

Lecture Notes for TFY4245
Solid State Physics, Advanced Course

Norwegian University of

Science and Technology

Faculty of Natural Sciences and Technology

Department of Physics

Contents

Preface

This document contains the lecture notes defining the curriculum for TFY4245 - Solid State Physics, Advanced Course at NTNU. The chapters in these lecture notes are heavily based on (and several parts of the text is taken directly from) the following references:

- Chapter 1: Introduction [3].
- Chapter 2: Electrons in the periodic crystal and second quantization [3, 6].
- Chapter 3: Optical processes and excitations [3].
- Chapter 4: Phonons, plasmons, polaritons, polarons [3, 4, 5, 7, 8].
- Chapter 5: Dielectrics and ferroelectrics [7, 9, 11, 12].
- Chapter 6: Superconductivity [2, 10].
- Chapter 7: Diamagnetism, paramagnetism, and ordered magnetic states [6, 1, 3].

A few paragraphs and footnotes with extra details have also been adapted from particularly clear explanations on Wikipedia. These lecture notes should therefore be viewed as a compilation of various key topics in condensed matter physics extracted from scientific literature across the web, in particular the references mentioned above which should be given the full credit for the content in these notes. We will use SI units throughout this document.

Chapter 1

Introduction

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

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

$$H = H_{\text{e}} + H_{\text{ion}} + H_{\text{ion-e}} \tag{1.1}$$

where

$$\begin{aligned} H_{\text{e}} &= \sum_i \frac{\hat{\boldsymbol{p}}^2}{2m} + \frac{1}{2}\sum_{i\neq i'} \frac{e^2}{4\pi\varepsilon_0|\boldsymbol{r}_i - \boldsymbol{r}_{i'}|}, \\ H_{\text{ion}} &= \sum_j \frac{\hat{\boldsymbol{P}}_j^2}{2M_j} + \frac{1}{2}\sum_{j\neq j'} \frac{e^2 Z_j Z_{j'}}{4\pi\varepsilon_0|\boldsymbol{R}_i - \boldsymbol{R}_{j'}|}, \\ H_{\text{ion-e}} &= -\sum_{ij} \frac{Z_j e^2}{4\pi\varepsilon_0|\boldsymbol{r}_i - \boldsymbol{R}_j|}. \end{aligned} \tag{1.2}$$

Here, m is the electron mass, M_j and Z_j is the mass and atomic (charge) number of ion j, and e the elementary charge. Although being accurate, in practice such a Hamiltonian is wildly impractical, and in fact impossible, to use for a macroscopic number of particles that is present in a real material. Fortunately, there exists a number of methodologies that are based on clever approximations made to H which can be used to describe the fascinating physics taking place in materials.

In condensed matter physics, we are dealing with phenomena occurring at room temperature $T \simeq 300$ K or below, corresponding to a characteristic thermal energy of about $E = k_B T \simeq 0.03$ eV $= 30$ meV, which is much smaller than the energy scale that particle physics concerns itself with

(MeV and higher). Correspondingly, the important length scales in condensed matter physics are given by the extension of the system and of the spatial behavior of the electronic wave functions. The focus is thus quite different from the one of high-energy physics. There, a highly successful phenomenological theory for "low energies" in a particle physics context, the so-called standard model, exists, whereas the underlying theory for higher energies is unknown. In solid state physics, the situation is reversed. The Hamiltonian we gave above describes the behavior of the particles at "high-energies" in a condensed matter physics context and the aim is instead to describe the low-energy properties using reduced (effective, phenomenological) theories. Both tasks are far from trivial. Among the various states of condensed matter that solid state theory seeks to describe are metals, semiconductors, and insulators. Furthermore, there are phenomena such as magnetism, superconductivity, ferroelectricity, charge ordering, and the quantum Hall effect. All of these states share a common origin: Electrons interacting among themselves and with the ions through the Coulomb interaction.

The microscopic formulation in Eq. (1.2) is too complicated to allow an understanding of the low-energy behavior. Consequently, the formulation of effective (reduced) theories is an important step in condensed matter theory. On the one hand, characterizing the ground state of a system is an important goal in itself. However, measurable quantities are influenced by excited states, so that the concept of 'elementary excitations' takes on a central role. Some celebrated examples are Landau's quasiparticles for Fermi liquids, the phonons connected to lattice vibrations, and magnons in ferromagnets. The idea is to treat the ground state as an effective vacuum in the sense of second quantization, with the elementary excitations as particles on that vacuum. Depending on the system, the vacuum may be the Fermi sea or some state with a broken symmetry, like a ferromagnet, a superconductor, or the crystal lattice itself. According to P. W. Anderson, the description of the properties of materials rests on two principles: the principle of adiabatic continuity and the principle of spontaneously broken symmetry.

By adiabatic continuity we mean that complicated systems may be replaced by simpler systems that have the same essential properties in the sense that the two systems may be adiabatically deformed into each other without changing qualitative properties. Arguably the most impressive example is Landau's Fermi liquid theory mentioned above. The low-energy properties of strongly interacting electrons are the same as those of non-interacting fermions with renormalized parameters.

On the other hand, phase transitions into states with qualitatively different properties can often be characterized by broken symmetries. In magnetically ordered states the rotational symmetry and the time-reversal invariance are broken, whereas in the superconducting state the global gauge symmetry is. In many cases the violation of a symmetry is a guiding principle which helps to simplify the theoretical description considerably. Moreover, in recent years some systems have been recognized as having topological order which may be considered as a further principle to characterize low-energy states of matter. A famous example for this is found in the context of the quantum Hall effect.

The goal of these lectures is to introduce these basic concepts on which virtually all more elaborate methods are building up. In the course of this, we will cover a wide range of frequently encountered ground states, starting with the theory of metals and semiconductors, proceeding with dielectrics, ferroelectrics, superconductors, and finally magnetically ordered states. During this journey, we will encounter several types of emergent quasiparticles that appear in these types of materials, formed by the interplay of physical electrons and ions, such as plasmons, phonons, polarons, polaritons,

magnons, and Cooper pairs.

1.1 Is fundamental physics useful for anything?

Is the fundamental research done on quantum condensed matter physics only justified by its potential to create new devices and technology?

This author believes the answer is no: gaining knowledge on how nature works has intrinsic value. There is nothing wrong with a long-term vision of technological applications emerging in the future from basic research, but discovering the solution to what is currently unknown and pushing the frontier of human knowledge forward, and disseminating these advances to others, is a key to the very understanding of the world around us.

With this in mind, it is interesting to note that there exist several examples of *major technological breakthroughs that much later originate from what at the time was considered fundamental research without any orientation toward applications.* Indeed, fundamental research has directly led to game-changing applications that were not anticipated when the research was undertaken. The challenge for investors is that since most applications arising in this way cannot be foreseen, the time interval between a basic research discovery and its application in technology may be years and decades. Below are some examples with text partially from *"The value of fundamental research"* by Cris dos Remedios (2006) and Wikipedia.

- Magnetic resonance imaging technology is now used widely for diagnosis (a non-invasive technique) and for research, with an increasingly wide variety of applications. Key developments include Carr and Purcell, who in 1954 described the use of magnetic field gradients to relate NMR frequencies to spatial position. Subsequently, magnetic resonance imaging became a multi-million industry although none of the early scientists had any idea that their work would lead to medical imaging technology of such practical significance.

- It is well known that fundamental studies of electromagnetic fields by Hertz and Maxwell underpin radio and television, developments that they clearly could not have imagined. The laser also falls into that category. The first working laser was built by Theodore Maiman in 1960. This idea came directly from atomic physics and, in particular, from principles of stimulated emission discovered by Einstein several decades before. With the subsequent development of gas lasers, these intense and coherent light sources found applications in experimental physics, enabled holographic and interferometric studies, and later were used for range finding and surveying. But the most important applications came with the development of solid state lasers and their use in fibre optic communications. Lasers are now used throughout the world in medicine, in consumer electronics, and in scanning and printing technology. Modern communications and data storage technologies depend on laser optics and in the next decade or so, optronic laser-based computers will supersede the electronic systems of today.

- The giant magnetoresistance effect is a good example of how an unexpected fundamental scientific discovery can quickly give rise to new technologies and commercial products, which is a strong argument for both public and private support of basic research. It has been noted that GMR can "be considered one of the first real applications of the promising field of nanotechnology."

- The exactness of the quantized Hall conductance was discovered during routine resistance

measurements. Actual measurements of the Hall conductance have been found to be integer or fractional multiples of e^2/h to nearly one part in a billion. It has allowed for the definition of a new practical standard for electrical resistance, based on the resistance quantum given by the von Klitzing constant. This is named after Klaus von Klitzing, the discoverer of exact quantization. Since 1990, a fixed conventional value RK-90 is used in resistance calibrations worldwide. The quantum Hall effect also provides an extremely precise independent determination of the fine structure constant, a quantity of fundamental importance in quantum electrodynamics.

- The first working transistor was invented in 1947, and its discoverers shared the Nobel Prize in Physics in 1956. The transistor originated from experiments performed at Bell Labs in New Jersey where researchers performed experiments and saw that when two gold point contacts were applied to the material germanium, a signal was produced with the output power greater than the input. The group leader for solid state physics William Shockley saw the potential in this, and over the next few months worked to greatly expand the knowledge of semiconductors. The most widely used transistor is the metal-oxide-semicondutor field-effect transistor (MOSFET) which properties is a direct interplay between the fundamental physics of three types of materials: metals, semiconductors, insulators. Transistors are the key active components in practically all modern electronics. Their primary use is to amplify or switch electrical signals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal. Many people consider transistors one of the 20th century's greatest inventions, as they enable compact, powerful computers and cell phones.

Chapter 2

Electrons in the periodic crystal and second quantization

One of the characteristic features of many solids is the regular arrangement of their atoms forming a crystal. Electrons moving in such a crystal are subject to a periodic potential which originates from the lattice of ions and an averaged electron-electron interaction (like Hartree-Fock approximation). The spectrum of extended electronic states, i.e. delocalized eigenstates of the Schrödinger equation, form bands of allowed energies and gaps of "forbidden" energies. One can understand band formation in two complementary ways: (1) the free electron gas whose continuous spectrum is broken up into bands under the influence of a periodic potential and (2) independent atoms are brought together into a lattice until the outer-most electronic states overlap and lead to delocalized states turning a discrete set of states into continua of electronic energies - bands.

In this chapter we will address the emergence of band structures of electrons in a periodic crystal. The band structure of electrons is essential for the basic classification of materials into metals and insulators (semiconductors).

2.1 Symmetries of crystals

Most solids consist of a regular lattice of atoms with a perfectly repeating structure. Defects may be present in realistic experimental samples, such as missing atoms, which can be treated as impurities in the system. However, we first consider a perfect crystal to establish the foundation for describing the behavior of electrons in such materials. The minimal repeating unit is the unit cell. The symmetries of a crystal are contained in the space group $\mathcal{R}$, a group of symmetry operations (translations, rotations, the inversion or combinations) under which the crystal is left invariant. In three dimensions, there are over 200 different space groups.

All symmetry transformations form together a set which has the properties of a group. A group $\mathcal{G}$ combined with a multiplication $*$ has the following properties:

- The product of two elements of $\mathcal{G}$ are also in $\mathcal{G}$: $a, b \in \mathcal{G} \rightarrow a * b = c \in \mathcal{G}$.
- Multiplications are associative: $a * (b * c) = (a * b) * c$.
- A unit element $e \in \mathcal{G}$ exists with $e * a = a * e = a$ for all $a \in \mathcal{G}$.
- For every element $a \in \mathcal{G}$ there is an inverse $a^{-1} \in \mathcal{G}$ satisfying $a^{-1} * a = a * a^{-1} = e$.

A group with $a * b = b * a$ for all pairs of elements is called an Abelian group, otherwise it is known as non-Abelian (meaning the elements do not commute with each other). A subset of elements $\mathcal{G}'$ from the group $\mathcal{G}$ is called a subgroup of $\mathcal{G}$ if it satisfies the criteria of a group as well.

We consider here a crystal displayed as a point lattice, each point symbolizing either an atom or a whole unit cell. See Fig. 2.1 for a square lattice geometry. Translations in the space group are represented by linear combinations of a basic set of translation vectors $\{\boldsymbol{a}_i\}$ (so-called primitive lattice vectors) connecting lattice points. Any translation between arbitrary lattice points is then a linear combination of $\{\boldsymbol{a}_i\}$ with integer coefficients:

$$\boldsymbol{R}_n = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3. \tag{2.1}$$

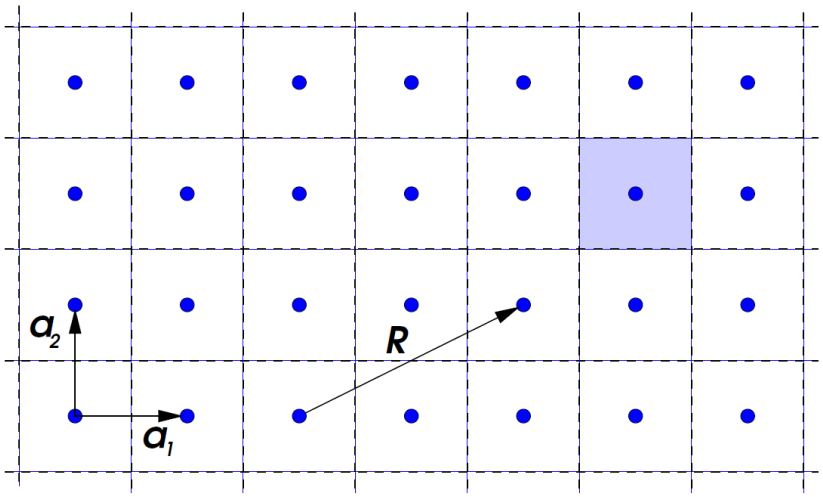

Figure 2.1: Crystal point lattice in two dimensions: the vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ form the basic set of translations and $\boldsymbol{R} = 2\boldsymbol{a}_1 + \boldsymbol{a}_2$ in the figure. The shaded area denotes the Wigner-Seitz cell which is obtained by drawing perpendicular lines (planes in three dimensions) through the center of all lines that are connecting neighboring lattice points. The Wigner-Seitz cell also constitutes a unit cell of the lattice. Figure taken from [3].

General symmetry transformations including general elements of the space group may be written in the notation due to Wigner

$$\boldsymbol{r}' = g\boldsymbol{r} + \boldsymbol{a} = \{g|\boldsymbol{a}\}\boldsymbol{r} \tag{2.2}$$

where g represents a rotation, reflection or inversion with respect to lattice points, axes, or planes. The elements g form the *generating point group* $\mathcal{P}$. In three dimensions there are 32 points groups.

We can distinguish between the following basic symmetry operations:

- Basic translations $\{E|\boldsymbol{a}\}$.
- Rotations, reflections, inversions: $\{g|\boldsymbol{0}\}$.
- Screw axes, glide planes: $\{g|\boldsymbol{a}\}$.

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

$$\{g|\boldsymbol{a}\}\{g'|\boldsymbol{a}'\} = \{gg'|g\boldsymbol{a}' + \boldsymbol{a}\} \tag{2.3}$$

form a group with unit element $\{E|\boldsymbol{0}\}$. In general, these groups are non-Abelian, i.e. the group elements do not commute with each other. However, an Abelian subgroup of the space group $\mathcal{R}$ always exists, namely the group of translations $\{E|\boldsymbol{a}\}$.

The elements $g \in \mathcal{P}$ do not necessarily form a subgroup, because of some of these elements (such as screw axes or glide planes) leave the lattice invariant only when combined with translation. If $\mathcal{P}$ is a subgroup of $\mathcal{R}$, then $\mathcal{R}$ is said to be symmorphic, in which case the space group contains only primitive translations and neither screw axes nor glide planes.
The most fundamental type of lattice is called a *Bravais lattice*. There are two equivalent definitions.

- (a) An infinite array of discrete points with an arrangement and orientation that appear exactly the same, from whichever of the points the array is viewed.
- (b) A lattice consisting of all points with positions vector $\boldsymbol{R}$ of the form

  $$\boldsymbol{R} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3 \tag{2.4}$$

  with integer n_j. This can also be restated as saying that every point of a Bravais lattice can be reached from any other point by a finite number of translations.

Non-Bravais lattice contains points which cannot be reached by translations only. Rotations and reflections must be used in addition to translation. Non-Bravais lattices can be represented as Bravais lattices with a basis consisting of more than one element. One example of this is graphene (see exercise).

2.2 Reciprocal lattice

We define the reciprocal lattice which is of importance for the electron band structure and x-ray diffraction on a periodic lattice. The reciprocal lattice is also perfectly periodic with a translation symmetry with a basic set $\{\boldsymbol{b}_i\}$ defining arbitrary reciprocal lattice vectors as

$$\boldsymbol{G}_m = m_1\boldsymbol{b}_1 + m_2\boldsymbol{b}_2 + m_3\boldsymbol{b}_3. \tag{2.5}$$

Above, m_i are integers and

$$\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij},\ i,j = 1,2,3. \tag{2.6}$$

while we have

$$\boldsymbol{b}_i = 2\pi\frac{\boldsymbol{a}_j \times \boldsymbol{a}_k}{\boldsymbol{a}_i \cdot (\boldsymbol{a}_j \times \boldsymbol{a}_k)},\ \boldsymbol{a}_i = 2\pi\frac{\boldsymbol{b}_j \times \boldsymbol{b}_k}{\boldsymbol{b}_i \cdot (\boldsymbol{b}_j \times \boldsymbol{b}_k)}. \tag{2.7}$$

The reciprocal lattice can be equal to the real space lattice in the simplest cases, for instance for a square or cubic lattice. For more complicated lattices, this is not the case. For instance, a body

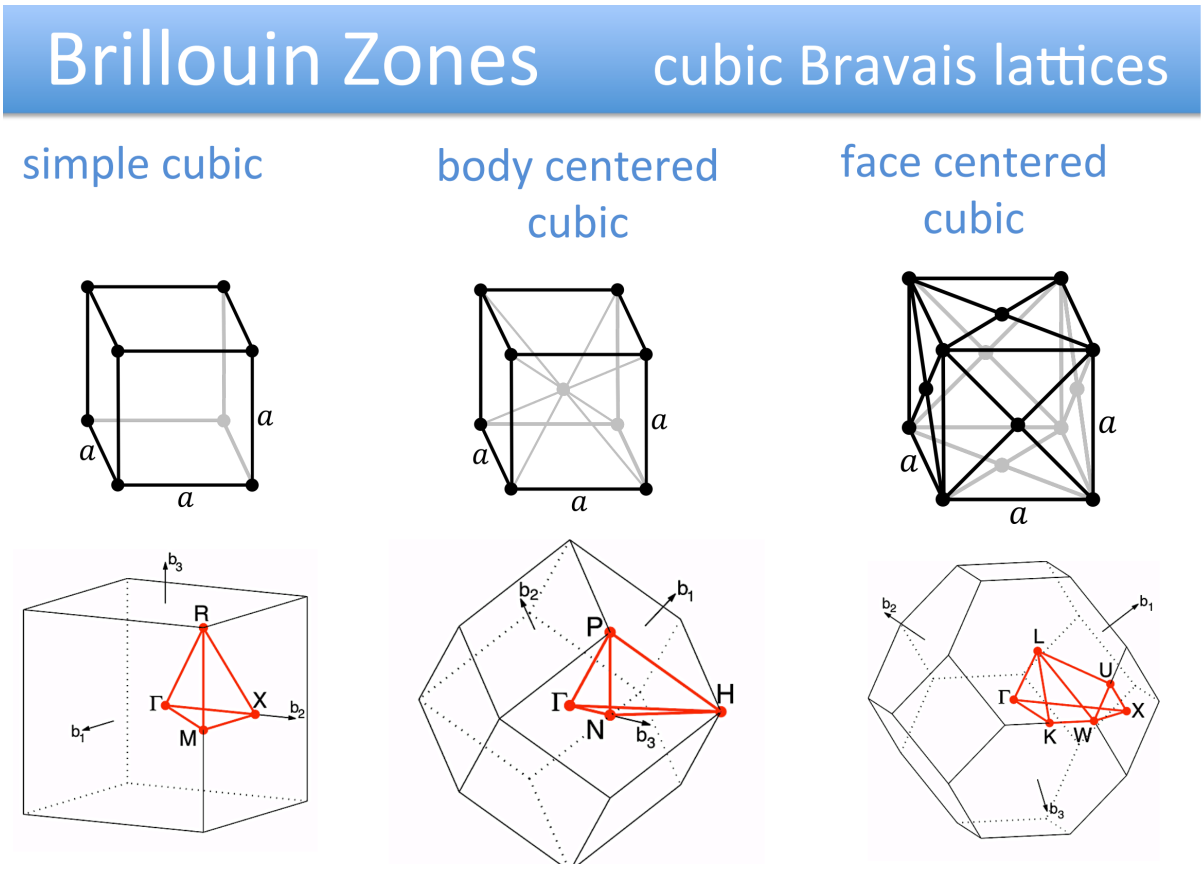

Figure 2.2: Real space unit cell and reciprocal space unit cells for cubic, bcc and fcc lattices. a is the lattice constant, whereas the points Γ, P, N, et.c. denote particular points in the BZ. These are points of high symmetry, called critical points. The center of the Brillouin zone is universally denoted Γ. The specific definition of the remaining points depends on the precise lattice structure. For instance, for a simple cube one has M: center of an edge, R: corner points, X: center of a face. For a body-centered cubic, H: corner point joining four edges, N: center of a face, P: corner point joining three edges. For a face-centered cubic one has K: middle of an edge joining two hexagonal faces, L: center of a hexagonal face, U: middle of an edge joining a hexagonal and a square face, W: corner point, X: center of a square face. Figure taken from [3].

centered cubic (bcc) lattice has a face centered cubic (fcc) reciprocal lattice and vice versa, as shown in Fig. 2.2. The Wigner-Seitz cell (unit cell) in reciprocal space is called the first Brillouin zone. From the above relations, it follows that any real space lattice vector $\boldsymbol{R}_n$ and any reciprocal lattice vector $\boldsymbol{G}_m$ have the property that

$$\boldsymbol{G}_m \cdot \boldsymbol{R}_n = 2\pi(m_1 n_1 + m_2 n_2 + m_3 n_3) = 2\pi N \tag{2.8}$$

where N is an integer. This permits us to expand any function $f(\boldsymbol{r})$ which is periodic in the real lattice as a Fourier series:

$$f(\boldsymbol{r}) = \sum_{\boldsymbol{G}} f_{\boldsymbol{G}} \mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{2.9}$$

since this ensures that the periodicity property $f(\boldsymbol{r}+\boldsymbol{R}_n)=f(\boldsymbol{r})$ holds. The coefficients in this expansion may be obtained via

$$f_{\boldsymbol{G}}=\frac{1}{\Omega_{\mathrm{UC}}}\int_{\mathrm{UC}}d^3r f(\boldsymbol{r})\mathrm{e}^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \tag{2.10}$$

where the integral runs over the unit cell of the real-space lattice with the volume Ω_{UC}.

2.3 Bloch's theorem and Bloch functions

We consider a Hamiltonian $\mathcal{H}$ of electrons invariant under a discrete set of lattice translations, meaning that the potential felt by the electrons must be periodic. Because of this symmetry, the translation operator $\hat{T}_{\boldsymbol{a}}$ commutes with the Hamiltonian H, $[\hat{T}_{\boldsymbol{a}},H]=0$. The translation operator is defined via

$$\hat{T}_{\boldsymbol{a}}|\boldsymbol{r}\rangle=|\boldsymbol{r}+\boldsymbol{a}\rangle. \tag{2.11}$$

Neglecting electron-electron interactions, we are faced with a single-particle problem:

$$H\to H_0+\hat{p}^2/2m+V(\hat{\boldsymbol{r}}). \tag{2.12}$$

Here, $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{p}}$ are position and momentum operators while

$$V(\boldsymbol{r})=\sum_j V_{\mathrm{ion}}(\boldsymbol{r}-\boldsymbol{R}_j) \tag{2.13}$$

describes the potential landscape that the electrons move in via the ionic background, and $\boldsymbol{R}_j$ is the position of the j-th ion. The potential is periodic, $V(\boldsymbol{r}+\boldsymbol{a})=V(\boldsymbol{r})$. It is clear that H_0 commutes with $\hat{T}_{\boldsymbol{a}}$. Therefore, these operators have a common set of eigenstates. Bloch's theorem states that the eigenvalues of $\hat{T}_{\boldsymbol{a}}$ lie on the unit circle in the complex plane, which makes its eigenstates extended as we'll see below. We have

$$\hat{T}_{\boldsymbol{a}}\psi(\boldsymbol{r})=\psi(\boldsymbol{r}-\boldsymbol{a})=\lambda_{\boldsymbol{a}}\psi(\boldsymbol{r}),\ \hat{T}_{l\boldsymbol{a}}\psi(\boldsymbol{r})=\psi(\boldsymbol{r}-l\boldsymbol{a})=\hat{T}^l_{\boldsymbol{a}}\psi(\boldsymbol{r})=\lambda^l_{\boldsymbol{a}}\psi(\boldsymbol{r}). \tag{2.14}$$

Above, l is an integer. On physical grounds, the probability density for electrons should be equal on each lattice sites, since the system is periodic. This gives

$$|\psi(\boldsymbol{r})|^2=|\psi(\boldsymbol{r}+l\boldsymbol{a})|^2=|\lambda^l_{\boldsymbol{a}}|^2|\psi(\boldsymbol{r})|^2, \tag{2.15}$$

requiring that $|\lambda_{\boldsymbol{a}}|=1$, so that $\lambda_{\boldsymbol{a}}=\mathrm{e}^{i\phi_{\boldsymbol{a}}}$. The eigenstate $\psi(\boldsymbol{r})$ satisfying the above equations is a product of an extended plane-wave $\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ and a Bloch function $u_{\boldsymbol{k}}(\boldsymbol{r})$ which must have the same periodicity as the lattice:

$$\psi(\boldsymbol{r})=\psi_{n,\boldsymbol{k}}(\boldsymbol{r})=\frac{1}{\sqrt{\Omega}}\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n,\boldsymbol{k}}(\boldsymbol{r}). \tag{2.16}$$

Thus, we have the relations

$$\hat{T}_{\boldsymbol{a}}u_{n,\boldsymbol{k}}(\boldsymbol{r})=u_{n,\boldsymbol{k}}(\boldsymbol{r}),\hat{T}_{\boldsymbol{a}}\psi_{n,\boldsymbol{k}}(\boldsymbol{r})=\psi_{n,\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{a})=\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{a}}\psi_{n,\boldsymbol{k}}(\boldsymbol{r}). \tag{2.17}$$

These states are also eigenstates of H_0, so we have $H_0\psi_{n,\boldsymbol{k}}(\boldsymbol{r})=\epsilon_{n,\boldsymbol{k}}\psi_{n,\boldsymbol{k}}(\boldsymbol{r})$. The integer n is a quantum number called the band index, $\boldsymbol{k}$ is the pseudo-momentum (also known as crystal-momentum or

wavevector) and Ω is the volume of the system. An important property of the eigenvalue of $\psi_{n,\boldsymbol{k}}(\boldsymbol{r})$ with respect to $\hat{T}_{\boldsymbol{a}}$, $\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{a}}$, is that it is periodic in reciprocal space. This is because $\mathrm{e}^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{a}} = \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{a}}$ for all reciprocal lattice vectors $\boldsymbol{G}$. Therefore, it is sufficient to consider values of $\boldsymbol{k}$ inside the first Brillouin zone. This means that any $\boldsymbol{k}$ outside the first Brillouin zone can be mapped back into it, making the first Brillouin zone sufficient to describe the electronic states.

Bloch's theorem then simplifies the initial problem to the so-called Bloch equation for the periodic function $u_{\boldsymbol{k}}$:

$$\Big(\frac{(\hat{\boldsymbol{p}}+\hbar\boldsymbol{k})^2}{2m} + V(\hat{\boldsymbol{r}})\Big)u_{\boldsymbol{k}}(\boldsymbol{r}) = \epsilon_{\boldsymbol{k}} u_{\boldsymbol{k}}(\boldsymbol{r}). \tag{2.18}$$

We suppressed the band index n to simplify the notation. There exists different numerical methods which allows one to efficiently compute the band energies $\epsilon_{\boldsymbol{k}}$ for a given Hamiltonian H.

2.4 Nearly free electron approximation

We here provide a solution strategy to obtain the eigenvalues and eigenenergies when one can assume that the periodic potential $V(\boldsymbol{r})$ is weak. First, we expand the periodic potential as follows

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}} \mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{2.19}$$

$$V_{\boldsymbol{G}} = \frac{1}{\Omega_{\mathrm{UC}}}\int_{\mathrm{UC}} d^3r V(\boldsymbol{r})\mathrm{e}^{-i\boldsymbol{G}\cdot\boldsymbol{r}}. \tag{2.20}$$

The potential is real, and assuming that it is also invariant under inversion $V(\boldsymbol{r}) = V(-\boldsymbol{r})$, leading to $V_{\boldsymbol{G}} = V^*_{-\boldsymbol{G}}$. The uniform component $V_{\boldsymbol{G}=0}$ corresponds to an irrelevant energy shift and may be set to zero. Because of its periodicity, we can also express the Bloch function as

$$u_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} c_{\boldsymbol{G}} \mathrm{e}^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \tag{2.21}$$

where the coefficients $c_{\boldsymbol{G}} = c_{\boldsymbol{G}}(\boldsymbol{k})$ depend on $\boldsymbol{k}$ in general. Inserting the above expansions into the Bloch equation gives a linear eigenvalue problem for the band energies $\epsilon_{\boldsymbol{k}}$:

$$\Big(\frac{\hbar^2}{2m}(\boldsymbol{k}-\boldsymbol{G})^2 - \epsilon_{\boldsymbol{k}}\Big)c_{\boldsymbol{G}} + \sum_{\boldsymbol{G}'} V_{\boldsymbol{G}'-\boldsymbol{G}}\, c_{\boldsymbol{G}'} = 0. \tag{2.22}$$

This is an eigenvalue problem with $c_{\boldsymbol{G}}(\boldsymbol{k})$ as components of the eigenvectors and $\epsilon_{\boldsymbol{k}}$ as eigenvalues (band energies). The latter include now corrections to the bare parabolic dispersion, $\epsilon^{(0)}_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ due to the potential $V(\boldsymbol{r})$. The dispersion is parabolic in the absence of $V(\boldsymbol{r})$ since the eigenstates are simply plane waves.

As a lowest order approach for the energy spectrum of electrons in a crystal, we neglect the term with $V_{\boldsymbol{G}'-\boldsymbol{G}}$. The energy spectrum then consists of all parabolic bands of the type $\epsilon_{\boldsymbol{k}}(\boldsymbol{G}) = \hbar^2(\boldsymbol{k}-\boldsymbol{G})^2/2m$ centered around the reciprocal wavevectors $\boldsymbol{G}$ (see dashed lines in Fig. 2.3). This is an "empty lattice approximation" which introduces the effect of the periodicity of the lattice on the band structure, but with zero potential.

We then proceed to illustrate the nearly free electron method on a 1D lattice when including a non-zero magnitude of the potential. Assume that the periodic potential is weak in the sense that $|V_G| \ll \hbar^2 G^2/2m$ for the reciprocal lattice vectors G. Consider first the lowest energy eigenvalue close to the center of the Brillouin zone $k = 0$, i.e. $|k| \ll \pi/a$. The solution of Eq. (2.22) is then obtained by using the values for $c_{G'}$ in the second term of Eq. (2.22) corresponding to lowest energy eigenvalue for $V_G = 0$ (since the right-hand side will otherwise be of order $\mathcal{O}(V_G^2)$). The lowest energy eigenvalue at $k = 0$ occurs at $G = 0$ in the absence of a potential. The belonging coefficients for this case are $c^{(0)}_{G=0} = 1$ (since the solution for u is just a plane-wave: the normalization constant is already accounted for by a factor $1/\sqrt{\Omega}$ that appeared earlier) and $c^{(0)}_G = 0$ otherwise. Inserting this into the second term of Eq. (2.22) then gives $V_{-\boldsymbol{G}} = V_{\boldsymbol{G}}$. Since we set $V_{\boldsymbol{G}=0} = 0$ in the beginning of this section, we obtain the solution:

$$c_G \simeq \begin{cases} 1 \text{ for } G = 0 \\ -\frac{2mV_G}{\hbar^2[(k-G)^2-k^2]} \ll 1 \text{ for } G \neq 0 \end{cases} \tag{2.23}$$

which gives the energy eigenvalue for $G = 0$:

$$\epsilon_k \simeq \frac{\hbar^2 k^2}{2m} - \sum_{G\neq 0} \frac{|V_G|^2}{\frac{\hbar^2}{2m}[(k-G)^2 - k^2]} \simeq \frac{\hbar^2 k^2}{2m^*} + E_0 \tag{2.24}$$

where

$$E_0 = -\sum_{G\neq 0} |\lambda_G|^2 \frac{\hbar^2 G^2}{2m},\ \lambda_G = \frac{V_G}{\hbar^2 G^2/2m} \tag{2.25}$$

and the effective mass $m \to m^* > m$:

$$\frac{1}{m^*} = \frac{1}{m}\Big[1 - 4\sum_{G\neq 0} |\lambda_G|^2\Big]. \tag{2.26}$$

This solution corresponds to the lowest branch of the band structure within this approach (see Fig. 2.3). The parabolic approximation of the band structure at a symmetry point with an effective mass m^*, is a standard way to approximate band tops or bottoms. Note that we have actually provided what can be thought of as a low-energy effective theory for the band structure, since we found the solution for $|k| \ll \pi/a$ where the band energies in our case is minimal.

The band structure can be extended beyond the first Brillouin zone since it is periodic through the relation $\epsilon_{k+G} = \epsilon_k$. We have $\epsilon_k = \epsilon_{-k}$ due to parity (and time reversal symmetry), like for free electrons.

Let us now consider higher energy bands. We stay at the zone center $k = 0$ and address the eigenstates which have their dominant contribution associated with the parabolas centered at $G_\pm = \pm 2\pi/a \equiv \pm\tilde{G}$ (see Fig. 2.3). In effect, the coefficients $c_{\pm\tilde{G}}$ essentially determine these eigenstates. The belonging energy bands cross for $k = 0$ in the case without potential at a value $\hbar^2\tilde{G}^2/2m$. Restricting ourselves to these two G-components, we obtain a two-dimensional eigenvalue equation system

$$\begin{pmatrix} \frac{\hbar^2}{2m}(k-\tilde{G})^2 - \epsilon_k & V^*_{2\tilde{G}} \\ V_{2\tilde{G}} & \frac{\hbar^2}{2m}(k+\tilde{G})^2 - \epsilon_k \end{pmatrix} = 0. \tag{2.27}$$

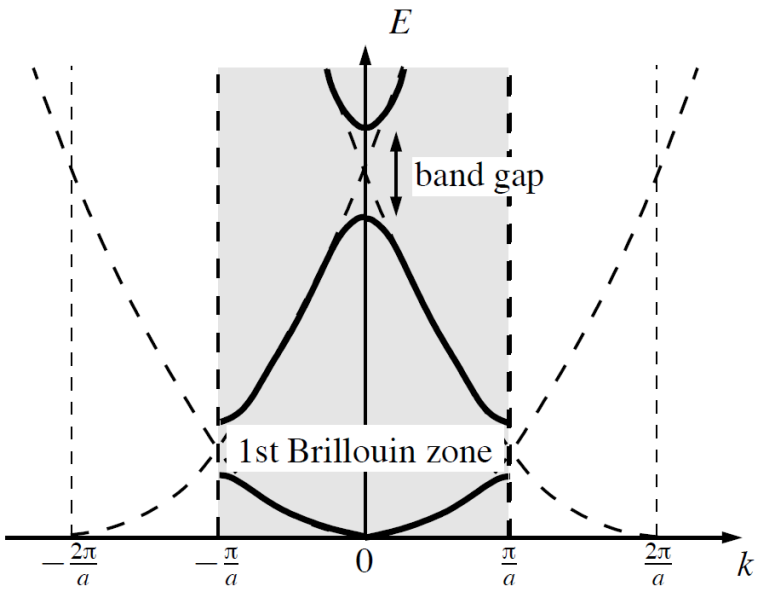

Figure 2.3: Band structure obtained by the nearly free electron approximation for a regular 1D lattice. The dashed lines are the allowed energy eigenvalues for zero potential, whereas the full lines are the solutions when a weak potential is taken into account. This leads to formation of band gaps. Figure taken from [3].

The eigenvalues are obtained by setting the determinant of the matrix equal to zero, as usual, and this gives two solutions:

$$\begin{aligned}\epsilon_{k\pm} &= \frac{1}{2}\Big[\frac{\hbar^2}{2m}\Big\{(k+\tilde{G})^2+(k-\tilde{G})^2\Big\} \pm \sqrt{\Big(\frac{\hbar^2}{2m}\Big\{(k+\tilde{G})^2-(k-\tilde{G})^2\Big\}\Big)^2+4|V_{2\tilde{G}}|^2}\Big] \\ &= \frac{\hbar^2}{2m}\tilde{G}^2 \pm |V_{2\tilde{G}}| + \frac{\hbar^2}{2m^*_\pm}k^2 \end{aligned} \tag{2.28}$$

with the effective mass

$$\frac{1}{m^*_\pm} \simeq \frac{1}{m}(1 \pm 2|\lambda_{\tilde{G}}|^{-1}). \tag{2.29}$$

Thus, we see that $m^*_+ > 0$ and $m^*_- < 0$ as $|\lambda_{\tilde{G}}| \ll 1$. This gives rise to an energy band gap separating two bands with opposite curvature, as shown in Fig. 2.3.

2.5 $\boldsymbol{k}\cdot\boldsymbol{p}$ approximation

Above, the periodic potential V was considered the perturbation. A different method called the $\boldsymbol{k}\cdot\boldsymbol{p}$-approximation ("k-dot-p") utilizes a different perturbation to approximate the band energies. It goes as follows. Consider the Bloch equation Eq. (2.18) which has the Hamiltonian

$$H = H_0 + H'_{\boldsymbol{k}} \tag{2.30}$$

where we let

$$H_0 = \frac{\hat{p}^2}{2m} + V(\hat{\boldsymbol{r}}),\ H'_{\boldsymbol{k}} = \frac{\hbar^2 k^2}{2m} + \frac{\hbar \boldsymbol{k}\cdot\boldsymbol{p}}{m}. \tag{2.31}$$

The unperturbed Hamiltonian is H_0, which includes the periodic potential, which is in fact exact at the Γ point $\boldsymbol{k}=0$. The perturbation is $H'_{\boldsymbol{k}}$. The result of this perturbation analysis is an expression for $\epsilon_{n,\boldsymbol{k}}$ and $u_{n,\boldsymbol{k}}$ in terms of the energies and wavefunctions at $\boldsymbol{k}=0$. Note that the perturbation term gets progressively smaller as $\boldsymbol{k}$ approaches zero. Therefore, $\boldsymbol{k}\cdot\boldsymbol{p}$ perturbation theory is most accurate for small values of $\boldsymbol{k}$.

Standard perturbation theory provides the following expression to lowest non-trivial order

$$\begin{aligned} u_{n,\boldsymbol{k}} &= u_{n,0} + \frac{\hbar}{m}\sum_{n'\neq n} \frac{\langle u_{n',0}|\boldsymbol{k}\cdot\boldsymbol{p}|u_{n,0}\rangle}{E_{n,0}-E_{n',0}} u_{n',0}, \\ E_{n,\boldsymbol{k}} &= E_{n,0} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m}\sum_{n'\neq n} \frac{|\langle u_{n,0}|\boldsymbol{k}\cdot\boldsymbol{p}|u_{n',0}\rangle|^2}{E_{n,0}-E_{n',0}}. \end{aligned} \tag{2.32}$$

The elements $\langle u_{n,0}|\boldsymbol{p}|u_{n',0}\rangle$ are known as optical matrix elements (note how $\boldsymbol{k}$ can be pulled outside of the expectation value since it is not an operator).

2.6 Tight-binding approximation

We consider now a regular lattice of atoms which are well separated such that their atomic orbitals have small overlaps only. Therefore, in a good approximation the electronic states are rather well represented by localized atomic orbitals, $\phi_n(\boldsymbol{r})$. The discrete spectrum of the atoms is obtained by solving the atomic Schrödinger equation

$$H_a(\boldsymbol{r},\boldsymbol{R})\phi_n(\boldsymbol{r}-\boldsymbol{R}) = \epsilon_n \phi_n(\boldsymbol{r}-\boldsymbol{R}) \tag{2.33}$$

for the wavefunction of an atom located at position $\boldsymbol{R}$, so that

$$H_a(\boldsymbol{r},\boldsymbol{R}) = \frac{\hat{\boldsymbol{p}}^2}{2m} + V_a(\boldsymbol{r}-\boldsymbol{R}) \tag{2.34}$$

where $V_a(\boldsymbol{r})$ is the atomic potential shown in Fig. 2.4(a). We let n denote all relevant quantum numbers, such as principal quantum number and angular momentum (l,m) and spin. We suppress the $\boldsymbol{r}$ dependence of H_a for brevity of notation in what follows. The total Hamiltonian for a singe particle combines all the potentials of the atoms

$$H = \frac{\hat{\boldsymbol{p}}^2}{2m} + \sum_{\boldsymbol{R}_j} V_a(\boldsymbol{r}-\boldsymbol{R}_j) = H_a(\boldsymbol{R}_j) + \Delta V_{\boldsymbol{R}_j}(\boldsymbol{r}), \tag{2.35}$$

where we singled out one atomic potential (the choice of $\boldsymbol{R}_j$ is arbitrary) and introduce the correction

$$\Delta V_{\boldsymbol{R}_j}(\boldsymbol{r}) = \sum_{\boldsymbol{R}_{j'}\neq\boldsymbol{R}_j} V_a(\boldsymbol{r}-\boldsymbol{R}_{j'}). \tag{2.36}$$

The eigenstates of the above total Hamiltonian will be either localized atomic orbitals or extended states, depending on their energy, as shown in Fig. 2.4(b).

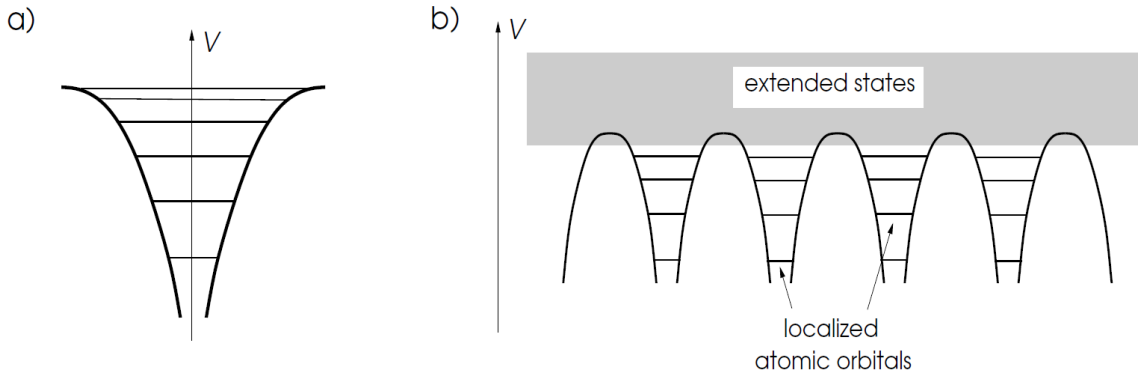

Figure 2.4: Potential landscape for **a)** a single atomic Coulomb potential yields a discrete spectrum electronic states and **b)** atoms arranged in a regular lattice give rise to a periodic potential which close to the atom sites look much like the attractive Coulomb-like potential. Electron states of low energy can be considered as practically localized at the atom sites, as the extension of their wave functions is very small. The higher energy states, however, extend further and can delocalize to form itinerant electron states which form bands. Figure taken from [3].

2.6.1 Linear combination of atomic orbitals (LCAO)

We use here a linear combination of atomic orbitals (LCAO) to approximate the extended Bloch states

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \sqrt{\frac{1}{N}} \sum_{\boldsymbol{R}_j} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}_j} \phi_n(\boldsymbol{r} - \boldsymbol{R}_j), \tag{2.37}$$

where N denotes the number of lattice sites. This superposition has the properties of a Bloch function since it satisfies $\psi_{n\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{a}) = \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{a}}\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ for all lattice vectors $\boldsymbol{a}$.
First, let us determine the norm of the Bloch function:

$$\begin{aligned}
\int d^3r \psi_{n\boldsymbol{k}}(\boldsymbol{r})^* \psi_{n'\boldsymbol{k}}(\boldsymbol{r}) \equiv \langle 1 \rangle_{nn'}(\boldsymbol{k}) &= \frac{1}{N} \sum_{\boldsymbol{R}_j, \boldsymbol{R}_{j'}} \int d^3r \mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{R}_{j'} - \boldsymbol{R}_j)} \phi_n^*(\boldsymbol{r} - \boldsymbol{R}_j) \phi_{n'}(\boldsymbol{r} - \boldsymbol{R}_{j'}) \\
&= \sum_{\boldsymbol{R}_j} \int d^3r \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j} \phi_n^*(\boldsymbol{r} - \boldsymbol{R}_j) \phi_{n'}(\boldsymbol{r}) \\
&= \delta_{nn'} + \sum_{\boldsymbol{R}_j \neq 0} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j} \alpha_{nn'}(\boldsymbol{R}_j).
\end{aligned} \tag{2.38}$$

Here, we were able to set $\boldsymbol{R}_{j'} = 0$ and eliminate the sum over $\boldsymbol{R}_{j'}$ due to translational invariance in the lattice, and thus cancelling also the factor $1/N$. The energy can now be estimated by first computing:

$$\begin{aligned}
\langle H \rangle_{nn'}(\boldsymbol{k}) &= \frac{1}{N} \sum_{\boldsymbol{R}_j, \boldsymbol{R}_{j'}} \int d^3r \mathrm{e}^{i\boldsymbol{k}\cdot(\boldsymbol{R}_{j'} - \boldsymbol{R}_j)} \phi_n^*(\boldsymbol{r} - \boldsymbol{R}_j) [H_a(\boldsymbol{R}_{j'}) + \Delta V_{\boldsymbol{R}_{j'}}(\boldsymbol{r})] \phi_{n'}(\boldsymbol{r} - \boldsymbol{R}_{j'}) \\
&= E_{n'} \langle 1 \rangle_{nn'}(\boldsymbol{k}) + \Delta E_{nn'} + \sum_{\boldsymbol{R}_j \neq 0} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j} \gamma_{nn'}(\boldsymbol{R}_j)
\end{aligned} \tag{2.39}$$

where

$$\Delta E_{nn'} = \int d^3r \phi_n^*(\boldsymbol{r}) \Delta V_{\boldsymbol{R}_{j'}=0}(\boldsymbol{r}) \phi_{n'}(\boldsymbol{r}) \tag{2.40}$$

and

$$\gamma_{nn'}(\boldsymbol{R}_j) = \int d^3r \phi_n^*(\boldsymbol{r} - \boldsymbol{R}_j) \Delta V_{\boldsymbol{R}_{j'}=0}(\boldsymbol{r}) \phi_{n'}(\boldsymbol{r}). \tag{2.41}$$

Now, the band energies may be computed through the secular equation

$$\det\Big[\langle H\rangle_{nn'}(\boldsymbol{k}) - \epsilon_{\boldsymbol{k}} \langle 1\rangle_{nn'}(\boldsymbol{k})\Big] = 0. \tag{2.42}$$

The merit of the approach is that the tightly bound atomic orbitals have only weak overlap such that both $\alpha_{nn'}(\boldsymbol{R}_j)$ and $\gamma_{nn'}(\boldsymbol{R}_j)$ fall off very quickly with growing $\boldsymbol{R}_j$. Mostly it is sufficient to take $\boldsymbol{R}_j$ connecting nearest-neighbor and sometimes next-nearest-neighbor lattice sites. This is for example fine for bands derived from 3d-orbitals among the transition metals such as Mn, Fe or Co etc. Also transition metal oxides are well represented in the tight-binding formulation. Alkali metals in the first row of the periodic table, Li, Na, K etc. are not suitable because their outermost s-orbitals have generally a large overlap. Note that the construction of the Hamiltonian matrix ensures that $\boldsymbol{k} \to \boldsymbol{k} + \boldsymbol{G}$ does not change $\epsilon_{\boldsymbol{k}}$ if $\boldsymbol{G}$ is a reciprocal lattice vector.

2.6.2 Band structure of s-orbitals

Let us apply this framework to the simplest case of a non-degenerate atomic orbital with s-wave symmetry, meaning it is rotationally invariant and has vanishing angular momentum $l = 0$. The rotation symmetry dictates that $\phi_s(\boldsymbol{r}) = \phi_s(|\boldsymbol{r}|)$, and the matrix elements only depend on the distance between the sites, $|\boldsymbol{R}_j|$.

As an example, consider a simple cubic lattice with nearest-neighbor vectors $\boldsymbol{R}_j = \{\pm(a,0,0), \pm(0,a,0), \pm(0,0,a)\}$ and next-nearest-neighbor vectors $\boldsymbol{R}_j = \{(\pm a, \pm' a, 0), (\pm a, 0, \pm' a), (0, \pm a, \pm' a)\}$. We will here neglect the overlap integrals $\alpha_{ss}(\boldsymbol{R}_j)$ as they are not important for the essential feature of the band structure. We thus have

$$\gamma_{ss}(\boldsymbol{R}_j) = \begin{cases} -t \text{ if } \boldsymbol{R}_j \text{ connects nearest neighbors} \\ -t' \text{ if } \boldsymbol{R}_j \text{ connects next-nearest neighbors.} \end{cases} \tag{2.43}$$

This provides us with the band energy

$$\begin{aligned} \epsilon_{\boldsymbol{k}} &= E_s + \Delta E_s - t \sum_{\boldsymbol{R}_j}^{\text{n.n}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j} - t' \sum_{\boldsymbol{R}_j}^{\text{n.n.n}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j} \\ &= E_s + \Delta E_s - 2t[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \\ &= -4t'[\cos(k_x a)\cos(k_y a) + \cos(k_y a)\cos(k_z a) + \cos(k_z a)\cos(k_x a)] \end{aligned} \tag{2.44}$$

Note that $\Delta V_{\boldsymbol{R}_j(\boldsymbol{r})} \leq 0$ in most cases due to the attractive ioniic potentials, so that $t, t' > 0$. We thus obtain a single band from this s-orbital. This is shown in Fig. 2.5. The nomenclature Γ, X, R, M refers to specific high-symmetry points in the reciprocal lattice as described previously.

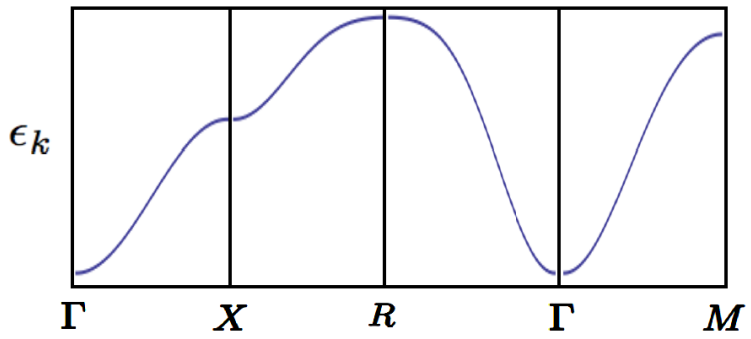

Figure 2.5: Band structure derived from atomic orbitals with s-wave symmetry. Here, $t' = 0.2t$ was chosen. Figure taken from [3].

We may obtain the effective mass of the electrons close to the $\boldsymbol{k} = 0$ point by expanding the cosines using $\cos x = 1 - x^2/2 + \ldots$, yielding

$$\epsilon_{\boldsymbol{k}} = E_s + \Delta E_s + 6t + 12t' + \frac{\hbar^2 \boldsymbol{k}^2}{2m^*} + \ldots \tag{2.45}$$

where

$$\frac{1}{m^*} = \frac{2}{\hbar^2}(t + 4t'). \tag{2.46}$$

Note that t and t' both diminish quickly when the lattice constant a increases due to the reduced overlap of atomic orbitals.

2.6.3 Wannier functions

An alternative approach to the tight-binding approximation is through Wannier functions. These are defined as the Fourier transformation of the Bloch wave functions,

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sqrt{1}\sqrt{N}\sum_{\boldsymbol{R}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}} w(\boldsymbol{r} - \boldsymbol{R}),\ w(\boldsymbol{r} - \boldsymbol{R}) = \sqrt{1}\sqrt{N}\sum_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \psi_{\boldsymbol{k}}(\boldsymbol{r}) \tag{2.47}$$

where the Wannier function $w(\boldsymbol{r} - \boldsymbol{R})$ is centered on the real-space lattice $\boldsymbol{R}$. We consider here a situation of a non-degenerate band analogous to the atomic s-orbital case, such that there is only one Wannier function per site. These functions obey the orthogonality relation

$$\begin{aligned}\int d^3r w^*(\boldsymbol{r} - \boldsymbol{R}')w(\boldsymbol{r} - \boldsymbol{R}) &= \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R} - i\boldsymbol{k}'\cdot\boldsymbol{R}'} \int d^3r \psi^*_{\boldsymbol{k}'}(\boldsymbol{r})\psi_{\boldsymbol{k}}(\boldsymbol{r}) \\ &= \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R} - i\boldsymbol{k}'\cdot\boldsymbol{R}'} \delta_{\boldsymbol{k}\boldsymbol{k}'} = \delta_{\boldsymbol{R}\boldsymbol{R}'}.\end{aligned} \tag{2.48}$$

The single-particle Hamiltonian has the usual form $H = -\hbar^2\nabla^2/2m + V(\boldsymbol{r})$, with a periodic potential $V(\boldsymbol{r})$. Then, we obtain the band energy

$$\begin{aligned}\epsilon_{\boldsymbol{k}} &= \int d^3r \psi^*_{\boldsymbol{k}}(\boldsymbol{r}) H \psi_{\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{N}\sum_{\boldsymbol{R},\boldsymbol{R}'} \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{R}' - \boldsymbol{R})} \int d^3r w^*(\boldsymbol{r} - \boldsymbol{R}')H w(\boldsymbol{r} - \boldsymbol{R}) \\ &= \sum_{\boldsymbol{R}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \int d^3r w^*(\boldsymbol{r} - \boldsymbol{R}) H w(\boldsymbol{r})\end{aligned} \tag{2.49}$$

where we used the translational invariance of the lattice. With the definitions

$$\epsilon_0 = \int d^3r w^*(\boldsymbol{r}) H w(\boldsymbol{r}),$$
$$t(\boldsymbol{R}) = \int d^3r w^*(\boldsymbol{r}-\boldsymbol{R}) H w(\boldsymbol{r}) \text{ for } \boldsymbol{R} \neq 0, \tag{2.50}$$

the band energy can be written as

$$\epsilon_{\boldsymbol{k}} = \epsilon_0 + \sum_{\boldsymbol{R}} t(\boldsymbol{R}) \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \tag{2.51}$$

This is the same type of tight-binding structure we derived above from the LCAO technique.
The Wannier function approach can be extended to the case of several bands, like p-orbital bands. In that case, one writes

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R},n'} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}} c_{nn'}(\boldsymbol{k}) w_{n'}(\boldsymbol{r}-\boldsymbol{R}) \tag{2.52}$$

where for all $\boldsymbol{k}$ we have

$$\sum_m c_{nm}(\boldsymbol{k}) c^*_{mn'} = \delta_{nn'}. \tag{2.53}$$

The matrix $c_{nn'}(\boldsymbol{k})$ performs a rotation from the band basis into the atomic orbital basis. For instance, in the p-orbital case with $l = 1$ it rotates from the energy bands into three Wannier functions with a p_x, p_y, and p_z-wave symmetry. The band energies can again be expressed in a tight-binding Hamiltonian form:

$$\epsilon_{m\boldsymbol{k}} = \sum_{\boldsymbol{R}} \sum_{nn'} c^*_{n'm}(\boldsymbol{k}) c_{mn}(\boldsymbol{k}) \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}} t_{n'n}(\boldsymbol{R}). \tag{2.54}$$

2.7 Second quantization

The tight-binding formulation of band electrons can also be implemented elegantly in a so-called second quantization language and provides a rather intuitive interpretation using creation and annihilation operators.

2.7.1 Introductory remarks and key results

First quantization

The original form of quantum mechanics where physical quantities such as position and momentum are converted into operators. A typical first quantized Hamiltonian could then look like:

$$\hat{H} = \sum_{i=1}^{N} \Big(-\frac{\hbar^2 \nabla_i^2}{2m} + U(\boldsymbol{r}_i) \Big) + \frac{1}{2} \sum_{i \neq j} V(\boldsymbol{r}_i - \boldsymbol{r}_j), \tag{2.55}$$

where U is the single-particle potential whereas V contains the interactions between particles at various positions.

Second quantization

Also known as *canonical quantization* in quantum field theory, one here upgrades the fields in the first quantization formalism to field operators. Second quantization is expressed via creation and annihilation operators in such a way that the matrix elements (expectation values of operators) remains the same both in the first and second quantized formalisms, as they have to since matrix elements may correspond to physical observables.

The main motivation for considering second quantization is that if we are working with an N-particle system where N is large, it is really inconvenient to work with huge wavefunctions which not only are products of each individual particle wavefunction, but which are linear combinations of such products (expressed as determinants) in order to satisfy *e.g.* the correct Pauli principle for fermions. Instead, it is far more convenient to work with the second quantized formalism which we explain below.

To be concrete, one converts from first to second quantization as follows. For single-particle operators, we have:

$$\hat{H} = \sum_{i=1}^{N} \hat{h}(\boldsymbol{r}_i) \rightarrow \hat{H} = \sum_{\alpha,\beta} \langle \alpha | \hat{h} | \beta \rangle c_\alpha^\dagger c_\beta, \tag{2.56}$$

where we defined the matrix element:

$$\langle \alpha | \hat{h} | \beta \rangle = \int d\boldsymbol{r} \phi_\alpha^*(\boldsymbol{r}) \hat{h}(\boldsymbol{r}) \phi_\beta(\boldsymbol{r}). \tag{2.57}$$

For two-particle operators, we have:

$$\hat{H} = \frac{1}{2} \sum_{i \neq j} \hat{v}(\boldsymbol{r}_i, \boldsymbol{r}_j) \rightarrow \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} \langle \alpha\beta | \hat{v} | \gamma\delta \rangle c_\alpha^\dagger c_\beta^\dagger c_\delta c_\gamma,$$
$$\text{where } \langle \alpha\beta | \hat{v} | \gamma\delta \rangle = \int \int d\boldsymbol{r} d\boldsymbol{r}' \phi_\alpha^*(\boldsymbol{r}) \phi_\beta^*(\boldsymbol{r}') \hat{v}(\boldsymbol{r}, \boldsymbol{r}') \phi_\gamma(\boldsymbol{r}) \phi_\delta(\boldsymbol{r}'). \tag{2.58}$$

This procedure holds for both fermions and bosons. The equivalence between the Hamiltonians may be verified by computing matrix elements $\langle \Psi_a | \hat{O} | \Psi_b \rangle$ for some operator $\hat{O}$ between any basis states $|\Psi_a\rangle$ and $|\Psi_b\rangle$ and see that the same element is obtained with both Hamiltonians. In the first quantization case, the basis states are Slater determinants (wavefunctions) whereas the basis states in the second quantization are of the type $|n\rangle = |n_1, n_2, \ldots\rangle$ (there is a unique state of this type for each Slater determinant).

Note that one can also use real-space *field operators* instead of creation and annihilation operators to formulate second quantization. In this case, one considers the field operators $\hat{\psi}(\boldsymbol{r})$ and $\hat{\psi}^\dagger(\boldsymbol{r})$ expressed via creation/annihilation operators in some other basis $\{|\gamma\}$ as:

$$\hat{\psi}_\sigma^\dagger(\boldsymbol{r}) = \sum_\gamma \langle \gamma | \boldsymbol{r}, \sigma \rangle c_\gamma^\dagger = \sum_\gamma \phi_{\gamma,\sigma}^*(\boldsymbol{r}) c_\gamma^\dagger,$$
$$\hat{\psi}_\sigma(\boldsymbol{r}) = \sum_\gamma \langle \boldsymbol{r}, \sigma | \gamma \rangle c_\gamma = \sum_\gamma \phi_{\gamma,\sigma}(\boldsymbol{r}) c_\gamma. \tag{2.59}$$

These field operators create and annihilate a particle at the position $\boldsymbol{r}$ with spin σ. The quantity $\phi_{\gamma,\sigma}$ is the single-particle wavefunction that emerges from the overlap between the single-particle

ket $|\gamma\rangle$ and the position-spin eigenstate $|\boldsymbol{r},\sigma\rangle$.

Let us now connect this to our earlier tight-binding framework. For simplicity, consider the single-orbital case and define the following fermionic operators:

$$\begin{aligned} &c_{js}^\dagger \text{ creates an electron of spin } s \text{ on lattice site } \boldsymbol{R}_j, \\ &c_{js} \text{ annihilates an electron of spin } s \text{ on lattice site } \boldsymbol{R}_j \end{aligned} \tag{2.60}$$

in the corresponding Wannier states. Our Hamiltonian then takes the form

$$H = \sum_{js} \epsilon_0 c_{js}^\dagger c_{js} + \sum_{ij} t_{ij} c_{is}^\dagger c_{js}, \tag{2.61}$$

where $t_{ij} = t_{ji}$ can be taken as real. The t_{ij}-coefficients are called hopping matrix elements, since $c_{is}^\dagger c_{js}$ annihilates an electron on site $\boldsymbol{R}_j$ and creates one on site $\boldsymbol{R}_i$. In this way, the electron moves (hops) from $\boldsymbol{R}_j$ to $\boldsymbol{R}_i$ and represents the kinetic energy of the electron. Since the Hamiltonian couples operators with different lattice indices, it is not diagonal. We can diagonalize it by the following Fourier transformation , which is equivalent to the transformation between Bloch and Wannier functions:

$$c_{js}^\dagger = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} a_{\boldsymbol{k}s}^\dagger \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j},\ c_{js} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} a_{\boldsymbol{k}s} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}_j}. \tag{2.62}$$

Here, $a_{\boldsymbol{k}s}^\dagger$ ($a_{\boldsymbol{k}s}$) creates (annihilates) an electron the Bloch state with pseudo-momentum $\boldsymbol{k}$ and spin s. Due to the Pauli principle, these fermion operators must satisfy anticommutation relations analogous to the commutation relations satisfied by the bosonic ladder operators for the harmonic oscillator:

$$\{c_{js}^\dagger, c_{j's'}\} = \delta_{jj'}\delta_{ss'},\ \{c_{js}^\dagger, c_{j's'}^\dagger\} = \{c_{js}, c_{j's'}\} = 0. \tag{2.63}$$

Inserting Eq. (2.62) in Eq. (2.61) gives

$$H = \sum_{\boldsymbol{k}\boldsymbol{k}'s} \Big[\frac{1}{N}\sum_i \epsilon_0 \mathrm{e}^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}_i} + \frac{1}{N}\sum_{ij} t_{ij}\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}_j - i\boldsymbol{k}'\cdot\boldsymbol{r}_u}\Big] a_{\boldsymbol{k}'s}^\dagger a_{\boldsymbol{k}s} = \sum_{\boldsymbol{k}s} \epsilon_{\boldsymbol{k}} a_{\boldsymbol{k}s}^\dagger a_{\boldsymbol{k}s}. \tag{2.64}$$

The combination $a_{\boldsymbol{k}s}^\dagger a_{\boldsymbol{k}s} = n_{\boldsymbol{k}s}$ is the number operator for electrons. The band energy is naturally the same as obtained above from the tight-binding approach. In what follows, we will mostly make use of the second quantized formulation of the Hamiltonian since it is highly suitable for dealing with many-body systems, as explained in detail in the following sections.

Additionally, the real-space formulation of the kinetic energy allows for the introduction of disorder which destroys periodicity (as is relevant for real experimental samples). This can be most straightforwardly implemented by site-dependent potentials $\epsilon_0 \to \epsilon_{0i}$ and to spatially (bond) dependent hopping matrix elements $t_{ij} = t(\boldsymbol{R}_i, \boldsymbol{R}_j) \neq t(\boldsymbol{R}_i - \boldsymbol{R}_j)$.

We now proceed to give a detailed derivation of the above results [6]. In terms of curriculum for this course, you are expected to be able to sketch the general strategy for deriving the second quantized expression, but not be able to reproduce all details in the derivation in the remainder of Sec. 2.7.1 from this point onward. The interested reader may of course delve into the full details of the derivation.

2.7.2 Systems of identical particles

Since quantum mechanics introduces the concept of *identical particles*, unlike classical physics where particle trajectories can in principle always be distinguished, we know that there is a constraint on the many-particle wavefunction describing a system of many identical particles (such as electrons). Let $\Psi(x_1, x_2, x_3, \ldots)$ be this wavefunction where we let $x_i = (\boldsymbol{r}_1, s_i)$ denote the position and spin coordinates of particle i. Since the probability density must be invariant under an exchange of two identical particles:

$$|\Psi(x_1, \ldots, x_j, \ldots, x_k, \ldots x_N)|^2 = |\Psi(x_1, \ldots, x_k, \ldots, x_j, \ldots x_N)|^2, \tag{2.65}$$

we are left with only two possibilities in 3D systems (we exclude here consideration of *anyons* which may exist in 2D):

$$\Psi \to +\Psi \text{ or } \Psi \to -\Psi. \tag{2.66}$$

This is the bosonic and fermionc case we have treated earlier, and this symmetry property has profound implications for the physical behavior of an ensemble of such particles at low temperatures. The energy distribution of such particles are in the non-interacting case described by the Bose-Einstein and Fermi-Dirac distribution, respectively.

2.7.3 Many-particle wavefunctions

Just like single-particle wavefunctions satisfying the SE are usually expanded in some complete basis, we can do the same for many-particle wavefunctions. An additional aspect to take into consideration compared to the single-particle case is, however, that the many-particle wavefunction basis set must have the appropriate symmetries under particle exchange as discussed above.

We start by considering a very general Hamiltonian of the type

$$\hat{H} = \hat{H}_0 + \hat{H}_I \tag{2.67}$$

where we defined

$$\hat{H}_0 = \sum_i \hat{h}(x_i), \; \hat{h}(x_i) = -\frac{\hbar^2}{2m}\nabla_i^2 + U(x_i) \tag{2.68}$$

and x_i is the coordinate and spin of particle i. This is a sum over the single-particle kinetic energy and potential energy $U(x_i)$ and an interaction part which will be a sum of terms that involve coordinates of multiple particles (such as Coulomb-interaction).

Our strategy will now be to set up a basis set where the many-particle wavefunctions in this basis set are eigenfunctions of a single-particle operator, such as $\hat{h}(x_i)$. Assume then that these are known:

$$\hat{h}(x)\phi_\nu(x) = \epsilon_\nu \phi_\nu(x). \tag{2.69}$$

The index ν describes a set of quantum number (assumed to be discrete for concreteness) which characterize the single-particle states ϕ_ν. These states are taken to form an orthonormal and complete set. We thus have for instance

$$\sum_\nu \phi_\nu^*(x)\phi_\nu(x') = \delta(x - x') = \delta(\boldsymbol{r} - \boldsymbol{r}')\delta_{s,s'}. \tag{2.70}$$

We will use the notation

$$\int dx \equiv \sum_s \int d\boldsymbol{r}. \tag{2.71}$$

Using these single-particle wavefunctions, we can now construct a many-particle wavefunction which is correctly symmetrized. For bosons, we have

$$\Phi^{\text{boson}}(x_1, \dots x_n) = \frac{1}{\sqrt{N!}\sqrt{\prod_\nu (n_\nu!)}} \sum_{P \in S_N} P\, \phi_{\nu_1}(x_1)\phi_{\nu_2}(x_2)\dots\phi_{\nu_N}(x_N). \tag{2.72}$$

The sum goes over all permutations P of the N coordinates $\{x_i\}$ while S_N is the set of these permutations. In total, there are $N!$ permutations. The prefactor nesures that the state is normalized to 1. Since bosons have no restriction regarding how many particles that can be in the same single-particle state, Φ^{boson} can in general be non-zero even if two indices ν_j are the same. For instance, if $\nu_7 = \nu_9$ there are two bosons in the single-particle state ν_7 so that $n_{\nu_7} = 2$.

For fermions, we instead have:

$$\Phi^{\text{fermion}}(x_1, \dots x_N) = \frac{1}{\sqrt{N!}} \sum_{P \in S_N} \text{sgn}(P) \cdot P\, \phi_{\nu_1}(x_1)\dots\phi_{\nu_N}(x_N). \tag{2.73}$$

The dependence on $n_\nu!$ is gone, compared to the bosonic case, since for fermions n_ν is 0 or 1, in which case $n_\nu! = 1$. Moreover, there is a sign factor accounting for the antisymmetric property under exchange of two fermions. The sign of P is defined in the following manner. A permutation has a positive sign if it can be arrived at by an even number of two-particle permutations. It has a negative sign if one requires an odd number of two-particle permutations.

As expected, we can also write the fermionic many-body wavefunction as a *Slater determinant*

$$\Phi^{\text{fermion}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{\nu_1}(x_1) & \dots & \phi_{\nu_1}(x_N) \\ \dots & \dots & \dots \\ \phi_{\nu_N}(x_1) & \dots & \phi_{\nu_N}(x_N) \end{vmatrix}. \tag{2.74}$$

The Pauli exclusion principle discussed earlier is manifested in that the determinant, and thus wavefunction, vanishes if $\nu_i = \nu_j$ for some $i \neq j$. It is not possible to put more than one fermion into a specific single-particle state.

Both in the fermionic and bosonic case, the many-particle wavefunctions $\Phi^{\text{fermion/boson}}$ are eigenfunctions of the non-interacting part $\hat{H}$ of our Hamiltonian. Their eigenvalues are

$$E = \sum_\nu \epsilon_\nu n_\nu \tag{2.75}$$

where n_ν is the number of particles that reside in quantum state ν. The point is now that we can use these wavefunctions Φ as a complete, orthonormal basis set for the wavefunctions Ψ that solve the *interacting* problem:

$$\Psi(x_1, \dots x_N) = \sum_a f_a \Phi_a(x_1, \dots x_N) \tag{2.76}$$

where f_a are expansion coefficients and the wavefunctions satisfy

$$\int d^N x \Phi_a^* \Phi_b = \delta_{a,b}. \tag{2.77}$$

2.7.4 Second quantization (*aka* occupation-number formalism)

The wavefunction formalism we have discussed in the preceeding sections is known as *first quantization*. Although we can in principle make use of the above treatment in order to deal with many-particle basis wavefunctions, there exists a more elegant and practical formalism. This is the *second quantization procedure* also known as the *occupation-number representation*. The essence of this formalism is that the many-body state can be described in a very succinct manner, as the relevant information consists of only two parts:

- The occupation number for each single-particle state (0 or 1 for fermions).
- The symmetry or antisymmetry of the state under an exchange of particles.

2.7.5 Fermionic case

To describe second quantization for fermions, we shall make use of creation and annihilation operators $c^\dagger$ and c which act on specified single-particle states. The antisymmetry propertey of fermions will then be taken care of via the anti-commutation relations that these operators satisfy.

The many-particle basis states that we used previously were Slater determinants. Now, we shall use products of creation operators acting on the "'vacuum"' state (with no fermions) as our many-particle basis state which then automatically has the correct anti-symmetry property under exchange of particles. Let us consider an example consisting of a 3-particle state where the single-particle states 2, 4, and 6 are occupied whereas the rest are empty. The analogue of the unique Slater determinant for this state (once we have agreed upon an ordering of the states 2,4,6) in the *occupation number representation* (rather than the position-space representation) is then:

$$|0,1,0,1,0,1,0,0,0,\ldots\rangle \tag{2.78}$$

which states that there 0 particles in single-particle state 1, 1 particle in state 2, and so forth. An arbitrary many-particle basis state can then be written as

$$|n_1,n_2,n_3,\ldots\rangle. \tag{2.79}$$

The creation operator $c_\nu^\dagger$, where ν labels the single-particle state, is defined as

$$c_\nu^\dagger|n_1,n_2,\ldots n_\nu,\ldots\rangle = (-1)^{\sum_{\mu<\nu} n_\mu}(1-n_\nu)|n_1,n_2,\ldots 1_\nu,\ldots\rangle. \tag{2.80}$$

Here, $\theta_\nu \equiv (-1)^{\sum_{\mu<\nu} n_\mu}$ is a phase factor equal to +1 or -1 which ensures the correct antisymmetric property of the state vector. To see this, we first realize that the phase factor depends on whether the total number of fermions in single-particle states with labels less than ν is even or odd. If we start with the vacuum state $|0\rangle = |0,0,0,\ldots\rangle$ and first create a fermion in state 1, we get

$$c_1^\dagger|0\rangle = |1,0,0,\ldots\rangle. \tag{2.81}$$

If we subsequently create a fermion in state 2, the final state is:

$$c_2^\dagger c_1^\dagger|0\rangle = -|1,1,0,\ldots\rangle \tag{2.82}$$

where the minus sign came from the phase factor $(-1)^{\sum_{\mu<\nu} n_\mu}$. But if we instead had created the fermions in the opposite order (first state 2, then state 1), we would have obtained

$$c_1^\dagger c_2^\dagger|0\rangle = |1,1,0,\ldots\rangle. \tag{2.83}$$

Thus, the creation operators *anticommutate* which ensures the antisymmetry property of the state vector: $\{c_i^\dagger, c_j^\dagger\} = 0$. An arbitrary many-particle basis state which has the correct antisymmetry property under exchange of particles may then generally be written as

$$|n_1, n_2, n_3, \ldots\rangle = (c_1^\dagger)^{n_1}(c_2^\dagger)^{n_2}\ldots|0\rangle. \tag{2.84}$$

Moreover, the factor $(1 - n_\nu)$ ensures that if a fermion already exists in state ν, then acting with $c_\nu^\dagger$ gives 0 due to the Pauli principle.

The *annihilation operator* is defined in a similar manner

$$c_\nu|n_1, n_2, \ldots n_\nu, \ldots\rangle = (-1)^{\sum_{\mu<\nu} n_\mu} n_\nu |n_1, \ldots 0_\nu, \ldots\rangle. \tag{2.85}$$

The result is non-zero only if there already existed a fermion in state ν due to the prefactor n_ν. This is reasonable as we cannot annihilate a fermion that does not exist in the first place.

Fock space

If we let $|\psi_N\rangle$ be a state with N fermions, then $c_\nu^\dagger|\psi_N\rangle$ and $c_n|\psi_N\rangle$ are states with $N+1$ and $N-1$ fermions, respectively. We can thus say that c_ν and $c_\nu^\dagger$ are mappings between fixed particle-number fermionic Hilbert spaces $\mathcal{H}_N$ whose particle numbers are different by 1:

$$\begin{aligned} c_\nu &: \mathcal{H}_N \to \mathcal{H}_{N-1}, \\ c_\nu^\dagger &: \mathcal{H}_N \to \mathcal{H}_{N+1}. \end{aligned} \tag{2.86}$$

These mappings hold for any value N and the collection (direct sum) of all these fixed particle-number fermionic Hilbert spaces $\{\mathcal{H}_N\}$ ($N = 0, 1, 2\ldots$) are referred to as *the Fock space* $\mathcal{F}$. It is therefore common to see many-particle states being referred to as Fock states in the literature and that they reside in Fock space. The operators c_ν and $c_\nu^\dagger$ are thus operators which map between states in Fock space.

From the way we defined the creation and annihilation operators in Eq. (2.80) and (2.85), we can prove that the operators satisfy the following anticommutation relations:

$$\begin{aligned} \{c_\mu, c_\nu\} &= 0, \\ \{c_\mu^\dagger, c_\nu^\dagger\} &= 0, \\ \{c_\mu^\dagger, c_\nu\} &= \delta_{\mu\nu}. \end{aligned} \tag{2.87}$$

Another useful and commonly encountered operator is the *number operator* $\hat{n}_\nu \equiv c_\nu^\dagger c_\nu$. To see why it earns its name, consider that

$$\begin{aligned} c_\nu^\dagger c_\nu|\ldots, n_\nu, \ldots\rangle &= \theta_\nu n_\nu c_\nu^\dagger|\ldots, 0_\nu, \ldots\rangle = \theta_\nu^2 n_\nu(1-0)|\ldots, 1_\nu, \ldots\rangle \\ &= n_\nu|\ldots, n_\nu, \ldots\rangle \end{aligned} \tag{2.88}$$

since $\theta_\nu^2 = 1$. Therefore, a basis state $|n\rangle$ is an eigenstate of $\hat{n}_\nu$ with eigenvalue n_ν: the operator $\hat{n}_\nu$ counts the number of fermions in the single-particle state ν. The total number operator is then:

$$\hat{N} = \sum_\nu \hat{n}_\nu \tag{2.89}$$

so that

$$\hat{N}|n\rangle = \sum_\nu n_\nu|n\rangle = N|n\rangle. \tag{2.90}$$

2.7.6 Second-quantized representation of operators

The next step is to figure out how operators in general are expressed in our newly formulated second-quantized language. We here confine our attention to the two most common and important types of operators: single-particle and two-particle operators.

Single-particle operators, such as the Hamiltonian $\hat{H}_0$ for a collection of free particles, can be written as a sum of terms and each term only involves the coordinates of a single particle. On the other hand, two-particle operators can be written as a sum of terms which each involve the coordinates of two particles. An example of the latter is the Coulomb interaction between electrons. We summarized the procedures for how to convert first-quantized operators to second-quantized operators initially in this chapter:

- Single-particle operators:

$$\hat{H}_0 = \sum_{i=1}^{N} \hat{h}(x_i) \Longrightarrow \sum_{\alpha\beta} \langle \alpha | \hat{h} | \beta \rangle c_\alpha^\dagger c_\beta \tag{2.91}$$

  where we defined

$$\langle \alpha | \hat{h} | \beta \rangle = \int dx \phi_\alpha^*(x) \hat{h}(x) \phi_\beta(x). \tag{2.92}$$

- Two-particle operators:

$$\hat{H}_I = \frac{1}{2} \sum_{i \neq j} \hat{v}(x_i, x_j) \Longrightarrow \frac{1}{2} \sum_{\alpha\beta\gamma\delta} \langle \alpha\beta | \hat{v} | \gamma\delta \rangle c_\alpha^\dagger c_\beta^\dagger c_\delta c_\gamma, \tag{2.93}$$

  where we defined

$$\langle \alpha\beta | \hat{v} | \gamma\delta \rangle = \int \int dx dx' \phi_\alpha^*(x) \phi_\beta^*(x') \hat{v}(x, x') \phi_\gamma(x) \phi_\delta(x'). \tag{2.94}$$

However, we not only wish to state this, but let us now *prove this*. The equivalence of the first- and second-quantized formalisms can be proven by considering matrix elements since these correspond to physically observable quantities (expectation values). The key point is that matrix elements of the type

$$\langle \Phi_a | \hat{O} | \Phi_b \rangle \tag{2.95}$$

for an operator $\hat{O}$ and two basis states $|\Phi_a\rangle$ and $|\Phi_b\rangle$ should yield the same value *regardless of which formalism (first or second quantization)* that is used to calculate it. We know that in the first quantized formalism, the basi states are Slater determinant wavefunctions. In the second quantized formalism, the basis states are the Fock states $|n\rangle = |n_1, n_2, \ldots\rangle$. There exists a unique correspondence between each Fock state and each Slater determinant.

It is clear from the outset that a matrix element involving a single-particle operator will be zero if the two Slater determinants differ in the occupation of more than two single-particle states (since such states are orthogonal). This can be seen easily in the second quantized version, where a single-particle operator is in general a sum of terms which each only changes (at most) the occupation of two single-particle states. For instance, $c_\alpha^\dagger c_\beta$ transfers a fermion from state β to state α. In

the same way, a matrix element for two-particle operators if the occupation between the two states differ with more than four single-particle states.

Before proving the above correspondence between first- and second-quantized operators, let us briefly go through the second quantization procedure for *bosons*.

2.7.7 Second quantization for bosons

Whereas the fundamental (anti)commutation relations for fermions were discussed above, bosonic operators b satisfy $[b, b^\dagger] = 1$. In the occupation-number representation, many-particle basis states for bosons read

$$|n\rangle = |n_1, n_2, \ldots\rangle. \tag{2.96}$$

Unlike fermions, bosons can have an occupancy of a particular single-particle state that exceeds 1. In fact, n_ν can be any nonnegative integer. Just like there was a unique correspondence between fermionic Slater determinants and fermionic many-particle basis states in the occupation number representation, the same one-to-one correspondence exists between Eq. (2.72) and states of the type shown in Eq. (2.96). The bosonic creation and annihilation operators are defined by

$$\begin{aligned} b_\nu^\dagger|\ldots, n_\nu, \ldots\rangle &= \sqrt{n_\nu + 1}|\ldots, n_\nu + 1, \ldots\rangle, \\ b_\nu|\ldots, n_\nu, \ldots\rangle &= \sqrt{n_\nu}|\ldots, n_\nu - 1, \ldots\rangle. \end{aligned} \tag{2.97}$$

The bosonic number operator for state ν is $\hat{n}_\nu = b_\nu^\dagger b_\nu$. An arbitrary and properly noramlized many-particle basis state can now be written as

$$|n_1, n_2, n_3, \ldots\rangle = \prod_\nu \frac{(b_\nu^\dagger)^{n_\nu}}{\sqrt{n_\nu!}}|0\rangle \tag{2.98}$$

where $|0\rangle$ as before is the vacuum state with no bosons. The operators satisfy

$$\begin{aligned} [b_\mu, b_\nu] &= 0, \\ [b_\mu^\dagger, b_\nu^\dagger] &= 0, \\ [b_\mu, b_\nu^\dagger] &= \delta_{\mu,\nu}. \end{aligned} \tag{2.99}$$

The conversion between first- and second-quantized single and two-particle operators for bosons is identical to that of fermions shown in the previous section.

2.7.8 Basis transformations

We need one final piece of the puzzle before we can prove the equivalence of first and second-quantization, namely how to do basis transformations. We have thus far only used a particular basis set $\{\phi_\alpha(x)\}$ of single-particle wavefunctions, but let us now investigate how to proceed if we wish to use some different single-particle basis set.

We let $|\alpha\rangle$ be the single-particle ket whose overlap with the position-spin eigenstate $|x\rangle$ is the the wavefunction $\phi_\alpha(x)$ we've been using:

$$\phi_\alpha(x) \equiv \langle x|\alpha\rangle. \tag{2.100}$$

Both in the bosonic and fermion case we may write

$$|\alpha\rangle = a_\alpha^\dagger |0\rangle \tag{2.101}$$

where $a_\alpha^\dagger$ is the creation operator for a particle in single-particle state α. Consider now a different single-particle basis set $\{\psi_{\tilde{\alpha}}(x)\}$ whose associated single-particle kets are denoted $|n\rangle$:

$$\psi_{\tilde{\alpha}}(x) = \langle x|\tilde{\alpha}\rangle, \ |\tilde{\alpha}\rangle \qquad\qquad = a_{\tilde{\alpha}}^\dagger |0\rangle. \tag{2.102}$$

Here, $a_{\tilde{\alpha}}^\dagger$ is the creation operator for a particle in the single-particle state characterized by quantum number(s) $\tilde{\alpha}$. For instance, $\{|\alpha\rangle\}$ could be the set of eigenstates for the kinetic energy operator $\hat{p}^2/(2m)$ whereas $\{|\tilde{\alpha}\rangle\}$ could be the set of eigenstates for the total Hamilton operator $\hat{p}^2/(2m) + U(\hat{x})$.

If we want to transform between the two single-particle basis sets, we make use of the completeness relation in single-particle Hilbert space:

$$1 = \sum_\alpha |\alpha\rangle\langle\alpha| = \sum_{\tilde{\alpha}} |\tilde{\alpha}\rangle\langle\tilde{\alpha}|. \tag{2.103}$$

It follows that

$$a_\alpha^\dagger |0\rangle = |\alpha\rangle = \sum_{\tilde{\alpha}} |\tilde{\alpha}\rangle\langle\tilde{\alpha}|\alpha\rangle = \sum_{\tilde{\alpha}} \langle\tilde{\alpha}|\alpha\rangle a_{\tilde{\alpha}}^\dagger |0\rangle. \tag{2.104}$$

By comparing the leftmost and rightmost side of the above equation, we see that

$$a_\alpha^\dagger = \sum_{\tilde{\alpha}} \langle\tilde{\alpha}|\alpha\rangle a_{\tilde{\alpha}}^\dagger. \tag{2.105}$$

Taking the adjoint produces the relation for how to transform annihilation operators:

$$a_\alpha = \sum_{\tilde{\alpha}} = \langle\alpha|\tilde{\alpha}\rangle a_{\tilde{\alpha}}. \tag{2.106}$$

Using these transformation rules, it is straightforward to show that the wavefunction $\phi_\alpha(x) = \langle x|\alpha\rangle$ transforms as:

$$\phi_\alpha(x) = \sum_{\tilde{\alpha}} \langle\tilde{\alpha}|\alpha\rangle \zeta_{\tilde{\alpha}}(x) \tag{2.107}$$

where $\zeta_{\tilde{\alpha}}(x) = \langle x|\tilde{\alpha}\rangle$. Such basis transformations as above are *unitary transformations*. This can be proven by defining the matrix D which have elements $D_{\tilde{\alpha}\alpha} \equiv \langle\tilde{\alpha}|\alpha\rangle$. If D is unitary, we must have $DD^\dagger = 1$ so that $(DD^\dagger)_{\alpha\beta} = \delta_{nn'}$. We obtain:

$$\begin{aligned}\delta_{\alpha\beta} = \langle\alpha|\beta\rangle &= \sum_\gamma \langle\alpha|\gamma\rangle\langle\gamma|\beta\rangle = \sum_\gamma \langle\alpha|\gamma\rangle\langle\beta|\gamma\rangle^* \\ &= \sum_\gamma D_{\alpha\gamma} D_{\beta\gamma}^* = \sum_\gamma D_{\alpha\gamma}(D^\dagger)_{\gamma\beta} = (DD^\dagger)_{\alpha\beta}.\end{aligned} \tag{2.108}$$

which shows the unitarity. It is also important to note that basis transformations preserve anti-commutation and commutation relations: the new creation/annihilation operators satisfy the same kind of relations as the ones in the old basis.

2.7.9 Proof of the second-quantized representation of single-particle operators

Finally, we are ready to give the proof for how to second-quantize single-particle operators. The proof for two-particle operators is performed analogously.

Consider an arbitrary single-particle operator

$$\hat{O} = \sum_{i=1}^{N} \hat{o}_i. \tag{2.109}$$

We choose the set $\{|\tilde{\alpha}\rangle\}$ to be the single-particle basis of eigenstates of $\hat{o}$, meaning that

$$\hat{o}|\tilde{\alpha}\rangle = o_{\tilde{\alpha}}|\tilde{\alpha}\rangle. \tag{2.110}$$

Thus, $o_{\tilde{\alpha}}$ is the eigenvalue of $\hat{o}$ associated with the eigenstate $|\tilde{\alpha}\rangle$.

Our first claim is that the second-quantized representation of $\hat{O}$ can be expressed as

$$\hat{O} = \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} \hat{n}_{\tilde{\alpha}} \tag{2.111}$$

where $\tilde{\alpha} = a^\dagger_{\tilde{\alpha}} a_{\tilde{\alpha}}$ is the number operator for the state $|\tilde{\alpha}\rangle$. To show this, we compute the matrix element of $\hat{O}$ in both the first- and second-quantized formalism and show that the result is identical. We use a many-particle basis that is built from single-particle basis $\{|\tilde{\alpha}\rangle\}$.

Consider the matrix element $\langle\Phi'|\hat{O}|\Phi\rangle$ where $|\Phi\rangle$ and $|\Phi'\rangle$ are two arbitrary basis states in this many-particle basis set. We obtain

$$\begin{aligned}
\langle\Phi'|\hat{O}|\Phi\rangle &= \langle n'_{\tilde{\alpha}_1}, n'_{\tilde{\alpha}_2}, \ldots | \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} \hat{n}_{\tilde{\alpha}} | n_{\tilde{\alpha}_1}, n_{\tilde{\alpha}_2}, \ldots \rangle \\
&= \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} n_{\tilde{\alpha}} \langle n'_{\tilde{\alpha}_1}, n'_{\tilde{\alpha}_2}, \ldots | n_{\tilde{\alpha}_1}, n_{\tilde{\alpha}_2}, \ldots \rangle \\
&= \delta_{\Phi,\Phi'} \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} n_{\tilde{\alpha}}.
\end{aligned} \tag{2.112}$$

If we instead compute the expectation value using first quantization, we obtain:

$$\begin{aligned}
\langle\Phi'|\hat{O}|\Phi\rangle &= \int d^N x \Phi'^*(x_1, \ldots, x_N) \sum_{i=1}^{N} \hat{o}(x_i) \Phi(x_1, \ldots, x_N) \\
&= \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} n_{\tilde{\alpha}} \int d^N x \Phi'^*(x_1, \ldots, x_N) \Phi(x_1, \ldots, x_N) \\
&= \delta_{\Phi,\Phi'} \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} n_{\tilde{\alpha}}.
\end{aligned} \tag{2.113}$$

where we made use of Eq. (2.75). This *proves that we get exactly the same result* when computing the expectation value both in first- and second-quantization for this particular basis choice.

We should in fairness state that we used a particlar single-particle basis here, namely that one that diagonalizes $\hat{o}$. The expression Eq. (2.111) is therefore not valid for an arbitrary basis. However,

we can start from that expression we can use the basis transformations considered in the previous subsection to derive an expression for the operator $\hat{O}$ that *is valid* in an arbitrary basis:

$$\begin{aligned}\hat{O} &= \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} a^\dagger_{\tilde{\alpha}} a_{\tilde{\alpha}} = \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} \Big(\sum_{\alpha} \langle \alpha | \tilde{\alpha} \rangle a^\dagger_\alpha \Big) \Big(\sum_{\beta} \langle \tilde{\alpha} | \beta \rangle a_\beta \Big) \\ &= \sum_{\alpha,\beta} \langle \alpha | \Big(\sum_{\tilde{\alpha}} |\tilde{\alpha}\rangle o_{\tilde{\alpha}} \langle \tilde{\alpha}| \Big) |\beta\rangle a^\dagger_\alpha a_\beta = \sum_{\alpha\beta} \langle \alpha | \hat{o} | \beta \rangle a^\dagger_\alpha a_\beta. \end{aligned} \tag{2.114}$$

Above, we used that

$$\hat{o} = \hat{o} \cdot 1 = \hat{o} \sum_{\tilde{\alpha}} |\tilde{\alpha}\rangle\langle\tilde{\alpha}| = \sum_{\tilde{\alpha}} o_{\tilde{\alpha}} |\tilde{\alpha}\rangle\langle\tilde{\alpha}| = \sum_{\tilde{\alpha}} |\tilde{\alpha}\rangle o_{\tilde{\alpha}} \langle\tilde{\alpha}|. \tag{2.115}$$

The matrix element $\langle \alpha | \hat{o} | \beta \rangle$ is

$$\langle \alpha | \hat{o} | \beta \rangle = \int\int dx dx' \langle \alpha | x \rangle \langle x | \hat{o} | x' \rangle \langle x' | \beta \rangle = \int dx \langle \alpha | x \rangle \hat{o}(x) \langle x | \beta \rangle = \int dx \phi^*_\alpha(x) \hat{o}(x) \phi_\beta(x), \tag{2.116}$$

where we used that $\langle x | \hat{o} | x' \rangle = \hat{o}(x)\delta(x - x')$. We know from the calculation above, using the $|\alpha\rangle$-basis, that $\hat{O}$ gives exactly the same expectation value as the first-quantized formalism, and therefore we have now derived a general prescription for how to second-quantize an operator in an arbitrary basis.

This framework provides a great flexibility in that any complete set of basis vectors can be used to second-quantize the Hamiltonian. The corresponding general eigenstates are then Fock-type state vectors which are characterized by the number of particles for each possible eigenvalue. However, it is worth noting that some basis choices are definitely more convenient than others in terms of evaluating what the expectation value $\langle \alpha | \hat{o} | \beta \rangle$ actually is. In the diagonal basis, as we saw initially, it is trivial. We can choose any other basis set, even a basis set which does not appear physically reasonable for the given system at hand, but the evaluation of this expectation value will then be much more cumbersome.

2.7.10 Field operators

Finally, we consider creation and annihilation operators in the single-particle basis consisting of the states $|x\rangle$ where $x = (\boldsymbol{r}, s)$ as before. In this basis, rather than the occupation number basis considered above, the creation and annihilation operators are referred to as *field operators* and denoted $\hat{\psi}^\dagger(x)$ and $\hat{\psi}(x)$. Note how we have put the state label as an argument rather than as a subscript (this is just notation). We know how to transform operators from one basis to another, so we can immediately write down

$$\begin{aligned}\hat{\psi}^\dagger(x) &= \sum_\alpha \langle \alpha | x \rangle a^\dagger_\alpha = \sum_\alpha \phi^*_\alpha(x) a^\dagger_\alpha, \\ \hat{\psi}(x) &= \sum_\alpha \langle x | \alpha \rangle a_\alpha = \sum_\alpha \phi_\alpha(x) a_\alpha. \end{aligned} \tag{2.117}$$

The field operators create and annihilate a particle at space-spin coordinate x and obey the proper (anti-)commutation relations:

$$\begin{aligned}[\hat{\psi}(x), \hat{\psi}(x')]_\mp = [\hat{\psi}^\dagger(x), \hat{\psi}^\dagger(x')]_\mp = 0, \\ [\hat{\psi}(x), \hat{\psi}^\dagger(x')]_\mp = \delta(x - x'). \end{aligned} \tag{2.118}$$

Here, $\mp$ refers to bosons/fermions. The above relations follow immediately from the commutation relations of the second quantized operators $\{a_\alpha, a_\alpha^\dagger\}$. The Dirac-delta function comes from $\langle x|x'\rangle$ which cannot be normalized to 1 since x is continuous. Using the $\{|x\rangle\}$ basis, the second-quantized representation of single- and two-particle operators is

$$\begin{aligned}\hat{H}_0 &= \int dx \hat{\psi}^\dagger(x)\hat{h}(x)\hat{\psi}(x),\\ \hat{H}_I &= \frac{1}{2}\int\int dx dx' \hat{\psi}^\dagger(x)\hat{\psi}^\dagger(x')\hat{v}(x,x')\hat{\psi}(x')\hat{\psi}(x).\end{aligned} \tag{2.119}$$

This can be verified by inserting Eq. (2.117) into Eq. (2.119) and observing that this reproduces the general second-quantized form of operators mentioned earlier in the occupation number basis. Alternatively, one can insert the inverse transformations

$$a_\alpha = \int dx \phi_\alpha^*(x)\psi(x),\ a_\alpha^\dagger = \int dx \phi_\alpha(x)\psi^\dagger(x) \tag{2.120}$$

into the previously obtained expression

$$H = \sum_{\alpha\beta}\langle\alpha|h|\beta\rangle a_\alpha^\dagger a_\beta \tag{2.121}$$

which gives

$$H = \int\int dx\ dx' \sum_{\alpha\beta}\langle\alpha|h|\beta\rangle\langle x|\alpha\rangle\psi^\dagger(x)\langle\beta|x'\rangle\psi(x'). \tag{2.122}$$

Now, all the expectation values are scalars and commute with the operators. Specifically, $\phi_\alpha(x)$ and $\psi^\dagger(x)$ commute just like a wavefunction $u_{\boldsymbol{k}}$ and operator $c_{\boldsymbol{k}}^\dagger$ commute since $c_{\boldsymbol{k}}^\dagger$ acts on the Fock states $|n_{\boldsymbol{k}}\rangle$ while $\psi^\dagger(x)$ acts on the states $|x\rangle$. Therefore, we move terms inside the integral to obtain:

$$H = \int\int dx\ dx' \sum_{\alpha\beta}\psi^\dagger(x)\langle\langle x|\alpha\rangle\alpha|h|\beta\rangle\langle\beta|x'\rangle\psi(x'). \tag{2.123}$$

Using now the completeness relations for $|\alpha\rangle$ and $|\beta\rangle$, we obtain

$$H = \int\int dx\ dx'\psi^\dagger(x)\langle x|h|x'\rangle\psi(x'). \tag{2.124}$$

But as shown *e.g.* in the QM II lecture notes, we have

$$\langle x|h|x'\rangle = \hat{h}(x')\delta(x-x'). \tag{2.125}$$

Therefore, we end up with

$$H = \int\int dx\psi^\dagger(x)\hat{h}(x)\psi(x). \tag{2.126}$$

2.8 Symmetry properties of the band structure

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

$$\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}}(\{g|\boldsymbol{a}\}^{-1}\boldsymbol{r}) = \psi_{\boldsymbol{k}}(g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}). \tag{2.127}$$

This relation is seen to hold by using the Dirac notation where we write for the Bloch state wavefunction with pseudo-momentum $\boldsymbol{k}$

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \langle \boldsymbol{r}|\psi_{\boldsymbol{k}}\rangle. \tag{2.128}$$

The action of the operator $\hat{S}$ on the state $|\boldsymbol{r}\rangle$ is given by

$$\hat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{r}\rangle = |g\boldsymbol{r} + \boldsymbol{a}\rangle \text{ and } \langle \boldsymbol{r}|\hat{S}_{\{g|\boldsymbol{a}\}} = \langle g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}|. \tag{2.129}$$

The latter relation follows by noting that since

$$\begin{aligned} 1 &= \langle g\boldsymbol{r} + \boldsymbol{a}|g\boldsymbol{r} + \boldsymbol{a}\rangle \\ &= \langle g\boldsymbol{r} + \boldsymbol{a}|g\boldsymbol{r} + \boldsymbol{a}\rangle^* \\ &= \langle g\boldsymbol{r} + \boldsymbol{a}|S_{\{g|\boldsymbol{a}\}}|\boldsymbol{r}\rangle^* \\ &= \langle \boldsymbol{r}|S^\dagger_{\{g|\boldsymbol{a}\}}|g\boldsymbol{r} + \boldsymbol{a}\rangle, \end{aligned} \tag{2.130}$$

we have $S^\dagger_{\{g|\boldsymbol{a}\}}|g\boldsymbol{r} + \boldsymbol{a}\rangle = |\boldsymbol{r}\rangle$. Taking the adjoint of this equation gives $\langle g\boldsymbol{r} + \boldsymbol{a}|S_{\{g|\boldsymbol{a}\}} = \langle \boldsymbol{r}|$, and so with $\boldsymbol{r} = g^{-1}\boldsymbol{x} - g^{-1}\boldsymbol{a}$ we find precisely $\langle \boldsymbol{x}|\hat{S}_{\{g|\boldsymbol{a}\}} = \langle g^{-1}\boldsymbol{x} - g^{-1}\boldsymbol{a}|$. Therefore, we are finally able to prove that:

$$\langle \boldsymbol{r}|\hat{S}_{\{g|\boldsymbol{a}\}} \cdot |\psi_{\boldsymbol{k}}\rangle = \langle g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}|\psi_{\boldsymbol{k}}\rangle = \psi_{\boldsymbol{k}}(g^{-1}\boldsymbol{r} - g^{-1}\boldsymbol{a}). \tag{2.131}$$

Because $\{g|\boldsymbol{a}\}$ belongs to the space group of the crystal, it must commute with the Hamilton-operator since the Hamilton-operator must have the same symmetries as the crystal structure it describes: $[\hat{S}_{\{g|\boldsymbol{a}\}}, H] = 0$. Applying now a pure translation $\hat{T}_{\boldsymbol{a}'} = \hat{S}_{\{E|\boldsymbol{a}\}}$ to this new wavefunction, we obtain using our previously mentioned associative multiplication rule Eq. (2.3) that:

$$\begin{aligned} \hat{T}_{\boldsymbol{a}'}\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) &= \hat{S}_{\{g|\boldsymbol{a}\}}\hat{T}_{g^{-1}\boldsymbol{a}'}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \hat{S}_{\{g|\boldsymbol{a}\}}\mathrm{e}^{-i\boldsymbol{k}\cdot(g^{-1}\boldsymbol{a}')}\psi_{\boldsymbol{k}}(\boldsymbol{r}) \\ &= \hat{S}_{\{g|\boldsymbol{a}\}}\mathrm{e}^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}\psi_{\boldsymbol{k}}(\boldsymbol{r}) \\ &= \mathrm{e}^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}). \end{aligned} \tag{2.132}$$

Thus, we find that $\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r})$ is an eigenfunction of $\hat{T}_{\boldsymbol{a}'}$ with eigenvalue $\mathrm{e}^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}'}$. According to the Bloch theorem, we know that our basis $\{\psi_{\boldsymbol{k}}\}$ are eigenstates of both $\hat{T}_{\boldsymbol{a}'}$ and H. Thus, the effect of a symmetry transformation $\{g|\boldsymbol{a}\}$ on the wavefunction $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ corresponds to a rotation from $\boldsymbol{k}$

to $g^{-1}\boldsymbol{k}$ and multiplication with a phase factor [1]:

$$\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \lambda_{\{g|\boldsymbol{a}\}}\psi_{g\boldsymbol{k}}(\boldsymbol{r}), \tag{2.134}$$

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

$$\hat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{k}\rangle = \lambda_{\{g|\boldsymbol{a}\}}|g\boldsymbol{k}\rangle \tag{2.135}$$

in Dirac notation. The point is now that it follows:

$$\epsilon_{g\boldsymbol{k}} = \langle g\boldsymbol{k}|H|g\boldsymbol{k}\rangle = \langle \boldsymbol{k}\hat{S}^{-1}_{\{g|\boldsymbol{a}\}}H\hat{S}_{\{g|\boldsymbol{a}\}}|\boldsymbol{k}\rangle = \langle \boldsymbol{k}|H|\boldsymbol{k}\rangle = \epsilon_{\boldsymbol{k}}. \tag{2.136}$$

Consequently, there is a set of equivalent points $g\boldsymbol{k}$ with the same band energy ($\rightarrow$ degeneracy) for each $\boldsymbol{k}$ in the Brillouin zone, as shown in Fig. 2.6.

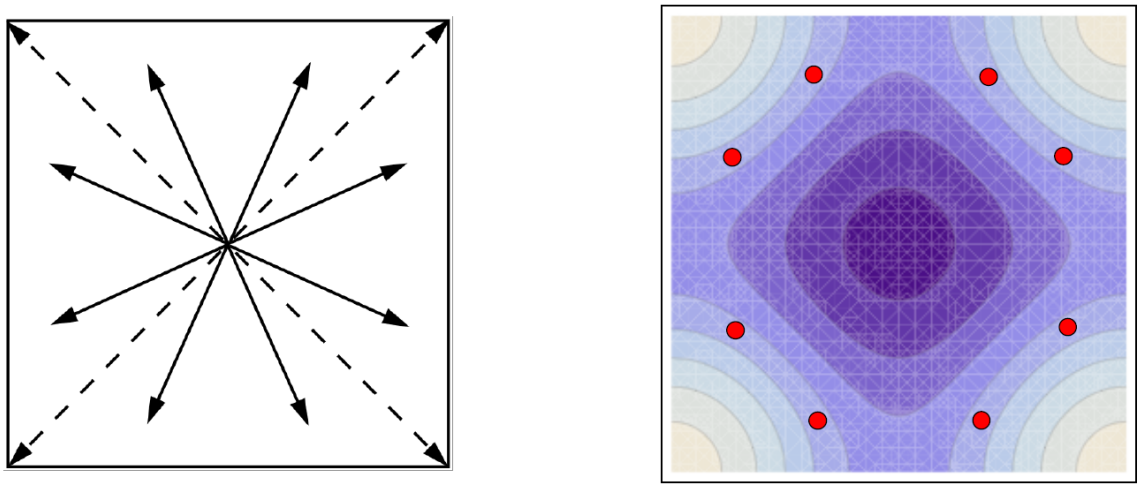

Figure 2.6: A "star" of $\boldsymbol{k}$-points in the Brillouin zone with degenerate band energies. **Left panel:** star of $\boldsymbol{k}$. **Right panel:** contour plot of a 2D band $\epsilon_{\boldsymbol{k}} = -2t[\cos(k_x a) + \cos(k_y a)] + 4t' \cos(k_x a)\cos(k_y a)$. The dots correspond to the star of $\boldsymbol{k}$ with degenerate energy values, demonstrating $\epsilon_{nk} = \epsilon_{n,g\boldsymbol{k}}$. Figure taken from [3].

The symmetry analysis of the band structure is useful to identify energy-degenerate points $\boldsymbol{k}$ in the BZ, which in the case of multiple bands can lead to band crossings. Changing the symmetry of the system, and thus the band structure, by external stimuli such as strain can thus change the band crossings and induce gaps in the band structure. This can have important consequences with regard to *e.g.* how conductive the material is.

[1] The detailed calculation is as follows

$$\begin{aligned}
\hat{S}_{\{g|\boldsymbol{a}\}}\psi_{\boldsymbol{k}}(\boldsymbol{r}) &= \frac{1}{\sqrt{\Omega}}\hat{S}_{\{g|\boldsymbol{a}\}}\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(\boldsymbol{k})\mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}} = \frac{1}{\sqrt{\Omega}}\mathrm{e}^{i\boldsymbol{k}\cdot(g^{-1}\boldsymbol{r}-g^{-1}\boldsymbol{a})}\sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(\boldsymbol{k})\mathrm{e}^{i\boldsymbol{G}\cdot(g^{-1}\boldsymbol{r}-g^{-1}\boldsymbol{a})} \\
&= \frac{1}{\sqrt{\Omega}}\mathrm{e}^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}}\mathrm{e}^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(\boldsymbol{k})\mathrm{e}^{i(g\boldsymbol{G})\cdot\boldsymbol{r}} = \mathrm{e}^{-i(g\boldsymbol{k})\cdot\boldsymbol{a}}\frac{1}{\sqrt{\Omega}}\mathrm{e}^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}} c_{g^{-1}\boldsymbol{G}}(\boldsymbol{k})\mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}} \\
&= \mathrm{e}^{-i(\hat{g}\boldsymbol{k})\cdot\boldsymbol{a}}\frac{1}{\sqrt{\Omega}}\mathrm{e}^{i(g\boldsymbol{k})\cdot\boldsymbol{r}}\sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(g\boldsymbol{k})\mathrm{e}^{i\boldsymbol{G}\cdot\boldsymbol{r}} = \lambda_{\{g|\boldsymbol{a}\}}\psi_{g\boldsymbol{k}}(\boldsymbol{r}),
\end{aligned} \tag{2.133}$$

where we used that $c_{\boldsymbol{G}} = c_{\boldsymbol{G}}(\boldsymbol{k})$ is a function of $\boldsymbol{k}$ with the property $c_{g^{-1}\boldsymbol{G}}(\boldsymbol{k}) = c_{\boldsymbol{G}}(g\boldsymbol{k})$, in effect $\hat{S}_{\{g|\boldsymbol{a}\}}u_{\boldsymbol{k}}(\boldsymbol{r}) = u_{g\boldsymbol{k}}(\boldsymbol{r})$.

2.9 Band-filling and materials properties

Due to the fermionic character of electrons each of the band states $|n, \boldsymbol{k}, s\rangle$ can be occupied with one electron taking also the spin quantum number into account with spin $s =\uparrow$ and $\downarrow$ (Pauli exclusion principle). The count of electrons has profound implications on the properties 23 of materials. Here we would like to look at the most simple classification of materials based on independent electrons.

2.9.1 Electron counting and band filling

We consider here the most simple band structure in the one-dimensional tight-binding model with nearest-neighbor coupling. The lattice has N sites (N even) and we assume periodic boundary conditions. The Hamiltonian is given by

$$H = -t \sum_{j=1}^{N} \sum_{s=\uparrow,\downarrow} (c^{\dagger}_{j+1,s} c_{js} + c^{\dagger}_{js} c_{j+1,s}) \tag{2.137}$$

Note that $H = H^{\dagger}$, as it should. This Hamiltonian is diagonalized by a Fourier transformation

$$c_{js} = \frac{1}{\sqrt{N}} \sum_{k} a_{ks} \mathrm{e}^{iR_j k} \tag{2.138}$$

which leads to

$$H = \sum_{ks} \epsilon_k a^{\dagger}_{ks} a_{ks} \text{ with } \epsilon_k = -2t \cos(ka). \tag{2.139}$$

To implement periodic boundary conditions, we impose the equivalence $j + N = j$ between lattice sites. Therefore, we demand that

$$\mathrm{e}^{iR_j k} = \mathrm{e}^{i(R_j+L)k} \rightarrow Lk = Nak = 2\pi n \rightarrow k = \frac{2\pi}{L} n = \frac{2\pi n}{Na}. \tag{2.140}$$

The pseudo-momentum $\boldsymbol{k}$ lies within the first Brillouin zone $-\pi/a < k \leq \pi/a$ and n is an integer. On the real-space lattice, an electron can take $2N$ different states: N difference lattice sites and a factor 2 for spin. Thus, for k we find that n should take the values $n = -N/2+1, -N/2+2, \ldots, N/2$ in order to preserve the number of degrees of freedom. Note that $k = -\pi/a$ and $k = \pi/a$ differ by a reciprocal lattice vector $G = 2\pi/a$ and are therefore identical: that is why $k = -\pi/a$ is excluded from the values k can take if $k = \pi/a$ is permitted.

We now fill these states with electrons following Pauli's exclusion principle. In Fig. 2.7, we show two typical situations. (1) N electrons correspond to half of the possible electrons which can be accommodated leading to a half-filled band. (2) $2N$ electrons exhausting all possible states representing a completely filled band. In the case of half-filling we define the Fermi energy as the energy ϵ_F of the highest occupied state (here $\epsilon_F = 0$). This corresponds to the chemical potential μ at $T = 0$: the energy necessary to add an electron to the system. At finite temperature, ϵ_F and μ are different: μ is the energy at which there is a 50% chance that an electron occupies the state, as given by the Fermi-Dirac distribution function.

An important difference between (1) and (2) is that the former allows for many different states which may be separated from each other by a very small energy. For example, considering the

ground state for $N = 8$ shown in case (1) in Fig. 2.7 and the excited states obtained by moving one electron from $k = \pi/2a$ to $k' = 2\pi(1 + N/4)/Na$ [corresponding to the quantized values $n = N/4$ and $n = 1 + N/4$], we find the energy difference

$$\Delta E = \epsilon_{k'} - \epsilon_k \simeq 2t\frac{2\pi}{N} \propto \frac{1}{N} \tag{2.141}$$

which shrinks to zero for $N \to \infty$. On the other hand, for case (2) there is only one electron configuration possible and no excitations within the one-band picture.

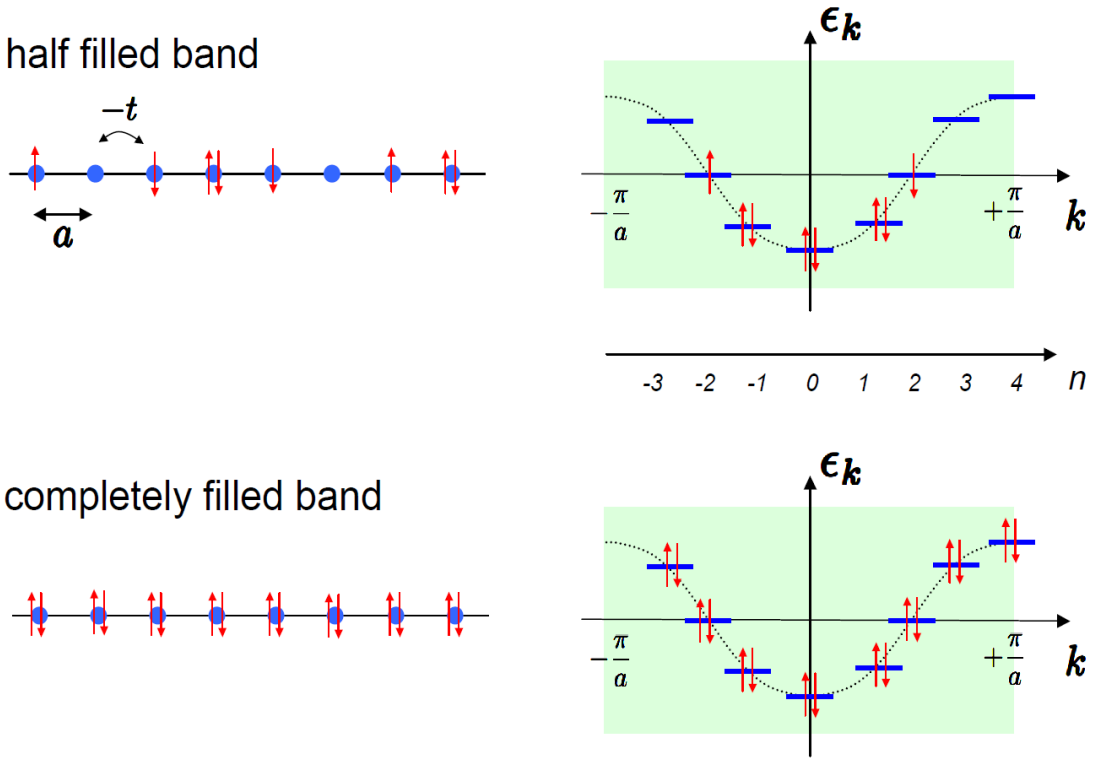

Figure 2.7: One-dimensional tight-binding model with $N = 8$ and periodic boundary conditions. The dispersion has eight different k-levels whereby it has to be noticed that π/a and $-\pi/a$ are equivalent. The condition of half-filling and complete filling are shown, where for half-filling a ground state configuration is shown (note there are 4 degenerate states). For the completely filled band all k levels are occupied by two electrons of opposite spin. This means in real space that also all sites are occupied by two electrons. This is a non-degenerate state. Figure taken from [3].

2.9.2 Metals, semiconductors, and insulators

The two situations depicted in Fig. 2.7 show each atom (site) in a lattice contributing an integer number of electrons to the system. We can distinguish the behavior of the conduction properties of the system when there is an odd or an even number of electrons per unit cell (note that for more complicated lattice geometries, the unit cell may contain more than one atom, unlike the situation

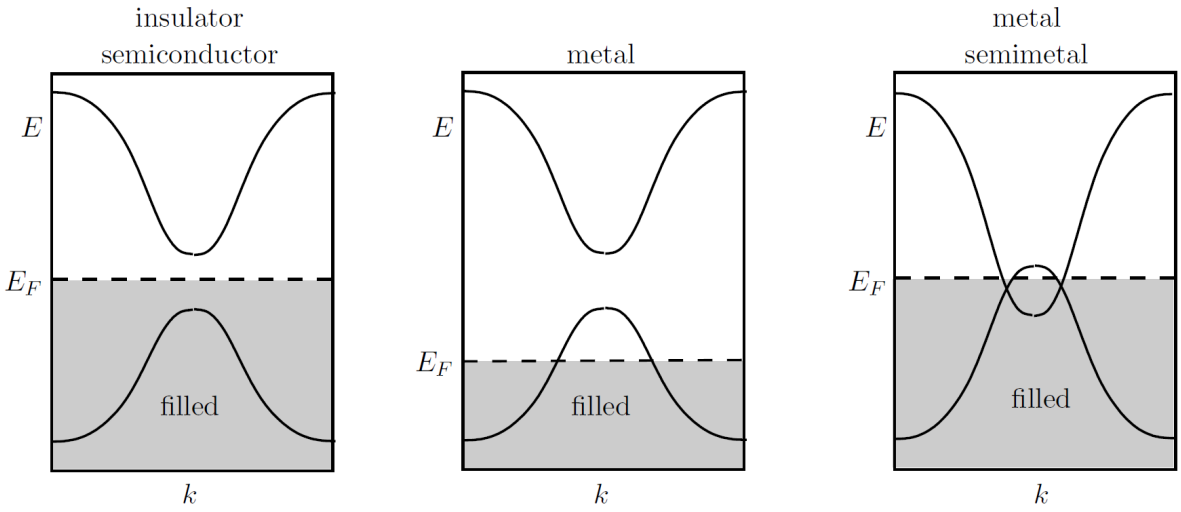

Figure 2.8: Material classes according to band filling. **Left panel:** insulator or semiconductor (partially filled bands with the Fermi level in band gap). **Center panel:** metal (Fermi level inside band). **Right panel:** metal or semimetal (Fermi level inside two overlapping bands). Figure taken from [3].

shown in our tight-binding example above). This is because the number of electrons per unit cell says something about whether an energy is fully or partly occupied.

- An important point is most easily illustrated using our 1D chain of equidistant atoms above. Due to conservation of degrees of freedom, there are N permitted k-values in the BZ if there are N atoms (unit cells). But there are $2N$ possible electron states due to spin. A single band therefore has room for 2 electrons per unit cell. If there is an even number of electrons per unit cell, the band will therefore be completely empty or full. If there is an odd number of electrons per unit cell, the band is partially filled. This means that it is possible for electrons to move in response to an external field since excitations (bumping electrons up to higher energy states) with arbitrarily small energy can be made. The following rule may then be formulated generally (which holds when electron-electron interactions can be neglected, as we have done here): *A material with an odd number of electrons per unit cell is a metal.*

- If the material has an even number of electrons per unit cell, it can be either a metal, insulator or semiconductor depending on the details of the band structure, in particular whether or not a gap is present between energy bands. In Fig. 3.2(a), the lower band is completely filled whereas the upper band is completely empty. The chemical potential then lies within the energy gap separating the highest filled and lowest empty band. If the band gap E_g is much smaller than the bandwidth (in practical terms, gaps smaller than 1 eV), we call the material a semiconductor. If E_g is comparable to the bandwidth (greater than 1 eV), it is a band insulator. We underline *band* insulator here, because we will later encounter another form of an insulator called Mott insulator whose insulating behavior is not governed by a band structure effect, but by a correlation effect through strong Coulomb interaction. For both a semiconductor and an insulator, at temperatures $T \ll E_g/k_B$ the the application of a small electric voltage will not produce an electric transport. The highest filled band is called valence

band, whereas the lowest empty band is termed conduction band. Examples for insulators are C as diamond and for semiconductors Si and Ge. They have diamond lattice structure with two atoms per unit cell. As these atoms belong to the group IV in the periodic table, each provides an even number of electrons suitable for completely filling bands.

- If the number of electrons per unit cell is odd, the uppermost non-empty band is half filled, see Fig. 3.2(b). Then the system is a metal, in which electrons can move and excitations with arbitrarily small energies are possible. The electrons remain mobile down to arbitrarily low temperatures. A standard example of a metal are the Alkali metals in the first column of the periodic table (Li, Na, K, Rb, Cs), as all of them have one mobile electron per ion.

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

2.9.3 Experimental measurement of electron energy-bands

Angle-resolved photoemission spectroscopy (ARPES) is an experimental technique used to measure the allowed energies and momenta of the electrons in a material. It is based on the photoelectric effect: an incoming photon of sufficient energy ejects an electron from the surface of a material. By directly measuring the kinetic energy and emission angle distributions of the emitted photoelectrons, the technique can subsequently determine the properties of the state in the material that the electron was ejected out of. In this way, one maps the electronic band structure and Fermi surfaces of the material. This technique is commonly used in modern research been used by physicists to investigate a wealth of different materials.

The energy of the incident photon needs to be higher than the binding energy of the electron in the solid in order to eject it. In the process, the electron's momentum remains virtually intact (at least if it is scattered from the surface), except for its component perpendicular to the material's surface. As the electron crosses the surface barrier, losing part of its energy due to the surface work function [2], only the component of momentum that is parallel to the surface is preserved. This is consistent with the absence of translational invariance in the direction $\perp$ to the surface.

By measuring the freed electron's kinetic energy, its velocity and momentum magnitude can be calculated. By measuring the emission angle with respect to the surface normal, ARPES can also determine the two in-plane components of momentum that are in the photoemission process preserved.

[2] The work function is the minimum energy required to remove an electron from a solid to a point in the vacuum right outside the solid surface.

Chapter 3

Optical processes and excitations in semiconductors

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

We can gain insight in the temperature dependence of the conductivity by considering the intuitively appealing Drude-model. Consider a material with a density n of charge carriers that have charge q. When a uniform dc electric field $\boldsymbol{E}$ is applied. The electron scatters on impurities after (on average) a time interval τ. At any given instant in time, an electron will then on average have been travelling for a time τ since its last collision, and accumulated a momentum

$$\Delta\langle \boldsymbol{p} \rangle = q\boldsymbol{E}\tau. \tag{3.1}$$

However, during its last collision, the electron will just as likely have scattered forward as backward, so all earlier contributions to the electron's momentum can be disregarded. This results in

$$\langle \vec{p} \rangle = q\boldsymbol{E}\tau. \tag{3.2}$$

We can then relate the current density to the electric field via $\langle \boldsymbol{p} \rangle = m\langle \boldsymbol{v} \rangle$ and $\boldsymbol{J} = nq\langle \boldsymbol{v} \rangle$, resulting in Ohm's law:

$$\boldsymbol{J} = \Big(\frac{nq^2\tau}{m}\Big)\boldsymbol{E}. \tag{3.3}$$

This is the Drude formula, and the conductivity is

$$\sigma = \frac{nq^2\tau}{m}. \tag{3.4}$$

This argument can be generalized if there are more than one type of carrier density n_i (which also can have different charges q_i). The more frequent the impurity scattering, the smaller τ and the smaller the conductivity.

We see that temperature now enters both via the density n and the impurity scattering time τ. In metals, n is independent of temperature, whereas the scattering time τ decreases with increasing

temperature. Thus, τ determines majorly the temperature dependence of the conductivity in metals. On the other hand, insulators and semiconductors have no mobile charges at $T = 0$. At finite temperature, charges are induced by thermal excitations which have to overcome the band gap E_g between the valence and the conduction band, yielding. One may derive the following expression for the temperature-dependent density of electrons in the conduction band (see exercise):

$$n = 2\Big(\frac{2\pi m_e^* k_B T}{h^2}\Big)^{3/2} \mathrm{e}^{-E_g/2k_B T} \tag{3.5}$$

For insulators, the energy gap E_g is huge, for instance 5.5 eV for diamond. The resulting charge carrier density at room temperature $T = 300$ K is around $n = 10^{-27}$ cm^{-3}. To achieve an appreciable density of charge carriers in the conduction band, smaller gaps 0.5 - 1 eV are thus necessary. Materials with a band gap in this regime are termed semiconductors. However, the carrier densities of both insulators and semimconductors are dwarfed by the electron density in metals that contributes to charge transport ($n_{\mathrm{metal}} \simeq 10^{23} - 10^{24}$ cm^{-3}). Nonetheless, adding a small amount of impurities in semiconductors in the form of atoms from a differen tmaterial, a process called doping witha cceptors or donors, the conductivity can be manipulated in different ways, which makes semiconductors immensely useful as components in technological applications.

3.1 Semiconductors/insulators in group IV

The most important semiconductor in technology is probably silicon (Si) which - like carbon (C) and germanium (Ge) - belong to group IV of the periodic table. These elements have four electrons in their outermost shell belonging to orbitals $(ns)^2$ and $(np)^2$ where $n = 2$ for C, $n = 3$ for Si and $n = 4$ for Ge. All these elements form crystals with a diamond structure: a face-centered cubic lattice with a two-atom basis, so that a unit cell contains two atoms located at $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. This is shown in Fig. 3.1.

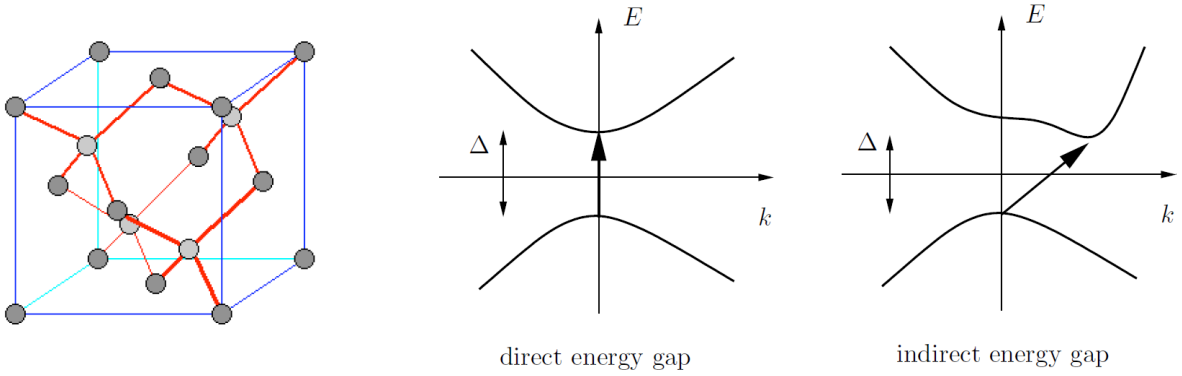

Figure 3.1: **Left:** crystal structure of diamond corresponding to two face-centered cubic lattices shifted by a quarter of lattice spacing along the (1,1,1) direction. **Middle and right:** illustration of direct and indirect band gaps. Figure taken from [3].

The band structure of these materials can be found around the Γ-point by applying the free-electron approximation discussed in the previous chapter, and one finds that the materials display both direct and indirect band gaps (see Fig. 3.1). Because of these gaps, the materials are not conducting at $T = 0$. When the gap is indirect, it means that the minimum of the conduction band

and the maximum of the valence band lie at different points in the Brillouin zone. An example of a semiconductor with a direct band gap is GaAs.

- Carbon has an energy gap of 5.5 eV in the diamond structure, causing it to classify as an insulator. The large energy gap means that diamond is transparent to light in the visible range (1.6 - 3.3 eV), since the EM energy cannot be absorbed by electrons.
- The energy gap of silicon is 1.12 eV and is indirect.
- Germanium has an indirect gap of 0.67 eV.
- GaAs has a direct energy gap of 1.43 eV.

3.2 Elementary excitations

We now develop a simple two-band model to illustrate basic properties of excitations in semiconductors. The Hamiltonian is

$$H = \sum_{\boldsymbol{k},s} \epsilon_{V,\boldsymbol{k}} c^\dagger_{V,\boldsymbol{k},s} c_{V,\boldsymbol{k},s} + \sum_{\boldsymbol{k},s} \epsilon_{C,\boldsymbol{k}} c^\dagger_{C,\boldsymbol{k},s} c_{C,\boldsymbol{k},s} \tag{3.6}$$

where $\epsilon_{V(C),\boldsymbol{k}}$ are the band energies of the valence and conduction band, respectively. The operators $c^\dagger_{n,\boldsymbol{k},s}$ creates an electron with crystal momentum $\boldsymbol{k}$ and spin s in the band $n \in \{V, C\}$. The ground-state is

$$|\Phi_0\rangle = \prod_{\boldsymbol{k},s} c^\dagger_{V,\boldsymbol{k},s} |0\rangle \tag{3.7}$$

so that the valence band is completely filled whereas the conduction band is empty. The product on the right hand side runs over all wavevectors in the first BZ. The ground state energy is $E_0 = 2\sum_{\boldsymbol{k}} \epsilon_{V,\boldsymbol{k}}$ whereas the total momentum and spin of the ground state vanish. We next consider excitations above the ground-state.

3.2.1 Electron-hole excitations

A simple excitation of the system consists of removing an electron, and thus creating a hole, from the valence band and inserting it into the conduction band. We write this excitation as

$$|\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle = c^\dagger_{C,\boldsymbol{k}+\boldsymbol{q},s} c_{V,\boldsymbol{k},s} |\Phi_0\rangle \tag{3.8}$$

since we remove an electron with momentum $\boldsymbol{k}$ and spin s' in the valence band and replace it by an electron with $\boldsymbol{k}+\boldsymbol{q}$ and s in the conduction band. We thus include the possibility to alter both the momentum and spin of the electron in the process. We assume that the state $|\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle$ has been normalized. The electron-hole pair may either be in a spin-singlet state (pure charge excitation $s = s'$) or a spin-triplet state (spin excitation $s \neq s'$).

A particle-hole excitation costs energy, a so-called excitation energy. To compute this energy, we first note that filling an electron state at $\boldsymbol{k}+\boldsymbol{q}$ costs an energy $\epsilon_{\boldsymbol{k}+\boldsymbol{q}}$. However, removing an electron at a filled state $\boldsymbol{k}$ also *costs* energy. Let us explain why this is so.

Hole excitations

Excitations can occur both as electrons or holes: absence of electrons. Consider for simplicity first a single band where all states up to the Fermi level are filled at $T = 0$. An excited electron should find its place outside the Fermi sphere whereas a hole (an unoccupied state) corresponds to removing an electron occur inside the sphere. In the latter case, it might appear at first glance that we are actually removing energy from the system. It is then unclear how this can be considered an excitation. The resolution lies in the fact that the remaining state with the vacant electron state is actually an excited state compared to the ground-state of the remaining electrons. The hole may be formally considered as a quasiparticle, and we here consider its properties more carefully.

A good place to start in terms of understanding how hole excitations behave is energy balance. The excitation energy of the system is:

$$E - E_0 = \sum_k E(k) f_k - \sum_{k<k_F} E(k), \tag{3.9}$$

where E_0 (the last term) is the energy of the system at $T = 0$ with all states below k_F being occupied. We have omitted the spin index in the argument above for simplicity. The quantities $E(k)$ and f_k represent the energies of electrons with momentum k and the distribution function (takes a value between 0 and 1, the probability that a state $E(k)$ will be occupied by an electron), respectively. We can rewrite this equation as

$$E - E_0 = \sum_{k>k_F} E(k) f_k - \sum_{k<k_F} E(k)(1 - f_k) = \sum_{k>k_F} \epsilon(k) f_k + \sum_{k<k_F} |\epsilon(k)|(1 - f_k), \tag{3.10}$$

where we have defined the relative excitation energy with regard to the Fermi level $\epsilon(k) = E(k) - E_F$. To get this equation, we made use of the fact that

$$\sum_k f_k = \sum_{k<k_F} 1 \tag{3.11}$$

due to conservation of particle number. Thus, the total excitation energy of the system can be written as the sum of the energies of all electrons and holes, these energies being defined as the *positive energy* difference between the energy $E(k)$ of the particle and the Fermi energy.

Let us now explain why creating a hole, corresponding to removing an electron below the Fermi level, requires energy even though it actually looks like we are removing energy from the system. Consider first a normal metal at $T = 0$. The ground-state consists of all states up to k_F filled with electrons. Let us denote the total energy of this system as E_0, as above. We now add an extra electron to the metal, which will occupy one of the states $k > k_F$ since the others are filled. The ground-state of the system with this extra electron is

$$E_g = E_0 + k_F^2/2m \tag{3.12}$$

since the lowest energy-state of the system with an extra electron would have been to put that electron at the Fermi level. To be very accurate, we would have to put it at $k = k_F^+$, but we assume that next available state is infinitesimally close to k_F so that we can set $k_F^+ = k_F$. Instead, with the electron put at $k > k_F$ the energy of the system is

$$E_{\text{syst}} = E_0 + k^2/2m. \tag{3.13}$$

The excitation energy of the system is now the difference in energy between the energy of the system and the ground-state of the system:

$$E_{\text{ex}} = E_{\text{syst}} - E_g = k^2/2m - E_F, \tag{3.14}$$

where $E_F = k_F^2/2m$. This energy increase is equal to the energy of the elementary excitation and the system's momentum will be k, namely the momentum of the elementary excitation. Thus, adding an electron at $k > k_F$ is a positive excitation with energy E_{ex} relative the ground-state.

Now we turn to the hole excitation case. Consider again the ground-state of the normal metal, but this time we remove an electron from $k < k_F$. The resulting state can also be treated as an elementary excitation, because its energy is greater than the ground-state energy of the *remaining* electrons. The energy of the system after removing the electron is

$$E_{\text{syst}} = E_0 - k^2/2m. \tag{3.15}$$

But the ground-state of the remaining electrons (after we have removed the electron at $k < k_F$) is:

$$E_g = E_0 - k_F^2/2m. \tag{3.16}$$

This can be understood by taking an electron at the Fermi level and putting it into the place of the missing electron in order to obtain the ground-state for the remaining electrons. Now, just like in the electron case, we obtain the excitation energy of the system by subtracting the ground-state energy from the system energy:

$$E_{\text{ex}} = E_{\text{syst}} - E_g = E_F - k^2/2m. \tag{3.17}$$

This is an energy increase (since $k < k_F$) and it is is equal to the energy of the elementary hole excitation. The system's momentum will be $-k$ since we removed an electron, and $-k$ is then the momentum of the elementary hole excitation. Thus, removing an electron at $k < k_F$ is a positive excitation with energy E_{ex} relative the ground-state.

Note that the total crystal momentum $\sum_k k f_k$ is not a conserved quantity as one increases temperature in an out of equilibrium situation. Increasing temperature changes the distribution function and increases the total momentum of the gas. This can be understood from the fact that increasing temperature increases the kinetic energy of the particles in the system, corresponding to an increase in total momentum for a free electron gas. However, if $f_k = f_{-k}$, then the total crystal momentum is still zero.

Properties of holes

A hole is a quasiparticle representing a missing electron. Consider a normal metal. If the electron has mass m, momentum $\boldsymbol{k}$, charge $-e$, and spin σ, the hole has the following properties:

- Opposite mass, $-m$.
- Opposite momentum, $-\boldsymbol{k}$.
- Opposite spin, $-\sigma$.
- Opposite charge, $+e$.

Going now back to our electron-hole excitation in the system with valence and conduction band, we first determine the chemical potential. Assume a simple band structure for a direct-gap semiconductor based on the $\boldsymbol{k}\cdot\boldsymbol{p}$ approximation:

$$\epsilon_{V,\boldsymbol{k}} = -\frac{\hbar^2\boldsymbol{k}^2}{2m_V},\ \epsilon_{C,\boldsymbol{k}} = E_g + \frac{\hbar^2\boldsymbol{k}^2}{2m_C}. \tag{3.18}$$

Next, compute the electron distribution. The overall electron density is n:

$$n = \frac{1}{V}\sum_{\boldsymbol{k}\in\mathrm{BZ}}\sum_s 1 = 2\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)3}1. \tag{3.19}$$

The density of the electrons in the conduction and valence band are then given by an integral over the probability of occupation, i.e. the Fermi-Dirac distribution function:

$$\begin{aligned}
n_c = 2\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}n_F(\epsilon_{C,\boldsymbol{k}}) &= \int_{\mathrm{BZ}}\frac{d^3k}{4\pi^3}\frac{1}{\mathrm{e}^{\beta(\epsilon_{C,\boldsymbol{k}}-\mu)}+1} \\
&\overset{\text{Assume low temperature } T}{\simeq} \int_{\mathrm{BZ}}\frac{d^3k}{4\pi^3}\mathrm{e}^{-\beta(\epsilon_{C,\boldsymbol{k}}-\mu)} = \frac{\mathrm{e}^{\beta(\mu-E_g)}}{4\pi^3}(2\pi m_C k_B T)^{3/2}, \\
n - n_V = \int_{\mathrm{BZ}}\frac{d^3k}{4\pi^3}[1-n_F(\epsilon_{V,\boldsymbol{k}})] &= \int_{\mathrm{BZ}}\frac{d^3k}{4\pi^3}\frac{1}{\mathrm{e}^{-\beta(\epsilon_{V,\boldsymbol{k}}-\mu)}+1} \\
&\overset{\text{Assume low temperature } T}{\simeq} \int_{\mathrm{BZ}}\frac{d^3k}{4\pi^3}\mathrm{e}^{\beta(\epsilon_{V,\boldsymbol{k}}-\mu)} = \frac{\mathrm{e}^{-\beta\mu}}{4\pi^3}(2\pi m_V k_B T)^{3/2}.
\end{aligned} \tag{3.20}$$

By low temperature, we explicitly mean that both μ and $E_g-\mu$ are much larger than $k_BT = 1/\beta$. The total electron density naturally has to satisfy $n_C + n_V = n$, which means that we can equate the above two lines and find the chemical potential satisfying this equation:

$$\mu = \frac{E_g}{2} + \frac{3}{4}k_BT\cdot\ln\Big(\frac{m_V}{m_C}\Big). \tag{3.21}$$

At $T = 0$, the chemical potential is exactly in the center of the band gap and then moves with increasing T, if the effective masses are different.

So the chemical potential μ lies inside the band gap at low temperatures and has some positive value. Therefore, the excitation energy according to our above reasoning is

$$\begin{aligned}
E &= E_{\text{electron}} + E_{\text{hole}} \\
&= (E_g + \frac{\hbar^2\boldsymbol{k}^2}{2m_C} - \mu) + (\frac{\hbar^2\boldsymbol{k}^2}{2m_V} + \mu) \\
&= \epsilon_{C,\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{V,\boldsymbol{k}}.
\end{aligned} \tag{3.22}$$

The spectrum of the electron-hole excitations with a specific momentum $\boldsymbol{q}$ can be quantitatively described using the spectral function which we define as

$$I(\boldsymbol{q},E) = \sum_{\boldsymbol{k},s,s'}|\langle\boldsymbol{k}+\boldsymbol{q},s;\boldsymbol{k},s'|c^\dagger_{C,\boldsymbol{k}+\boldsymbol{q},s}c_{V,\boldsymbol{k},s'}|\Phi_0\rangle|^2\delta[E-(\epsilon_{C,\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{V,\boldsymbol{k}})]. \tag{3.23}$$

It can be formally derived using Green functions, but this is beyond the content of the present course. Excitations exist for all values of $E,\boldsymbol{q}$ for which $I(\boldsymbol{q},E)\neq 0$. Finite values of the spectral weight requires a $\boldsymbol{q}$-dependent threshold for the energy E. If the band gap is direct, the distance between conduction and valence band is minimal at $\boldsymbol{q}=0$, which corresponds to the energy threshold for creating an excitation. If the bandgap is indirect, the threshold energy is obtained for $\boldsymbol{q}\neq 0$.

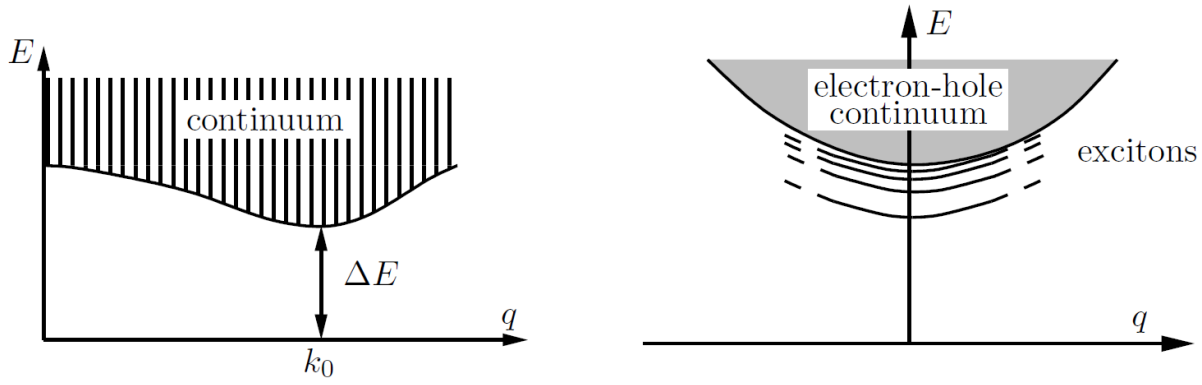

Figure 3.2: **Left:** electron-hole excitation spectrum for $\boldsymbol{k}_0 \neq 0$. Excitations exist in the shaded region where $I(q, E) \neq 0$. **Right:** qualitative form of the exciton spectrum below the electron-hole continuum. Figure taken from [3].

3.2.2 Excitons

When we include the Coulomb interaction between the electrons, another class of excitations arises called *excitons*. To show this, we include the Coulomb interaction in our Hamiltonian through a term

$$V = \sum_{ss'} \int d^3r d^3r' \Psi_s^\dagger(\boldsymbol{r}) \Psi_{s'}^\dagger(\boldsymbol{r}') \frac{e^2}{4\pi\varepsilon_0 |\boldsymbol{r} - \boldsymbol{r}'|} \Psi_{s'}(\boldsymbol{r}') \Psi_s(\boldsymbol{r}), \tag{3.24}$$

where the field operators are defined via (see the chapter on second quantization)

$$\Psi_s(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{n=V,C} \sum_{\boldsymbol{k}} u_{n,\boldsymbol{k}}(\boldsymbol{r}) \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} c_{n,\boldsymbol{k},s} \tag{3.25}$$

and $u_{n,\boldsymbol{k}}$ are the Bloch functions of the band C, V. Consider now a general particle-hole state

$$|\Phi_{\boldsymbol{q}}\rangle = \sum_{\boldsymbol{k}} A(\boldsymbol{k}) c^\dagger_{C,\boldsymbol{k}+\boldsymbol{q},s} c_{V,\boldsymbol{k},s'} |\Phi_0\rangle = \sum_{\boldsymbol{k}} A(\boldsymbol{k}) |\boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s'\rangle \tag{3.26}$$

and demand that it satisfies the stationary Schrödinger equation $(H + V)|\Phi_{\boldsymbol{q}}\rangle = (E_0 + E_{\boldsymbol{q}})|\Phi_{\boldsymbol{q}}\rangle$. This two-body (electron and hole) problem can then be expressed as

$$\sum_{\boldsymbol{k}'} \langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s' | H + V | \boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s' \rangle A(\boldsymbol{k}') = (E_0 + E_{\boldsymbol{q}}) A(\boldsymbol{k}) \tag{3.27}$$

The matrix elements are given by

$$\langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s' | H | \boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s' \rangle = \delta_{\boldsymbol{k},\boldsymbol{k}'} (E_0 + \epsilon_{C,\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{V,\boldsymbol{k}}), \tag{3.28}$$

and

$$\begin{aligned} &\langle \boldsymbol{k}+\boldsymbol{q}, s; \boldsymbol{k}, s' | V | \boldsymbol{k}'+\boldsymbol{q}, s; \boldsymbol{k}', s' \rangle = \\ &\quad \frac{2\delta_{ss'}}{\Omega^2} \int d^3r d^3r' u^*_{C,\boldsymbol{k}+\boldsymbol{q}}(\boldsymbol{r}) u_{V,\boldsymbol{k}}(\boldsymbol{r}) u_{C,\boldsymbol{k}'+\boldsymbol{q}}(\boldsymbol{r}') u^*_{V,\boldsymbol{k}'}(\boldsymbol{r}') \mathrm{e}^{-i\boldsymbol{q}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \frac{e^2}{4\pi\varepsilon_0|\boldsymbol{r}-\boldsymbol{r}'|} \\ &\quad - \frac{1}{\Omega^2} \int d^3r d^3r' u^*_{C,\boldsymbol{k}+\boldsymbol{q}}(\boldsymbol{r}) u_{V,\boldsymbol{k}}(\boldsymbol{r}') u_{C,\boldsymbol{k}'+\boldsymbol{q}}(\boldsymbol{r}) u^*_{V,\boldsymbol{k}'}(\boldsymbol{r}') \times \mathrm{e}^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot(\boldsymbol{r}-\boldsymbol{r}')} \frac{e^2}{4\pi\varepsilon_0|\boldsymbol{r}-\boldsymbol{r}'|}. \end{aligned} \tag{3.29}$$

In the last expression, the first term is the exchange term, and the second term is the direct term of the Coulomb interaction between an electron in the conduction band and an electron in the valence band. This is seen from the spatial argument of the u_C and u_V factors. The exchange term arises as a quantum mechanical effect due to the indistinguishibility of electrons. In our expression for V, the $c^\dagger_{C,V}$ operator from $\psi^\dagger_s(\boldsymbol{r})$ can combine with the $c_{C,V}$ operator from either $\psi_{s'}(\boldsymbol{r}')$ or $\psi_s(\boldsymbol{r})$, which is what provides the two terms in Eq. (3.29).

The electron and hole can now be bound together by the Coulomb interaction, forming a so-called exciton. In principle, one could proceed with Eq. (3.27) and solve this for the energy eigenvalue E. We will instead a simpler and more intuitive picture, including the effective mass approximation, to find the exciton spectrum. Consider the Schrödinger equation for a two-body wavefunction Ψ describing the electron and hole at coordinates $\boldsymbol{r}_e$ and $\boldsymbol{r}_h$:

$$\Big[\Big(\frac{\boldsymbol{p}_e^2}{2m_e^*} + E_G\Big) + \frac{\boldsymbol{p}_h^2}{2m_h^*} - \frac{e^2}{4\pi\epsilon|\boldsymbol{r}_e - \boldsymbol{r}_h|}\Big]\Psi(\boldsymbol{r}_e, \boldsymbol{r}_h) = E\Psi(\boldsymbol{r}_e, \boldsymbol{r}_h). \tag{3.30}$$

This includes the kinetic energy operators for the electrons and holes ($\boldsymbol{p}_{e,h}$ are momentum operators), the extra energy shift E_G due to the electron residing in the conduction band, and the Coulomb interaction between the electron and the hole. Note that the semiconductor is a dielectric medium with a dielectric constant ε, which is why we have used the effective, screened Coulomb interaction in the Schrödinger equation right away.

Now introduce a center of mass coordinate $\boldsymbol{R}$ and relative coordinate $\boldsymbol{r}$:

$$\begin{aligned}\boldsymbol{R} &= \frac{m_e^*\boldsymbol{r}_e + m_h^*\boldsymbol{r}_h}{m_e^* + m_h^*},\\ \boldsymbol{r} &= \boldsymbol{r}_e - \boldsymbol{r}_h.\end{aligned} \tag{3.31}$$

The first describes the movement of the electron and hole as a whole, whereas the latter describes their relative separation distance. Since

$$\frac{\partial}{\partial \boldsymbol{r}_e} = \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{r}_e}\frac{\partial}{\partial \boldsymbol{R}} + \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{r}_e}\frac{\partial}{\partial \boldsymbol{r}} \tag{3.32}$$

and similarly for $\frac{\partial}{\partial \boldsymbol{r}_h}$, we obtain

$$\Big[\frac{\boldsymbol{p}_{\boldsymbol{R}}^2}{2M} + E_G + \frac{\boldsymbol{p}_{\boldsymbol{r}}^2}{2m_r} - \frac{e^2}{4\pi\epsilon|\boldsymbol{r}|}\Big]\Psi(\boldsymbol{r}_e, \boldsymbol{r}_h) = E\Psi(\boldsymbol{r}_e, \boldsymbol{r}_h). \tag{3.33}$$

Here, we have defined the total mass $M = m_e^* + m_h^*$ and the reduced mass $1/m_r = 1/m_e^* + 1/m_h^*$.

We may now observe that the SE permits a separable solution

$$\Psi(\boldsymbol{r}_e, \boldsymbol{r}_h) = A(\boldsymbol{r})B(\boldsymbol{R}) \tag{3.34}$$

where $A(\boldsymbol{r})$ describes the relative motion of the electron and hole whereas $B(\boldsymbol{R})$ describes their center of mass motion. Inserting this Ψ into the SE gives us an eigenvalue equation for $A(\boldsymbol{r})$:

$$\Big[\frac{\boldsymbol{p}_{\boldsymbol{r}}^2}{2m_r} + E_G - \frac{e^2}{4\pi\epsilon|\boldsymbol{r}|}\Big]A(\boldsymbol{r}) = E_A A(\boldsymbol{r}). \tag{3.35}$$

This is recognized as the SE for a hydrogen atom, and we immediately know that the eigenvalues E_A are characterized by quantum numbers $n = 1, 2, \ldots$ so that

$$E_A = E_G + E_n = E_G - \frac{m_r e^4}{32\pi^2\epsilon^2\hbar^2 n^2}. \tag{3.36}$$

The eigenvalue equation for $B(\boldsymbol{R})$ takes the form

$$\frac{\boldsymbol{p}_{\boldsymbol{R}}^2}{2M} B(\boldsymbol{R}) = E_B B(\boldsymbol{R}). \tag{3.37}$$

Again, we immediately know the solution: this is a free particle SE with eigenvalues

$$E_B = \frac{\hbar^2 \boldsymbol{q}^2}{2M} \tag{3.38}$$

where $\boldsymbol{q}$ is the wavevector of the plane wave eigenstate. The total energy eigenvalue $E = E_A + E_B$ is then

$$E = E_G - \frac{m_r e^4}{32\pi^2\epsilon^2\hbar^2 n^2} + \frac{\hbar^2 \boldsymbol{q}^2}{2M}, n = 1, 2, \ldots \tag{3.39}$$

This shows that excitations exist right below the conduction band (E_g at $\boldsymbol{q} = 0$), see right panel of Fig. 3.2. This non-trivial quasiparticle is called an *exciton*.

Its wavefunction has a free-particle term, showing that the exciton can move freely as a unit through the crystal. It also has a bound-state term, describing that the electron and hole must move together in the bound-state known as the exciton. In the present approximation, the exciton thus takes the form of a two-particle (particle-hole) excitation where screening is taken into account. In this sense it can be viewed as a collective excitation, since the modified dielectric constant includes the polarization by all electrons.

Note that when screening is neglected, meaning one uses the vacuum permittivity, which moves the energies of the excitons far below the band edge. Instead, for the case of a weak binding energy which shifts the exciton energy only slightly below the continuum of states, the excitation is called a Wannier exciton. Using reasonable values for $\mu_{\text{ex}} \simeq m/10$ and $\varepsilon = 10\varepsilon_0$, as well as for the electron density n, one obtains that the excitons reside in the meV range below the gap edge. The reduction in energy can be understood as a bound-state formation of the oppositely charged electron and hole lowering their energy due to the Coulomb interaction. The exciton is thus a composite boson since it consists of two fermion excitations (particle and hole).

The exciton levels are dispersive, and their spectrum becomes increasingly dense with increasing energy, similar to the hydrogen atom. When the levels merge with the continuum band, the exciton state is ionized so that the electron and hole dissociate and behave like independent particles, since there are now available states in the conduction band. Strongly bound excitons are called Frenkel excitons. For very strong binding, the pair is almost local in the sense that the excitation is restricted to a single atom rather than involving the whole semiconductor band structure.

Excitons are mobile, but carry no charge. Their spin depends on s and s'. If $s = s'$, the exciton is a spin-singlet, while for $s \neq s'$ it has a spin triplet character. For small densities, excitons approximately obey Bose-Einstein statistics, and in special cases Bose-Einstein condensation of excitons can be observed experimentally. Excitons are important in the optical spectra of semiconductors, in particular at low temperatures, but also with respect to device applications. This is because light emitting diodes and semiconductor lasers often involve excitons.

3.3 Optical properties

Excitations in semiconductors can occur via absorption of EM radiation. The energy and momentum imparted by the photon is $\hbar\omega$ and $\hbar\boldsymbol{q}$, respectively. With the light dispersion $\omega = c|\boldsymbol{q}|$ and setting $E_g \simeq 1$ eV, meaning that the photon must carry an energy of at least 1 eV to excite an electron from the valence to conduction band, the corresponding momentum transfer is estimated to

$$q = \frac{\omega}{c} = \frac{\hbar\omega}{hc} 2\pi \simeq \frac{E_g}{hc} 2\pi a = \alpha \frac{2\pi}{a} \ll \frac{2\pi}{a}. \tag{3.40}$$

Here, c is the speed of light, a the lattice constant and $\alpha \simeq 1/137$ is the fine structure constant. We see that the momentum transfer from a photon to the excited electron is much less than the 1BZ boundaries and can be neglected. In other words, purely electromagnetic excitations lead only to "direct" excitations.

For semiconductors with a direct energy gap (such as GaAs), the photo-induced electron-hole excitation yields absorption rates with the characteristics (computed from Fermi's golden rule)

$$\Gamma_{\text{abs}}(\omega) \propto \begin{cases} (\hbar\omega - E_g)^{1/2}, \text{ dipole-allowed} \\ (\hbar\omega - E_g)^{3/2}, \text{ dipole-forbidden.} \end{cases} \tag{3.41}$$

Here, the terms dipole-allowed and dipole-forbidden have a similar meaning as in the excitation of atoms regarding whether matrix elements of the type $\langle u_{V,\boldsymbol{k}} | \boldsymbol{r} | u_{C,\boldsymbol{k}} \rangle$ are finite or vanish, respectively. We see that dipole-allowed transitions occur at a higher rate for photon energies right above the energy gap E_g.

For semiconductors with indirect energy gap (such as Si and Ge), the lowest energy transition connecting the top of the valence band to the bottom of the conduction band is not allowed without the help of phonons (quantized lattice vibration). These phonons contribute little energy $\omega_{\boldsymbol{Q}}$, but much momentum transfer $\boldsymbol{Q}$ since $\hbar\omega_{\boldsymbol{Q}} \ll \hbar\omega$ since $\omega_{\boldsymbol{Q}} = c_S|\boldsymbol{Q}|$ and the sound velocity $c_s \ll c$. The phonon-assisted photon absorption in a semiconductor with indirect gap is shown in the left panel of Fig. 3.3.

The requirement of a phonon assisting in the transition changes the transition rate to

$$\Gamma_{\text{abs}}(\omega) \propto c_+(\hbar\omega + \hbar\omega_{\boldsymbol{Q}} - E_g)^2 + c_-(\hbar\omega - \hbar\omega_{\boldsymbol{Q}} - E_g)^2. \tag{3.42}$$

Here, $c_\pm$ are constants and $\boldsymbol{Q}$ corresponds to the wave vector of the phonon connecting the top of the valence band and the bottom of the conduction band. There are two ways this can happen: either the phonon is absorbed (c_+-process) or it is emitted (c_--process).

In addition to the absorption into the particle-hole spectrum, absorption processes involving our exciton states discussed above exist. They lead to discrete absorption peaks below the absorption continuum, as shown in the right panel of Fig. 3.3. It is also possible for electrons and holes to recombine (de-excitation of an electron). In a radiative recombination, a photon is emitted. Other recombination channels, such as recombination at impurities and interfaces are also possible (due to scattering). The radiative recombination for direct-gap semiconductors is most relevant for applications.

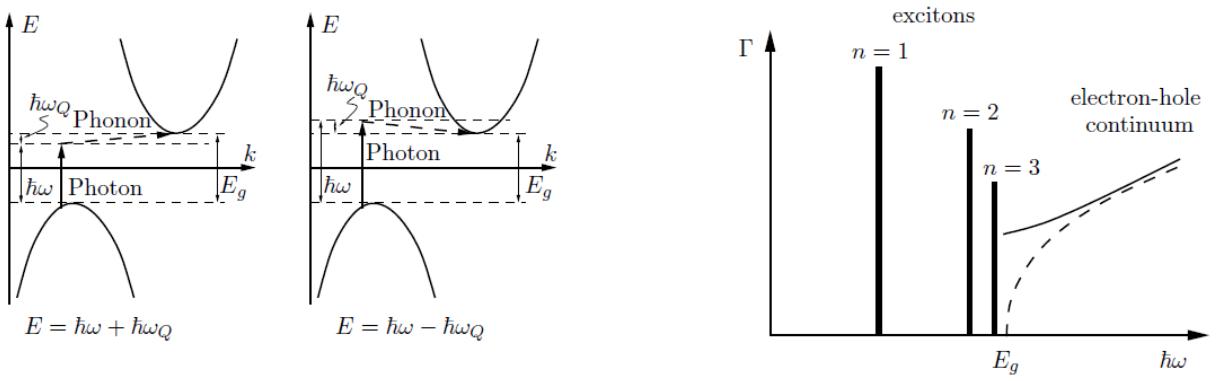

Figure 3.3: **Left:** phonon-assisted photon absorption in a semiconductor with indirect gap; phonon absorption in the left panel and phonon emission in the right panel. **Right:** absorption spectrum including the exciton states for a direct-gap semiconductor. The exciton states appear as sharp lines, due to their discrete nature, below the e-h continuum starting at $\hbar\omega = E_g$. Figure taken from [3].

3.3.1 Experimental measurement of excitons

To determine the position of an excitonic energy level, one has to identify the difference between the electrical band gap (energy difference between valence and conduction band) and the optical gap. The latter is the energy difference between the valence band and the lowest-lying exciton level. The optical gap is thus smaller than the electrical band gap, and the difference between the two can be experimentally measured via optical absorption. As shown in Fig. 3.3, the absorption displays sharp peaks when the photon energy matches the optical gap to different exciton levels, and instead turns into a continuous curve as the photon energy becomes large enough to excite across the electrical band gap. The binding energy of the exciton is then identified as the difference between the electrical band gap and the optical gap.

3.4 Doping semiconductors

Let us replace a Si atom in a Si semiconductor by aluminium Al (group III) or phosphorous P (group V), which then act as impurities in the crystal lattice. Their electron configurations are given by

$$\begin{aligned} &\text{Al: } [(1s)^2(2s)^2(2p)^6](3s)^2(3p) \\ &\text{P: } [(1s)^2(2s)^2(2p)^6](3s)^2(3p)^3. \end{aligned} \tag{3.43}$$

Thus, the compound Al (P) has one electron less (more) than Si.

3.4.1 Impurity state

Consider the case of a P-impurity contributing an additional electron whose dynamics is governed by the conduction band of the semiconductor. For simplicity, let the conduction band be a single isotropic band with effective mass m_C:

$$\epsilon_{\boldsymbol{k}} = \frac{\hbar^2 \boldsymbol{k}^2}{2m_C} + E_g. \tag{3.44}$$

Compared to the Si background, the phosphorous ion is a positively charged center which attracts its additional electron. We can describe this situation by a Schrödinger equation for the effective "hydrogen" atom (positively charged center)

$$\Big[-\frac{\hbar^2\nabla^2}{2m_C}-\frac{e^2}{4\pi\varepsilon|\boldsymbol{r}|}\Big]F(\boldsymbol{r})=EF(\boldsymbol{r}) \tag{3.45}$$

where screening of the ionic potential by th surrounding electrons is accounted for via ε. Analogous to the discussion of the exciton states, $F(\boldsymbol{r})$ is an envelope wavefunction of the electron. Therefore, we know that the low-energy states of the additional electron must be bound states around the P ion. The electron becomes mobile when this "reduced hydrogen atom" is ionized. The binding energy relative to the minimum of the conduction band is given by

$$E_n-E_g=-\frac{m_Ce^4}{32\pi^2\hbar^2\varepsilon^2n^2} \tag{3.46}$$

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

$$r_1=\frac{4\pi\hbar^2\varepsilon}{m^*e^2}=\frac{\varepsilon m}{m_C}a_B. \tag{3.47}$$

Here, $a_B=0.53$ Åis the Bohr radius for the hydrogen atom. For Si, we find with $m_c=0.2m$ and $\varepsilon\simeq 12$ that

$$E_1-E_g\simeq -20\text{ meV and }r_1\simeq 30\text{Å}. \tag{3.48}$$

The resulting state is thus weakly bound, with energies inside the band gap. The net effect of the P-impurities is therefore to inroduce additional electrons into the crystal whose energies are situated just below the conduction band ($E_g\sim 1$ eV while $E_g-E_1\sim 10$ meV). These impurity states can be easily transferred to the conduction band by thermal excitation. One then speaks of an n-doped semiconductor, where n stands for negative charge. In full analogy, one can instead consider Al-impurities, thereby replacing electrons with holes. An Al-atom introduces an additional hole into the lattice which is weakly bound to the Al-ion (its energy is slightly above the band edge of the valence band), and may dissociate from the impurity by thermal excitation. This case is called p-doping, where p stands for positive charge. In both cases, the chemical potential is tied to the dopand levels: in effect, it likes between the dopand level and the valence band for p-doping and between the dopand level and the conduction band in the case of n-doping.

The electric conductivity of semiconductors, in particular at room temperature where thermal excitations are more likely than at lower temperatures, can be tuned strongly by doping with so-called donors (n-doping) and acceptors (p-doping). Practically all dopand atoms are ionized, with the electrons/holes becoming mobile. Combining differently doped semiconductors, the possibility to engineer electronic properties is enhanced further. The is the basic reason for semiconductors being so frequently used in modern electronics.

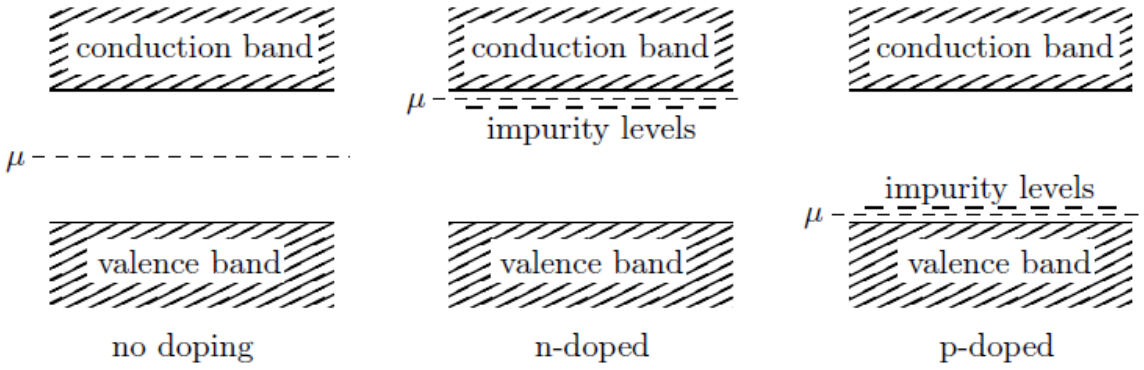

Figure 3.4: Position of the chemical potential in semiconductors. Figure taken from [3].

Chapter 4

Plasmons, phonons, polarons, polaritons

The electronic states in a periodic atomic lattice which are extended have an energy spectrum forming energy bands. In the ground state these energy states are filled successively starting at the bottom of the electronic spectrum until the number of electrons is exhausted. Metallic behavior occurs whenever in this way a band is only partially filled. The fundamental difference that distinguishes metals from insulators and semiconductors is the absence of a gap for electron-hole excitations, which we will explain in this chapter. In metals, the ground state can be excited at arbitrarily small energies which has profound phenomenological consequences.

We will consider a basic model suitable for the description of simple metals like the Alkali metals Li, Na, or K, where the (atomic) electron configuration consists of closed shell cores and one single valence electron in an ns-orbital. Neglecting the core electrons (completely filled bands), we consider the valence electrons only and apply the approximation of nearly free electrons. The lowest band around the Γ-point is then half-filled. First, we will also neglect the in uence of the periodic lattice potential and consider the problem of a free electron gas subject to mutual (repulsive) Coulomb interaction.

An important point is that there exists not only excitations corresponding to electrons, but also with respect to the behavior of the crystal lattice (ion) itself, called phonons. There can also exist hybrid excitations which are combinations of electrons and phonons, called polarons. Even light can be part of the excitations of a metal when combined with a phonon, called a polariton. The electron excitations themselves can also come together and form a so-called collective excitation in which the electrons do not act individually, but are each part of a larger excitation called a plasmon. We now proceed to describe all of these excitations in detail.

4.1 Fundamental quantum properties of metals

To introduce quasiparticles such as plasmons, we first need to talk a bit about some of the properties of metals. The *jellium* model is probably the simplest possible model of a metal that can capture qualitative and to some extent even quantitative aspects of simple metals. The main simplifictation made is to replace the ionic lattice by a homogeneous positively charged background (the so-called jellium) which the electrons move in. The uniform charge density en_{ion} is chosen such that the whole system - electrons and ionic background - is charge neutral: $n_{\text{ion}} = n$ where n is the electron density. Since the periodic lattice is now gone, the system is fully translationally invariant, and the

plane waves

$$\psi_{\boldsymbol{k},s}(\boldsymbol{r}) = \sqrt{\frac{1}{\Omega}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{4.1}$$

represent the single-particle wavefunctions of the free electrons. Here, Ω is as before the volume of the system and $\boldsymbol{k}, s$ denote the wavevector and spin, respectively. Assume a cubic system of side length L and volume $\Omega = L^3$, and impose periodic boundary conditions for the wavefunction

$$\psi_{\boldsymbol{k},s}(\boldsymbol{r} + L\boldsymbol{e}_i) = \psi_{\boldsymbol{k},s}(\boldsymbol{r}),\ i = x, y, z. \tag{4.2}$$

Then, reciprocal lattice space is discretized as

$$\boldsymbol{k} = \frac{2\pi}{L}(n_x, n_y, n_z) \tag{4.3}$$

where $n_i \in \mathbb{Z}^3$. The energy of a single-electron state is $\epsilon_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$ and the ground-state of non-interacting electrons is obtained by filling all single particle states up to the Fermi energy with two electrons. Expressed through second quantization, we thus have

$$|\Psi_0\rangle = \prod_{|\boldsymbol{k}|\leq k_F} \prod_s c^\dagger_{\boldsymbol{k},s}|0\rangle. \tag{4.4}$$

where $c^\dagger_{\boldsymbol{k},s}, c_{\boldsymbol{k},s}$ create and annihilate an electron state. The Fermi wavevector k_F is determined by equating the filled electronic states with the electron density n:

$$n = \frac{1}{\Omega}\sum_{|\boldsymbol{k}|\leq k_F} 1 = 2\int \frac{d^3k}{(2\pi)^3} 1 = \frac{8\pi}{3}\frac{k_F^3}{(2\pi)^3} \tag{4.5}$$

so that $k_F = (3\pi^2 n)^{1/3}$ is the radius of the Fermi sphere in $\boldsymbol{k}$-space around $\boldsymbol{k} = 0$.

4.1.1 Basic thermodynamic properties of metals

The thermodynamic properties of a free electron gas are obtained using the Fermi-Dirac distribution function $f(\epsilon_{\boldsymbol{k}})$, which tells us the probability that an electron state $\epsilon_{\boldsymbol{k}}$ is occupied:

$$f(\epsilon_{\boldsymbol{k}}) = \frac{1}{\mathrm{e}^{(\epsilon_{\boldsymbol{k}}-\mu)/k_BT+1}}. \tag{4.6}$$

The density of states is also useful:

$$\begin{aligned} N(E) &= \sum_{\boldsymbol{k},s} \delta(E - \epsilon_k) = 2\Omega \int \frac{d^3k}{(2\pi)^3}\delta(E - \hbar^2\boldsymbol{k}^2/2m) = \frac{\Omega}{4\pi^3}\int d\Omega dk k^2 \frac{m}{\hbar^2 k}\delta(k - \sqrt{2mE}\hbar) \\ &= \frac{\Omega}{2\pi^2}\Big(\frac{2m}{\hbar^2}\Big)^{3/2} E^{1/2} = \frac{3}{2}\frac{N}{\epsilon_F}\Big(\frac{E}{\epsilon_F}\Big)^{1/2} \end{aligned} \tag{4.7}$$

for $E > 0$ and with $N = n\Omega$ being the total number of electrons. Note that the $d\Omega$ inside the integral in the upper line does not mean integration over volume, but over solid angle element. The chemical potential can be expressed analytically up to second order in temperature at a fixed electron number N, by using

$$N = \sum_{\boldsymbol{k},s} f(\epsilon_{\boldsymbol{k}}) = \int_0^\infty f(E)N(E) = \int_0^\mu dE N(E) + \frac{\pi^2}{6}(k_BT)^2 N'(\mu) + \ldots \tag{4.8}$$

where we utilized the Sommerfeld expansion [1] assuming $T \ll T_F = \epsilon_F/k_B$. We now use

$$\int_0^{\mu} dEN(E) \simeq \int_0^{\epsilon_F} dEN(E) + (\mu - \epsilon_F)N(\epsilon_F) = N + (\mu - \epsilon_F)N(\epsilon_F) \tag{4.13}$$

leading to

$$N \simeq N + (\mu - \epsilon_F)N(\epsilon_F) + \frac{\pi^2}{6}(k_BT)^2N'(\epsilon) \to \mu(T) = \epsilon_F - \frac{\pi^2}{6}(k_BT)^2\frac{N'(\epsilon_F)}{N(\epsilon_F)} + \dots \tag{4.14}$$

where $N'(\epsilon_F)/N(\epsilon_F) = 1/2\epsilon_F$. We can also determine the internal energy

$$U(T) = \int_0^{\infty} dEEN(E)f(E) \overset{\text{See exercise for derivation}}{\simeq} U_0 + \frac{\pi^2}{6}(k_BT)^2N(\epsilon_F). \tag{4.15}$$

The specific heat is then given by

$$C = \frac{1}{\Omega}\frac{\partial U}{\partial T} = \gamma T \text{ where } \gamma = \frac{\pi^2}{3\Omega}k_B^2N(\epsilon_F) \tag{4.16}$$

and is thus linear in T where γ is the Sommerfeld coefficient, proportional to the DOS at the Fermi energy.

Consider now the effect of a magnetic field coupling to the electron spin so that $\epsilon_{\boldsymbol{k}} \to \epsilon_{\boldsymbol{k},s} = \epsilon_{\boldsymbol{k}} - s\mu_B B$ with μ_B the Bohr magneton and B the external magnetic field with $s = \pm 1$. The magnetization generated due to the spin polarization of the electrons can then be computed as

$$\begin{aligned} M = \mu_B(N_+ - N_-) &= \frac{\mu_B}{2}\Big[\int_0^{\infty} dEN(E)f(E - \mu_B B) - \int_+^{\infty} dEN(E)f(E + \mu_B B)\Big] \\ &\simeq \frac{\mu_B}{2}\int_0^{\infty} dEN(E)\Big(-\frac{\partial f(E)}{\partial E}\Big)2\mu_B B \\ &\simeq \mu_B^2 BN(\epsilon_F)\int_0^{\infty}\Big(-\frac{\partial f(E)}{\partial E}\Big) = \mu_B^2 BN(\epsilon_F). \end{aligned} \tag{4.17}$$

[1] The Sommerfeld expansion consists of the following. In the limit $k_BT \ll \epsilon_F$, the derivative $\partial f(E)/\partial E$ is very sharp and narrow (almost Dirac-delta function-like) around $E = \mu$. Therefore:

$$\begin{aligned} \int_{-\infty}^{\infty} dEg(E)\Big(-\frac{\partial f(E)}{\partial E}\Big) &= \int_{-\infty}^{\infty} dE\Big[g(\mu) + (E-\mu)g'(\mu) + \frac{(E-\mu)^2}{2}g''(\mu) + \dots\Big]\Big(-\frac{\partial f(E)}{\partial E}\Big) \\ &= g(\mu) + \frac{g''(\mu)}{2}\int_{-\infty}^{\infty} dE(E-\mu)^2\Big(-\frac{\partial f(E)}{\partial E}\Big) + \dots = g(\mu) + \frac{\pi^2}{6}g''(\mu)(k_BT)^2 + \dots \end{aligned} \tag{4.9}$$

and analogous

$$\int_{-\infty}^{\infty} dE\Big(\frac{\partial g(E)}{\partial E}\Big)f(E) = g(\mu) + \frac{\pi^2}{6}g''(\mu)(k_BT)^2 + \dots \tag{4.10}$$

Here we used the definition

$$g(E) = \int_{-\infty}^{E} dE'\Gamma(E') \to \int_{-\infty}^{\infty} dE\Gamma(E)f(E) = \int_{-\infty}^{\mu} dE\Gamma(E) + \frac{\pi^2}{6}(k_BT)^2\Gamma'(\mu) + \dots \tag{4.11}$$

Note that

$$\int_{-\infty}^{\infty} dx\frac{\beta x^2 e^{\beta x}}{(e^{\beta x}+1)^2} = \frac{\pi^2}{3\beta^2}. \tag{4.12}$$

Taking the derivative with respect to B we find for the Pauli paramagnetic susceptibility

$$\chi_p = \frac{\partial M}{\partial H} \simeq \mu_0 \frac{\partial M}{\partial B} = \mu_0 \mu_B^2 N(\epsilon_F). \tag{4.18}$$

Here, we used that $B = \mu H \simeq \mu_0 H$ for a small susceptibility $\chi_p \ll 1$ since $\mu/\mu_0 = 1 + \chi_p$. The Pauli susceptibility is indeed typically small. We will discuss paramagnetism of the electron gas in more detail in a later chapter.

4.1.2 Stability of metals: a Hartree-Fock approach

For the jellium model to make sense, it has to be a stable physical system. To check this, we compute its ground state energy variationally, using the density n as a variational parameter. This will give is an understanding of the stability of the metal, in effect the cohesion of the ion lattice through itinerant electrons. The variational ground state is denoted $|\Psi_0\rangle$ from Eq. (4.4) for a given k_F. The Hamiltonian consists of four terms:

$$H = H_{\rm kin} + H_{\rm ee} + H_{\rm ei} + H_{\rm ii} \tag{4.19}$$

where the various terms describe the kinetic energy of the electrons, electron-electron interactions, electron-ion interactions, and ion-ion interactions:

$$\begin{aligned}
H_{\rm kin} &= \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k},s} c_{\boldsymbol{k},s} \\
H_{\rm ee} &= \frac{1}{2} \sum_{ss'} \int d^3r d^3r' \psi^\dagger_s(\boldsymbol{r}) \psi^\dagger_{s'}(\boldsymbol{r}') \frac{e^2}{4\pi\epsilon_0 |\boldsymbol{r} - \boldsymbol{r}'|} \psi_s(\boldsymbol{r}') \psi_s(\boldsymbol{r}) \\
H_{\rm ei} &= -\sum_s \int d^3r d^3r' \frac{ne^2}{4\pi\epsilon_0 |\boldsymbol{r} - \boldsymbol{r}'|} \psi^\dagger_s(\boldsymbol{r}) \psi_s(\boldsymbol{r}) \\
H_{\rm ii} &= \frac{1}{2} \int d^3r d^3r' \frac{n^2 e^2}{4\pi\epsilon_0 |\boldsymbol{r} - \boldsymbol{r}'|}
\end{aligned} \tag{4.20}$$

where we introduced the second quantized electron field operators, as discussed previously:

$$\begin{aligned}
\psi^\dagger_s(\boldsymbol{r}) &= \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k},s} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \\
\psi_s(\boldsymbol{r}) &= \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{k}} c_{\boldsymbol{k},s} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}}.
\end{aligned} \tag{4.21}$$

We now want to minimize the variational energy $E_g = \langle\Psi_0|H|\Psi_0\rangle$ with respect to n to see if a stable metallic state with non-zero n can really be obtained. The variational energy has four different contributions.

First, we have the kinetic energy

$$\begin{aligned}
E_{\rm kin} = \langle\Psi_0|H_{\rm kin}|\Psi_0\rangle &= \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} \overbrace{\langle\Psi_0|c^\dagger_{\boldsymbol{k},s} c_{\boldsymbol{k},s}|\Psi_0\rangle}^{=n_{\boldsymbol{k},s}} \\
&= 2\Omega \int \frac{d^3k}{(2\pi)^3} \epsilon_{\boldsymbol{k}} n_{\boldsymbol{k},s} = \frac{3}{5} N \epsilon_F.
\end{aligned} \tag{4.22}$$

where we used that

$$n_{\boldsymbol{k},s} = \begin{cases} 1 \text{ for } |\boldsymbol{k}| \leq k_F \\ 0 \text{ for } |\boldsymbol{k}| > k_F \end{cases} \tag{4.23}$$

Second, there is the Coulomb repulsion energy between the electrons

$$\begin{aligned} E_{\text{ee}} &= \frac{1}{2}\int d^3r d^3r' \frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}-\boldsymbol{r}'|} \sum ss' \langle\Psi_0|\Psi_s^\dagger(\boldsymbol{r})\Psi_{s'}^\dagger(\boldsymbol{r}')\Psi_{s'}(\boldsymbol{r}')\Psi_s(\boldsymbol{r})|\Psi_0\rangle \\ &\overset{\text{See exercise}}{=} \frac{1}{2}\int d^3r d^3r' \frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}-\boldsymbol{r}'|}[n^2 - G(\boldsymbol{r}-\boldsymbol{r}')] = E_{\text{Hartree}} + E_{\text{Fock}}. \end{aligned} \tag{4.24}$$

where

$$G(\boldsymbol{r}) = \frac{9n^2}{2}\Big(\frac{k_F|\boldsymbol{r}|\cos(k_F|\boldsymbol{r}|) - \sin(k_F|\boldsymbol{r}|)}{(k_F|\boldsymbol{r}|)^3}\Big). \tag{4.25}$$

The Coulomb repulsion H_{ee} between the electrons combined with the fermion statistics (Pauli principle) that the field operators $\Psi_s(\boldsymbol{r})$ have to satisfy thus leads to two terms. The first term is the *direct*, or Hartree, term describing the Coulomb energy of a uniformly spread charge distribution. The second term is the *exchange*, or Fock, term resulting from the exchange of two of the electron coordinates in the upper line of Eq. (4.24) which is possible due to the indistinguishibility of any two electrons.

The third contribution originates from the attractive interaction between the ionic background and the electrons

$$\begin{aligned} E_{\text{ei}} &= -\int d^3r d^3r' \frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}-\boldsymbol{r}'|} n \sum_s \langle\Psi_0|\Psi_s^\dagger(\boldsymbol{r})\Psi_s(\boldsymbol{r})|\Psi_0\rangle \\ &= -\int d^3r d^3r' \frac{e^2 n^2}{4\pi\epsilon_0|\boldsymbol{r}-\boldsymbol{r}'|} \end{aligned} \tag{4.26}$$

since $\langle\Psi_0|\Psi_s^\dagger(\boldsymbol{r})\Psi_s(\boldsymbol{r})|\Psi_0\rangle$ corresponds to the uniform density n.

Finally, there is the repulsive ion-ion interaction:

$$E_{\text{ii}} = \langle\Psi_0|H_{\text{ii}}|\Psi_0\rangle = \frac{1}{2}\int d^3r d^3r' \frac{n^2e^2}{4\pi\epsilon_0|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{4.27}$$

We can now verify that the three contributions E_{Hartree}, E_{ei}, and E_{ii} compensate each other to exactly zero. Interestingly, these three terms are the only ones that would arise in a classical electrostatic calculation. This implies that the stability of metals will be determined by quantum effects. The remaining terms are the kinetic energy and the Fock term. The latter is negative and is evaluated to

$$E_{\text{Fock}} = -\Omega\frac{9n^2}{4}\int d^3r \frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}|}\Big(\frac{k_F|\boldsymbol{r}|\cos(k_F|\boldsymbol{r}|) - \sin(k_F|\boldsymbol{r}|)}{(k_F|\boldsymbol{r}|)^3}\Big)^2 = -Nk_F\frac{3e^2}{4\pi}. \tag{4.28}$$

Therefore, we find the total variational energy per electron

$$\frac{E_g}{N} = \frac{3}{5}\frac{\hbar^2k_F^2}{2m} - \frac{3e^2}{16\pi^2\epsilon_0}k_F = \Big(\frac{2.21}{r_s^2} - \frac{0.916}{r_s}\Big)\ \text{Ry} \tag{4.29}$$

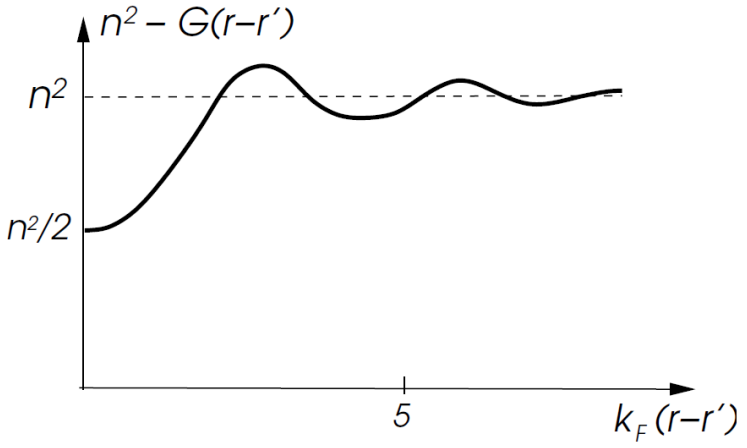

Figure 4.1: Pair correlation function in the electron-electron interaction energy. Figure taken from [3].

where 1 Ry $= e^2/(8\pi\epsilon_0 a_B) \simeq 13.6$ eV with a_B being the Bohr radius and the dimensionless quantity r_s defined via

$$r_s = \frac{d}{a_B} = \left(\frac{9\pi}{4}\right)^{1/3} \frac{me^2}{4\pi\epsilon_0\hbar^2 k_F}, \quad \text{where } d \text{ is defined via } n = \frac{3}{4\pi d^3}. \tag{4.30}$$

The meaning of d is thus the average radius of the volume occupied by one electron. Minimizing the energy per electron with respect to n is then equivalent to minimizing it with respect to r_S. This provides

$$r_{s,\min} = 4.83 \rightarrow d \simeq 2.5 \text{ Å}. \tag{4.31}$$

This corresponds to a lattice constant of

$$a = (4\pi/3)^{1/3} d \simeq 4 \text{ Å}. \tag{4.32}$$

Such an estimate is in good agreement with the lattice constants found in alkali metals: $r_{s,\text{Li}} = 3.22, r_{s,\text{Na}} = 3.96, r_{s,\text{K}} = 4.86$. Therefore, in metals the delocalized electorns are responsible for the cohesion of the positive background and yield a stable solid.

The good agreement of this simple estimate with the experimental values is due to the fact that the Alkali metals have only one valence electron in an s-orbital that is delocalized, whereas the the core electrons are in a noble gas configuration and, thus, relatively inert. In the variational approach outlined above correlation effects among the electrons due to the Coulomb repulsion have been neglected in the choice of variational ground-state. In particular, electrons can be expected to 'avoid' each other not just because of the Pauli principle, as in $|\Psi_0\rangle$, but also as a result of the repulsive interaction. However, for the problem under consideration the correlation corrections turn out to be small for $r_s \simeq r_{s,\min}$ which can be seen from a more sophisticated quantum field theoretical analysis.

4.1.3 Charge excitations and the dielectric function

In both metals and semiconductors, which we will consider in detail later, the elementary excitations are electron-hole excitations. For metals, these can have arbitrarily small energies due to the absence

of a gap in the band-energies at the Fermi level. One drastic consequence of this behavior is the strong screening of the long-ranged Coulomb potential. As we will show, a negative test charge in a metal reduces the electron density in its vicinity. This can be modelled as an induced cloud of positive charges relative a uniform charge density, which weaken the Coulomb potential in the following way

$$V(r) \propto \frac{1}{r} \to V'(r) \propto \frac{\mathrm{e}^{-r/l}}{r}. \tag{4.33}$$

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

$$V(r) \propto \frac{1}{r} \to V'(r) \propto \frac{1}{\varepsilon_r r} \tag{4.34}$$

where ε_r is the relative permittivity ($\varepsilon = \varepsilon_r \varepsilon_0$). The semiconductor acts as a dielectric medium and its screening effects are accounted for by the polarization of localized electric dipoles. This renormalizes the Coulomb potential inside the semiconductors by ε_r.

Dielectric response and Lindhard function

We now determine the response of an electron gas to a time- and position-dependent weak external potential $V_a(\boldsymbol{r}, t)$ based on the Heisenberg equation of motion. The total Hamiltonian is

$$H = H_{\text{kin}} + H_V = \sum_{\boldsymbol{k},s} \epsilon_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k},s} c_{\boldsymbol{k},s} + \sum_s \int d^3 r V_a(\boldsymbol{r}, t) \Psi^\dagger_s(\boldsymbol{r}) \Psi_s(\boldsymbol{r}) \tag{4.35}$$

and the second term is considered as a small perturbation. We first consider the linear response of the system to the external potential. We consider a harmonically varying potential (one Fourier component in the spatial and time dependence), and the math is more convenient by representing the potential as a complex number, similarly to how one works with complex numbers (phasors) in ac electric circuits with inductors and capacitors. The potential has an amplitude and a phase, and so we may write

$$V_a(\boldsymbol{r}, t) = V_a(\boldsymbol{q}, \omega) \mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{r} - i\omega t} \mathrm{e}^{\eta t} \tag{4.36}$$

where the physical potential is just the real part of this quantity. This complex representation works because we will work with linear response, i.e. linear differential equations, ensuring that the real and imaginary parts of the potential do not get mixed up in an incorrect way (which could happen if we had higher-order equations in the potential). Above, $\eta \to 0^+$ models that we adiabatically switch on the potential.

To linear response, this potential is assumed to induce a small modulation of the electron density $n(\boldsymbol{r}, t)$ so that

$$n(\boldsymbol{r}, t) = n_0 + \delta n_{\text{ind}}(\boldsymbol{r}, t) \tag{4.37}$$

where

$$\delta n_{\text{ind}}(\boldsymbol{r}, t) = \delta n_{\text{ind}}(\boldsymbol{q}, \omega) \mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{r} - i\omega t} \tag{4.38}$$

Any phase-shift between the oscillations in the induced density $\delta n_{\rm ind}$ and the applied potential is encoded in the relative phase between $V_a(\boldsymbol{q},\omega)$ and $\delta n_{\rm ind}(\boldsymbol{q},\omega)$.
Fourier-transforming the density operator $\rho(\boldsymbol{r})$, we obtain

$$\rho_{\boldsymbol{q}} = \sum_S \int d^3r \Psi_s^\dagger(\boldsymbol{r})\Psi_s(\boldsymbol{r})\mathrm{e}^{-i\boldsymbol{q}\cdot\boldsymbol{r}} = \frac{1}{\Omega}\sum_{\boldsymbol{k},s} c_{\boldsymbol{k},s}^\dagger c_{\boldsymbol{k}+\boldsymbol{q},s} = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\rho_{\boldsymbol{k},\boldsymbol{q},s} \tag{4.39}$$

where we defined $\rho_{\boldsymbol{k},\boldsymbol{q},s} = c_{\boldsymbol{k},s}^\dagger c_{\boldsymbol{k}+\boldsymbol{q},s}$. The perturbation term may then be rewritten as

$$H_V = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\rho_{\boldsymbol{k},\boldsymbol{q},s}^\dagger V_a(\boldsymbol{q},\omega)\mathrm{e}^{-i\omega t}\mathrm{e}^{\eta t}. \tag{4.40}$$

To describe the time-dependent electron density in the metal, we need to determine the time-dependent density operator $\rho_{\boldsymbol{q}}(t)$ in the Heisenberg representation. Recall that operators are time-independent and states are time-dependent in the Schrödinger picture, whereas operators are time-dependent and states are time-independent in the Heisenberg picture. Operators in these two pictures are related via $A_H(t) = U^\dagger(t)A_S(t)U(t)$ where $U(t)$ is a time-evolution operator and $A_S(t)$ is the operator in the Schrödinger picture. Thus, $A_H(t)$ gains a time-dependence even if A_S has no explicit time-dependence. The equation of motion determining the time-evolution of an operator $A_H(t)$ in the Heisenberg picture is generally

$$\frac{d}{dt}A_H(t) = \frac{i}{\hbar}[H_H(t), A_H(t)] + \Big(\frac{\partial A_S}{\partial t}\Big)_H. \tag{4.41}$$

Here, the density operator has no explicit time-dependence, so the partial derivative above is zero. We thus have the following Heisenberg equation of motion for the density operator (dropping the H subscript):

$$\begin{aligned} i\hbar\frac{d}{dt}\rho_{\boldsymbol{k},\boldsymbol{q},s} &= [\rho_{\boldsymbol{k},\boldsymbol{q},s}, H] \\ &= (\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}})\rho_{\boldsymbol{k},\boldsymbol{q},s} + (c_{\boldsymbol{k},s}^\dagger c_{\boldsymbol{k},s} - c_{\boldsymbol{k}+\boldsymbol{q},s}^\dagger c_{\boldsymbol{k}+\boldsymbol{q},s})V_a(\boldsymbol{q},\omega)\mathrm{e}^{-i\omega t}\mathrm{e}^{\eta t}. \end{aligned} \tag{4.42}$$

We now compute the thermal average (expectation value) of the above operator equation according to

$$\langle\hat{A}\rangle = \frac{\mathrm{Tr}[\hat{A}\mathrm{e}^{-\beta H}]}{\mathrm{Tr}(\mathrm{e}^{-\beta H})} \tag{4.43}$$

and following the linear response scheme by assuming the same time dependence for $\langle\rho_{\boldsymbol{k},\boldsymbol{q},s}(t)\rangle \propto \mathrm{e}^{-i\omega t+\eta t}$ as for the potential (there can still exist a constant phase-offset). The equation of motion then reads

$$(\hbar\omega + i\hbar\eta)\langle\rho_{\boldsymbol{k},\boldsymbol{q},s}\rangle = (\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}})\rho_{\boldsymbol{k},\boldsymbol{q},s} + (n_{0\boldsymbol{k},s} - n_{0\boldsymbol{k}+\boldsymbol{q},s})V_a(\boldsymbol{q},\omega) \tag{4.44}$$

where $n_{0\boldsymbol{k},s} = \langle c_{\boldsymbol{k},s}^\dagger c_{\boldsymbol{k},s}\rangle$ and therefore

$$\delta n_{\rm ind}(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\langle\rho_{\boldsymbol{k},\boldsymbol{q},s}\rangle = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\frac{n_{0\boldsymbol{k}+\boldsymbol{q},s} - n_{0\boldsymbol{k},s}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega - i\hbar\eta}V_a(\boldsymbol{q},\omega) \tag{4.45}$$

The dynamical linear response function which determines the induced electron density due to the external perturbation can then be defined as

$$\delta n_{\rm ind}(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)V_a(\boldsymbol{q},\omega) \tag{4.46}$$

where we thus have derived that

$$\chi_0(\boldsymbol{q},\omega) = \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\frac{n_{0\boldsymbol{k}+\boldsymbol{q},s} - n_{0\boldsymbol{k},s}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega - i\hbar\eta}. \tag{4.47}$$

The response function $\chi_0(\boldsymbol{q},\omega)$ is known as the Lindhard function. The Lindhard function thus expresses the charge response of a free electron gas to an external electric potential perturbation. This is closely related to so-called electron-hole excitations, which we will cover in more detail in the next section, since the major contributions to χ_0 are seen to come from $\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \hbar\omega = 0$ which corresponds exactly to a electron-hole excitation: the Fermi sea is imparted a little bit of energy ω which promotes a particle from $\boldsymbol{k}$ (thus creating a *hole*) to a state at $\boldsymbol{k}+\boldsymbol{q}$ (thus creating an electron).

In computing the linear response of the system due to an external perturbation, we have neglected any feedback effect due to the interaction among the electrons. In a more accurate treatment, the density fluctuation $\delta n(\boldsymbol{r},t)$ can be thought of as a source for an additional Coulomb potential $V_{\delta n}$. This potential can be determined by means of the Poisson equation arising from Maxwell's equation $\nabla\cdot\boldsymbol{E} = \rho/\epsilon_0$:

$$\nabla^2 V_{\delta n}(\boldsymbol{r},t) = -\frac{e^2}{\epsilon_0}\delta n_{\text{ind}}(\boldsymbol{r},t). \tag{4.48}$$

In Fourier-space, this reads

$$V_{\delta n}(\boldsymbol{q},\omega) = \frac{e^2}{\epsilon_0 q^2}\delta n_{\text{ind}}(\boldsymbol{q},\omega). \tag{4.49}$$

If we thus allow feedback effects in our system with the external perturbation V_a, the effective potential V experienced by the electrons is determined self-consistently via

$$\begin{aligned} V(\boldsymbol{q},\omega) &= V_a(\boldsymbol{q},\omega) + V_{\delta n}(\boldsymbol{q},\omega) \\ &= V_a(\boldsymbol{q},\omega) + \frac{e^2}{\epsilon_0 q^2}\delta n_{\text{ind}}(\boldsymbol{q},\omega) \end{aligned} \tag{4.50}$$

where we write

$$\delta n(\boldsymbol{q},\omega) = \tilde{\chi}(\boldsymbol{q},\omega)V_a(\boldsymbol{q},\omega) \tag{4.51}$$

Here, we have used our previous result Eq. (4.46) for the induced electron density but written the response function as $\tilde{\chi}$. This is because since there is now an extra term $V_{\delta n}(\boldsymbol{r},t)$ in the Hamiltonian, which couples to the electrons and depends on the induced density itself (hence the necessity for a self-consistent approach), the response function is in general different from the free-electron response function χ_0.

At this point, we can make an approximation that $\tilde{\chi} = \chi_0$. In effect, we are replacing the proper response function with the non-interacting response function. Then, the relation between V and V_a may be written as

$$V(\boldsymbol{q},\omega) = \frac{V_a(\boldsymbol{q},\omega)}{\varepsilon_r(\boldsymbol{q},\omega)} \tag{4.52}$$

where

$$\varepsilon_r(\boldsymbol{q},\omega) = 1 - \frac{e^2}{\epsilon_0 q^2}\chi_0(\boldsymbol{q},\omega). \tag{4.53}$$

Here, $\varepsilon_r(\boldsymbol{q},\omega)$ is the dynamical dielectric function (momentum and frequency dependent relative permittivity) and describes how the external potential is renormalized due to the dynamical response of the electrons in the metal. We can also write expression Eq. (4.51) in an alternative way so that the information about electron-electron interactions is contained in the response function itself:

$$\delta n(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)V(\boldsymbol{q},\omega) = \chi(\boldsymbol{q},\omega)V_a(\boldsymbol{q},\omega). \tag{4.54}$$

The response function $\chi(\boldsymbol{q},\omega)$ defined above is then

$$\chi(\boldsymbol{q},\omega) = \frac{\chi_0(\boldsymbol{q},\omega)}{1-\frac{e^2}{\epsilon_0 q^2}\chi_0(\boldsymbol{q},\omega)} \tag{4.55}$$

and contains information about both the renormalization of the electric potential as well as the excitation spectrum of the metal. This way of computing the dynamical response function $\chi(\boldsymbol{q},\omega)$ is known as the random phase approximation (RPA). The key approximation we made was thus to set the proper response function equal to the non-interacting one.

We briefly comment on the origin of the name of this approximation. The above equation can be written in the form of a geometric series

$$\chi(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)\Big[1+\frac{e^2}{\epsilon_0 q^2}\chi_0(\boldsymbol{q},\omega)+\Big(\frac{e^2}{\epsilon_0 q^2}\chi_0(\boldsymbol{q},\omega)\Big)^2+\dots\Big]. \tag{4.56}$$

From a perturbation theory perspective, this series corresponds to summing a limited subset of perturbative terms (which can be represented as Feynman diagrams) to infinite order. In effect, a more accurate treatment of $\chi(\boldsymbol{q},\omega)$ would yield even more terms than written in the above expansion. The particular summation over terms above is known as the RPA and is based on the assumption that the phase relation between some of the terms entering the perturbation series are random, meaning that interference terms vanish on average.

4.1.4 Electron-hole excitation

To describe simple particle-hole excitations in a metal, we may neglect the Coulomb interaction between electrons and study the bare response function $\chi_0(\boldsymbol{q},\omega)$. First, separate χ_0 into a real and imaginary part

$$\chi_0 = \chi_{01}(\boldsymbol{q},\omega) + i\chi_{02}(\boldsymbol{q},\omega). \tag{4.57}$$

Now use the Sokhotski-Plemelj theorem which states that when integrating over z, the following relation holds

$$\lim_{\eta\to 0^+}\frac{1}{z-i\eta} = \mathcal{P}\Big(\frac{1}{z}\Big)+i\pi\delta(z) \tag{4.58}$$

where $\mathcal{P}$ is Cauchy's principal value. The Lindhard function then is split into

$$\begin{aligned}\chi_{01}(\boldsymbol{q},\omega) &= \frac{1}{\Omega}\sum_{\boldsymbol{k},s}\mathcal{P}\Big(\frac{n_{0,\boldsymbol{k}+\boldsymbol{q}}-n_{0,\boldsymbol{k}}}{\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}-\hbar\omega}\Big),\\ \chi_{02}(\boldsymbol{q},\omega) &= \frac{1}{\Omega}\sum_{\boldsymbol{k},s}(n_{0,\boldsymbol{k}+\boldsymbol{q}}-n_{0,\boldsymbol{k}})\delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}}-\epsilon_{\boldsymbol{k}}-\hbar\omega).\end{aligned} \tag{4.59}$$

The real part will get more attention in a later subsection when discussing instabilities of metals. We focus here on the imaginary part which expresses the absorption of energy by electrons subject to a time-dependent external perturbation. In particular, note that χ_{02} corresponds to Fermi's golden rule known from time-dependent perturbation theory, i.e. the transition rate from the ground state to an excited state of energy $\hbar\omega$ and momentum $\boldsymbol{q}$.

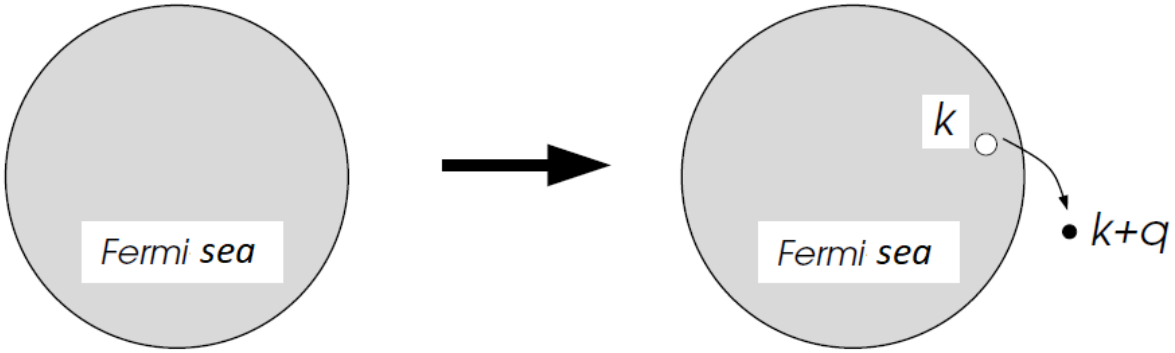

Figure 4.2: Schematic view of an particle-hole pair creation (electron-hole excitation). Figure taken from [3].

As mentioned earlier, the relevant excitations described by the Lindhard function are particle-hole excitations. Starting from the ground state of a completely filled Fermi sea, one electron with momentum $\boldsymbol{k}$ is removed and inserted again outside of the Fermi sea in some state with momentum $\boldsymbol{k}+\boldsymbol{q}$ (see Fig. 4.2).

The excitation energy is:

$$E_{\boldsymbol{k},\boldsymbol{q}} = \left[\frac{\hbar^2(\boldsymbol{k}+\boldsymbol{q})^2}{2m} - \epsilon_F\right] + \left[\epsilon_F - \frac{\hbar^2\boldsymbol{k}^2}{2m}\right] = \frac{\hbar^2\boldsymbol{q}^2}{2m} + \frac{\hbar^2\boldsymbol{k}\cdot\boldsymbol{q}}{m}. \tag{4.60}$$

The maximum value that $|\boldsymbol{k}|$ can have is k_F. The energy above is then maximized if $\boldsymbol{k} \parallel \boldsymbol{q}$, while it is minimized if $\boldsymbol{k} \parallel -\boldsymbol{q}$. Thus, the spectrum is for positive q limited according to $E_{\min} \leq E(\boldsymbol{k},\boldsymbol{q}) \leq E_{\max}$ where

$$E_{\min/\max} = \frac{\hbar^2}{q^2}2m \mp \frac{\hbar^2 q k_F}{m}. \tag{4.61}$$

This is shown by the shaded region in the left part of Fig. 4.3, which is the particle-hole excitation continuum. The leftmost parabola of the shaded region is $E_{\max}$ whereas the rightmost parabola is $E_{\min}$, and the parabolic shape of these boundaries follows by completing the square in Eq. (4.61) and plotting it as a function of q at a fixed k_F-value.

4.1.5 Kramers-Kronig relation for response functions

Response functions to external stimuli share common properties, which we describe in this section. The goal of response theory is to figure out how a system reacts to outside influences. These outside influences are things like applied electric and magnetic fields, or applied pressure, or an applied driving force. This is metaphorically like sticking a spoon into a quantum liquid and stirring. Our

purpose here is to explore the more general case of time dependent influences. As we will ll see, by studying the response of the system at different frequencies, we learn important information about what is going on inside the system itself.
Consider a simple classical dynamical system with some generalized coordinates $x_i(t)$ which depend on time. If left alone, these coordinates will obey some equations of motion:

$$\ddot{x}_i + g_i(\dot{x}, x) = 0. \tag{4.62}$$

The outside influence in this example arises from perturbing the system by the addition of some driving forces $F_i(t)$, so that the equations of motion become

$$\ddot{x}_i + g_i(\dot{x}, x) = F_i(t). \tag{4.63}$$

In this expression, $x_i(t)$ are dynamical degrees of freedom. This is what we are solving for. In contrast, $F_i(t)$ are not dynamical: they are forces that are under our control, like someone pulling on the end of a spring. We get to decide on the time dependence of each $F_i(t)$.

Instead, in quantum mechanics we introduce the outside influences in a slightly different manner. The observables of the system are now operators, $\mathcal{O}_i$. We will work in the Heisenberg picture, so that the operators are time dependent: $\mathcal{O} = \mathcal{O}(t)$. Left alone, the dynamics of these operators will be governed by a Hamiltonian $H(\mathcal{O})$. However, we have no interest in leaving the system alone. We want to give it a kick. Mathematically, this is achieved by adding an extra term to the Hamiltonian

$$H_{\text{source}}(t) = \phi_i(t)\mathcal{O}_i(t). \tag{4.64}$$

The $\phi_i(x)$ are referred to as sources. They are external fields that are under our control, analogous to the driving forces in the example above. Indeed, if we take a classical Hamiltonian and add a term of the form x then the resulting Euler-Lagrange equations include the source on the right-hand-side in the same way that the force appears in Eq. (4.63).

Linear response

We want to understand how our system reacts to the presence of the source or the driving force. To be concrete, we'll chose to work in the language of quantum mechanics, but everything that we discuss in this section will also carry over to classical systems. Our goal is to understand how the correlation functions of the theory change when we turn on a source (or sources) $\phi_i(x)$.

In general, it's a difficult question to understand how the theory is deformed by the sources. To figure this out, we really just need to sit down and solve the theory all over again. However, we can make progress under the assumption that the source is a small perturbation of the original system. This is fairly restrictive but it's the simplest place where we can make progress so, from now on, we focus on this limit. Mathematically, this means that we assume that the change in the expectation value of any operator is linear in the perturbing source. We write

$$\delta\langle\mathcal{O}_i(t)\rangle = \int dt' \chi_{ij}(t,t')\phi_j(t'). \tag{4.65}$$

Here, $\chi_{ij}(t,t')$ is known as a *response function*. We could write a similar expression for the classical dynamical system Eq. (4.63) where $\delta\langle\mathcal{O}_i\rangle$ where is replaced by $x_i(t)$ and ϕ is replaced by the driving force $F_j(t)$. In classical mechanics, it is clear from the form of the equation of motion Eq. (4.63) that

the response function is simply the Green function [2] for the system. For this reason, the response functions are often called Green functions and you will often see them denoted as G instead of χ.

From now on, assume that our unperturbed system is invariant under time translations. Then, we must have

$$\chi_{ij}(t,t') = \chi_{ij}(t-t') \tag{4.71}$$

and it is useful to perform a Fourier transform to work in frequency space. We define the Fourier transform of the function $f(t)$ to be

$$f(\omega) = \int dt \mathrm{e}^{i\omega t} f(t) \text{ and } f(t) = \int \frac{d\omega}{2\pi} \mathrm{e}^{-i\omega t} f(\omega). \tag{4.72}$$

In effect, we are using here the convention that the function and its Fourier-transform are distinguished by looking at their argument. In general, $f(\omega)$ is not the same function as $f(t)$ only with t replaced by ω, but should be denoted $\tilde{f}(\omega)$, but as explained above we simplify the notation here.

Taking the Fourier-transformation of Eq. (4.65) gives

$$\begin{aligned}
\delta\langle \mathcal{O}_i(\omega)\rangle &= \int dt' \int dt \mathrm{e}^{i\omega t} \chi_{ij}(t-t')\phi_j(t') \\
&= \int dt' \int dt \mathrm{e}^{i\omega(t-t')} \chi_{ij}(t-t') \mathrm{e}^{i\omega t'} \phi_j(t') \qquad (4.73)\\
&= \chi_{ij}(\omega)\phi_j(\omega). \qquad (4.74)
\end{aligned}$$

We see that the linear response is local in frequency space: if you shake something a frequency ω, it responds at frequency ω, just like we saw in our perturbative treatment of the electron gas in response to a time-dependent potential. Anything beyond this lies within the domain of nonlinear response.

[2] A Green function $G(x,s)$ of a linear differential operator $L = L(x)$ is any solution of

$$LG(x,s) = \delta(s-x). \tag{4.66}$$

Here, δ is the Dirac delta-function. This property of a Green function can be exploited to solve differential equations of the form

$$Lu(x) = f(x). \tag{4.67}$$

To see this, multiply $LG(x,s) = \delta(s-x)$ with $f(s)$ and integrate with respect to s. We obtain

$$\int LG(x,s)f(s)ds = \int \delta(x-s)f(s)ds = f(x). \tag{4.68}$$

Since $L = L(x)$ is a linear operator and only acts on the variable x, we may take it outside of the integration:

$$L\Big(\int G(x,s)f(s)ds\Big) = f(x). \tag{4.69}$$

This means that

$$u(x) = \int G(x,s)f(s)ds \tag{4.70}$$

is the solution to the equation $Lu(x) = f(x)$.

In the following section we will describe some of the properties of the response function $\chi(\omega)$ and how to interpret them. Many of these properties follow from very simple physical input. To avoid clutter, we will mostly drop both the i, j indices. When there is something interesting to say about them, we will put them back in.

Analyticity and causality

Working with a real source ϕ and a Hermitian operator $\mathcal{O}$ (which means a real expectation value $\langle\mathcal{O}\rangle\rangle$), then it follows that $\chi(t)$ must also be real. Let us see what this means for the Fourier transform $\chi(\omega)$. It's useful to introduce some new notation for the real and imaginary parts:

$$\begin{aligned}\chi(\omega) &= \mathrm{Re}\chi(\omega) + i\mathrm{Im}\chi(\omega) \\ &\equiv \chi'(\omega) + i\chi''(\omega).\end{aligned} \tag{4.75}$$

The real and imaginary parts of the response function $\chi(\omega)$ have different interpretations. Let us look at these in turn.

Imaginary part
We can write the imaginary part as

$$\begin{aligned}\chi''(\omega) &= -\frac{i}{2}[\chi(\omega) - \chi^*(\omega)] \\ &= -\frac{i}{2}\int_{-\infty}^{\infty} dt\chi(t)[\mathrm{e}^{i\omega t} - \mathrm{e}^{-i\omega t}] \\ &= -\frac{i}{2}\int_{-\infty}^{\infty} dt\mathrm{e}^{i\omega t}[\chi(t) - \chi(-t)].\end{aligned} \tag{4.76}$$

We see that the imaginary part of $\chi(\omega)$ is due to the part of the response function which is *not* invariant under time-reversal $t \to -t$. In other words, $\chi''(\omega)$ knows about the arrow of time. This suggests that χ'' should be related to dissipative processes in the system, since such processes are not invariant under time-reversal. Indeed, $\chi''(\omega)$ is called the dissipative or absorptive part of the response function. Note that it follows from the above equation that $\chi''(\omega)$ is an odd function: $\chi''(\omega) = -\chi''(-\omega)$.

Real part
The same analysis as above shows that

$$\chi'(\omega) = \frac{1}{2}\int_{-\infty}^{\infty} dt\mathrm{e}^{i\omega t}[\chi(t) + \chi(-t)]. \tag{4.77}$$

In effect, the real part stems from the part of the correlation function that does not care about the arrow of time. It is called the reactive part of the response function, and is even: $\chi'(\omega) = \chi'(-\omega)$. Before moving on, we briefly mention what happens when we put the labels i, j back on the response functions. In this case, a similar analysis to that above shows that the dissipative response function comes from the anti-Hermitian part:

$$\chi''_{ij}(\omega) = -\frac{i}{2}[\chi_{ij}(\omega) - \chi^*_{ji}(\omega)]. \tag{4.78}$$

An important point which is unfortunately often not mentioned in lecture notes discussing the interpretation of the response function is that the imaginary part of $\chi(\omega)$ is the dissipative response

only when the observable and applied stimulus have the same parity with respect to time-reversal. If they have opposite parity, then it is the real part of $\chi(\omega)$ which describes the dissipative response. Often, one the observable and stimulus have the same parity. This is the case for the magnetization (observable) induced by an magnetic field (stimulus): both are odd under time-reversal. It is also the case for the electric polarization (observable) induced by an electric field (stimulus): both are even under time-reversal.

But the observable and stimulus can have opposite parity. This is the case for conductivity σ. Consider the linear-response equations for polarization density $\boldsymbol{P}$ and current density $\boldsymbol{J}$, both induced by an electric field $\boldsymbol{E}$:

$$\begin{aligned}\boldsymbol{P}(t) &= \epsilon_0 \int dt' \chi(t-t')\boldsymbol{E}(t'),\\ \boldsymbol{J}(t) &= \int \sigma(t-t')\boldsymbol{E}(t')dt'.\end{aligned} \tag{4.79}$$

Here, we have for simplicity assumed that the response is along the same direction as the stimuli, whereas χ in general will be a tensor instead of a scalar. The current $\boldsymbol{J}$ is odd under time-reversal, but $\boldsymbol{E}$ is not. This causes a $\pi/2$ phase-shift (in effect, factor i) between the response functions σ and $\epsilon = \epsilon_0\chi$ in terms of how they affect a given physical quantity. To see an example of this, consider Ampere's law

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t}. \tag{4.80}$$

Now Fourier-transform this equation and use that a time-derivative amounts to multiplying with $i\omega$. Then express $\boldsymbol{J}$ and $\boldsymbol{D}$ in terms of the Fourier-transforms of the response functions: $\boldsymbol{J}(\omega) = \sigma(\omega)\boldsymbol{E}(\omega)$ and $\boldsymbol{D}(\omega) = \epsilon_0\boldsymbol{E}(\omega) + \boldsymbol{P}(\omega) = \epsilon(\omega)\boldsymbol{E}(\omega)$. This gives

$$\nabla \times \boldsymbol{H} = (\sigma + i\omega\epsilon)\boldsymbol{E}. \tag{4.81}$$

Now we can see what was stated above. An electric field generates $\boldsymbol{H}$ in an electromagnetic wave through both σ and ϵ, but they are shifted by $\pi/2$. So if the imaginary part of ϵ describes dissipation in this process, then so does the real part of σ.

Causality

We cannot affect the past. This statement of causality means that any response function should satisfy

$$\chi(t) = 0 \text{ for all } t < 0. \tag{4.82}$$

For this reason, χ is typically referred to as the retarded Green's function and is sometimes denoted $G_R(t)$. Let us see what this simply causality requirement means for the Fourier-expansion of χ:

$$\chi(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \mathrm{e}^{-i\omega t}\chi(\omega). \tag{4.83}$$

When $t < 0$, we can perform the integral by completing the contour in the upper-half place so that the exponent becomes $-i\omega \times (-i|t|) \to -\infty$. The answer has to be zero. Of course, the integral is given by the sum of the residues inside the contour. So if we want the response function to vanish for all $t < 0$, it must be that $\chi(\omega)$ has no poles in the upper-half plane. In other words, causality requires:

$$\chi(\omega) \text{ is analytic for } \mathrm{Im}\omega > 0. \tag{4.84}$$

Kramers-Kronig relation

The fact that χ is analytic in the upper-half plane means that there is a relationship between the real and imaginary parts, χ' and χ''. This is called the Kramers-Kronig relation. Our task in this section is to derive it.
Given any analytical function χ in the closed upper half-plane, the function $\chi(\omega')/(\omega'-\omega)$ where ω is real, will be analytic in the open upper half-plane (in effect, excluding the real line). The residue theorem consequently states that

$$\oint \frac{\chi(\omega)}{\omega'-\omega}d\omega' = 0 \tag{4.85}$$

for any closed contour within this region of the complex ω' plane. Choose now this contour to trace the real axis, a semicircular hump over the pole at $\omega' = \omega$ with a radius that we set to be infinitesimal, and a large semicircle in the upper half-plane. We now look at the contribution to the above integral along each of these three contour segments. The length of the semicircular segment increases proportionally to $|\omega'|$, but the integral over it vanishes as the radius of the semicircle is taken to the limit $\to \infty$. This is because $\chi(\omega')/(\omega'-\omega)$ vanishes faster than $1/|\omega'|$ so long that $\chi(\omega' \to \infty) \to 0$. It is reasonable that the response function should vanish at $\omega \to \infty$, since the system will not be able to respond to a stimuli that changes infinitely rapidly. We are then left with the segments along the real axis and the half-circle around the pole. The contribution from the half-circle C with infinitesimal radius around $\omega' = \omega$ is evaluated by using first

$$\int_C \frac{d\omega'\chi(\omega')}{\omega'-\omega} = \chi(\omega)\int_C \frac{d\omega'}{\omega'-\omega} \tag{4.86}$$

This is ok since χ will be constant on the infinitely small semicircle. Then, we substitute $\omega' = re^{i\phi}$ where r is infinitesimal and integrate over ϕ:

$$\chi(\omega)\int_C \frac{d\omega'}{\omega'-\omega} = \chi(\omega)\int_\pi^0 \frac{ire^{i\phi}d\phi}{re^{i\phi}} = -i\pi\chi(\omega). \tag{4.87}$$

Overall, we have therefore found:

$$0 = \oint \frac{\chi(\omega)}{\omega'-\omega}d\omega' = \mathcal{P}\int_{-\infty}^{\infty}\frac{\chi(\omega)}{\omega'-\omega}d\omega' - i\pi\chi(\omega). \tag{4.88}$$

Rearranging, we arrive at the compact form

$$\chi(\omega) = \frac{1}{i\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi(\omega)}{\omega'-\omega}d\omega'. \tag{4.89}$$

The i in the denominator is crucial. It means that we can write, by taking the real and imaginary parts,:

$$\mathrm{Re}\chi(\omega) = \mathcal{P}\int_{-\infty}^{\infty}\frac{d\omega'}{\pi}\frac{\mathrm{Im}\chi(\omega')}{\omega'-\omega} \tag{4.90}$$

and

$$\mathrm{Im}\chi(\omega) = -\mathcal{P}\int_{-\infty}^{\infty}\frac{d\omega'}{\pi}\frac{\mathrm{Re}\chi(\omega')}{\omega'-\omega} \tag{4.91}$$

These are the Kramers-Kronig relations. They follow from causality and a vanishing response function at $\omega \to \infty$. They tell us that the dissipative, imaginary part of the response function $\chi''(\omega)$ is determined in terms of the reactive, real part, $\chi'(\omega)$ and vice-versa. However, the relationship is not local in frequency space: you need to know $\omega'(\omega)$ for all frequencies in order to reconstruct χ'' for any single frequency. In effect, if you know either the real or imaginary part of χ, then you can construct the entire response function $\chi(\omega)$.

4.2 Plasmons

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

Assuming small momentum $|\boldsymbol{q}| \ll k_F$, the goal is to expand $\chi(\boldsymbol{q},\omega) = \chi_0(\boldsymbol{q},\omega)/\varepsilon_r(\boldsymbol{q},\omega)$ which we found earlier. This means we must find the low-momentum (long wavelength) expansion of χ_0 and ε. To do so, note that

$$\begin{aligned} \epsilon_{\boldsymbol{k}+\boldsymbol{q}} &= \epsilon_{\boldsymbol{k}} + \boldsymbol{q}\cdot\nabla_{\boldsymbol{k}}\epsilon_{\boldsymbol{k}} + \dots \\ n_{0,\boldsymbol{k}+\boldsymbol{q}} &= n_{0,\boldsymbol{k}} + \frac{\partial n_0}{\partial\epsilon}\boldsymbol{q}\cdot\nabla_{\boldsymbol{k}}\epsilon_{\boldsymbol{k}} + \dots \end{aligned} \tag{4.92}$$

Note that the Fermi-Dirac distribution function satisfies $\partial n_0/\partial\epsilon_{\boldsymbol{k}} = -\delta(\epsilon_{\boldsymbol{k}} - \epsilon_F)$ at $T = 0$ and $\nabla_{\boldsymbol{k}}\epsilon_{\boldsymbol{k}} \equiv \hbar\boldsymbol{v}_{\boldsymbol{k}}$ is the velocity. Focusing on states located near the Fermi energy, $\boldsymbol{v}_{\boldsymbol{k}} = v_F\boldsymbol{k}/|\boldsymbol{k}|$ is the Fermi velocity. We then obtain for the bare response function

$$\begin{aligned} \chi_0(\boldsymbol{q},\omega) &\simeq -2\int \frac{d^3k}{(2\pi)^3}\frac{\boldsymbol{q}\cdot\boldsymbol{v}_F\delta(\epsilon_{\boldsymbol{k}}-\mu)}{\boldsymbol{q}\cdot\boldsymbol{v}_F - \omega - i\eta} \\ &= \frac{2}{(2\pi)^2}\int_{-1}^{1} d(\cos\theta)\frac{k_F^2}{\hbar v_F}\Big[\frac{qv_F\cos\theta}{\omega+i\eta} + \Big(\frac{qv_F\cos\theta}{\omega+i\eta}\Big)^2 + \Big(\frac{qv_F\cos\theta}{\omega+i\eta}\Big)^3 + \dots\Big] \\ &\simeq \frac{k_F^3q^2}{3\pi^2m(\omega+i\eta)^2}\Big(1+\frac{3}{5}\frac{v_F^2q^2}{(\omega+i\eta)^2}\Big) \\ &= \frac{n_0q^2}{m(\omega+i\eta)^2}\Big(1+\frac{3}{5}\frac{v_F^2q^2}{(\omega+i\eta)^2}\Big). \end{aligned} \tag{4.93}$$

Since we previously derived $\epsilon_r(\boldsymbol{q},\omega) = 1 - \frac{e^2}{4\pi q^2\varepsilon_0}\chi_0(\boldsymbol{q},\omega)$, we get

$$\lim_{|\boldsymbol{q}|\to 0}\frac{\varepsilon(\boldsymbol{q},\omega)}{\varepsilon_0} = 1 - \frac{\omega_p^2}{\omega^2}\ (= \varepsilon_r,\ \text{relative electric permittivity}) \tag{4.94}$$

where

$$\omega_p^2 = \frac{e^2n_0}{\varepsilon_0 m}. \tag{4.95}$$

This allows us to finally approximate the response function that takes into account the Coulomb

interaction: $\chi(\boldsymbol{q},\omega)$. We get

$$\chi(\boldsymbol{q},\omega) \simeq \frac{n_0 q^2 [R(q,\omega)]^2}{(\omega+i\eta)^2 - e^2 n_0^2 R(q,\omega)/\epsilon_0} = \frac{n_0 q^2 R(q,\omega)}{2m\omega_p}\Big[\frac{1}{\omega+i\eta-\omega_p R(q,\omega)} - \frac{1}{\omega+i\eta+\omega_p R(q,\omega)}\Big]. \tag{4.96}$$

Here, we introduced

$$[R(q,\omega)]^2 = 1 + \frac{3v_F^2 q^2}{5\omega^2}. \tag{4.97}$$

Applying Eq. (4.58), we get

$$\mathrm{Im}\{\chi(\boldsymbol{q},\omega)\} \simeq \frac{\pi n_0 R(q,\omega_p)}{\omega_p}\Big[\delta(\omega-\omega_p R(q,\omega_p)) - \delta(\omega+\omega_p R(q,\omega_p))\Big]. \tag{4.98}$$

From this, it is clear that a sharp excitation modes exist when the argument of the delta-functions vanish:

$$\omega(\boldsymbol{q}) = \omega_p R(q,\omega_p) = \omega_p\Big[1 + \frac{3v_F^2 q^2}{10\omega_p^2} + \dots\Big], \tag{4.99}$$

which is called a *plasma resonance* with ω_p as the *plasma frequency*. Similar to the exciton considered in the semiconducting case, the plasma excitation has a well-defined energy-momentum relation (called dispersion relation) and may consequently be viewed as as a quasiparticle called the *plasmon*. When the plasmon dispersion merges with the electron-hole continuum (see left part of Fig. 4.3), it is gradually damped (called Landau dampaing) since it can decay into the electron-hole excitations. This results in a finite lifetime of the plasmons within the electron-hole continuum, which corresponds to a finite width of the resonance of the collective excitation when measuring the response function. In effect, it isn't infintely sharp as the δ-function of Eq. (4.98), but becomes smeared. Experimental values for the plasma frequency of different metals are shown in Table 4.2 [taken from Ref. [3]], in addition to the predicted ω_p value according to Eq. (4.95) where the free electron mass was used for m and n determined through $r_{s,\mathrm{Li}} = 3.22$, $r_{s,\mathrm{Na}} = 3.96$, $r_{s,\mathrm{K}} = 4.86$.

Metal	$\omega_p^{(\mathrm{exp})}$ [eV]	$\omega_p^{(\mathrm{theo})}$ [eV]
Li	7.1	8.5
Na	5.7	6.2
K	3.7	4.6

Table 4.1: Measured and predicted values for the plasma frequency for different metals [3].

The plasmon excitation can also be considered from a classical perspective. Consider negatively charged electrons in a positively charged ionic background. This is a *plasma*: a medium with an equal concentration of positive and negative charges where at least one charge type is mobile (in our case, the negatively charged electrons moving against the fixed positive charge of ion cores). Now displace the negative charge upward by a small distance u, creating a negative surface charge density $\sigma = -neu$ at the top and a negative charge density $-\sigma$ at the bottom (see right part of Fig. 4.3. Using the integral version of Gauss law, we find that the electric field inside the slab is $E = neu/\epsilon_0$ pointing upwards. The equation of motion for an electron in the system then takes the form:

$$m\frac{d^2u}{dt^2} = -eE. \tag{4.100}$$

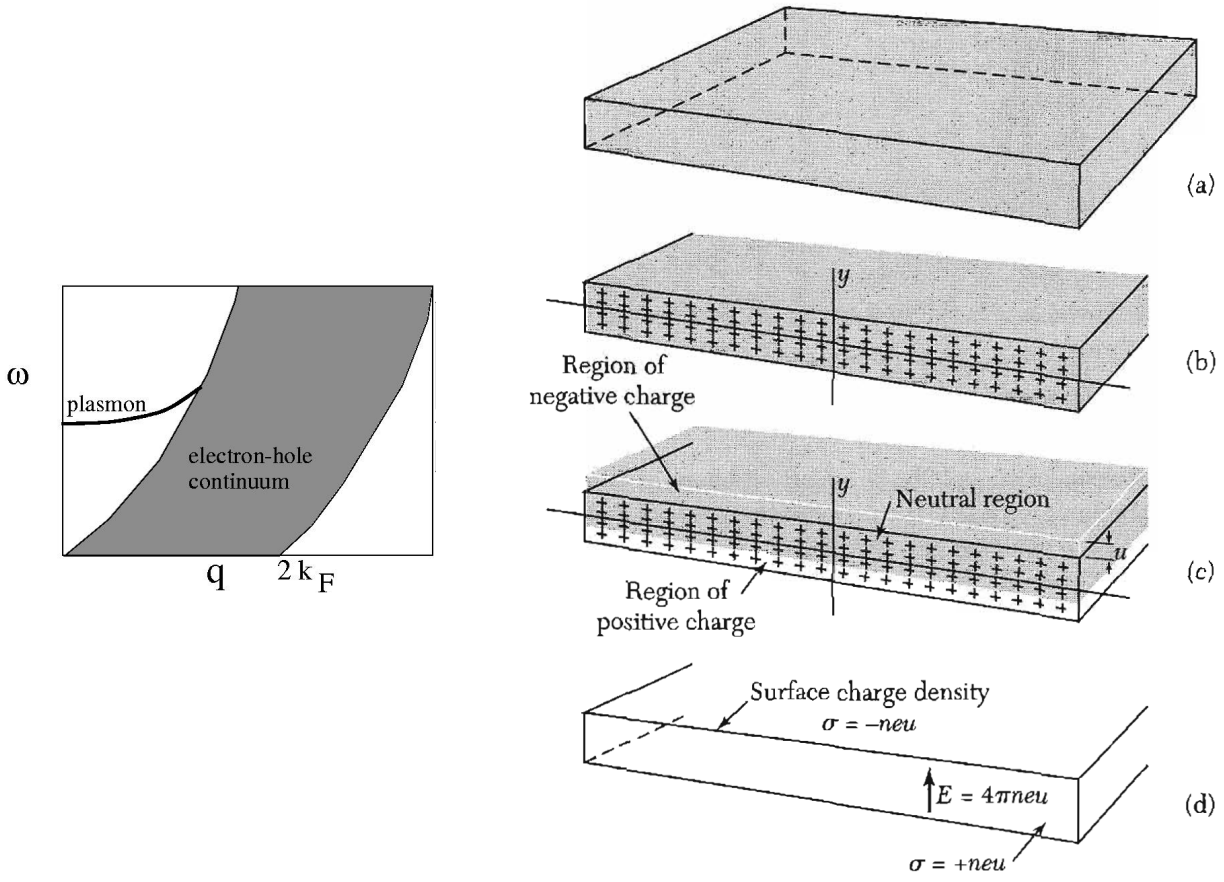

Figure 4.3: **Left:** schematic view of an particle-hole pair creation (electron-hole excitation). Figure taken from [3]. **Right:** (a) shows a thin slab or film of a metal. A cross section is shown in(b), and the whole slab is electricall neutral. In (c), the negative charge has been displaced upward uniformly by a small distance u. In (d), this displacement establisheds a surface charge density $-neu$ on the upper surface and opposite in the lower where n is the electron density. An electric field $E = neu/\epsilon_0$ is produced inside the slab, which tries to restore the electron sea to its equilibrium position. Figure taken Ref. [7].

This is a simple harmonic oscillator equation of motion with a frequency of precisely $\omega_p = ne^2/(\varepsilon_0 m)$: the plasma frequency. The electric field tries to restore the equilibrium configuration of the electron density. The plasma oscillation (of which the plasmon excitation is the quantum mechanical version) can therefore be thought of as a longitudinal oscillation of the whole electron gas cloud on top of a positively charged background. Based on this picture, it is clear that the plasmon is a different type of excitation than a single electron moving from inside the Fermi sea to outside of it. Instead, the plasmon is a *collective excitation* involving the whole electron gas cloud and can be shown to have a bosonic character rather than fermionic nature.

We can also derive an expression for the permittivity $\varepsilon(\omega)$ when a time-dependent electric field is applied to the above system. Assume that we are applying an oscillating electric field $E \propto \mathrm{e}^{-i\omega t}$, which according the above equation means that the displacement u should have the same time-

dependence. We get

$$-\omega^2 mu = -eE \rightarrow u = eE/m\omega^2. \tag{4.101}$$

The dipole moment generated by a single electron (against the ionic background) is $-eu = -e^2E/m\omega^2$. The polarization, defined as the dipole moment per unit volume, is then

$$P = -neu = -\frac{ne^2E}{m\omega^2} \tag{4.102}$$

where n is the electron density. Now, the electric permittivity at frequency ω is generally given by

$$\frac{\varepsilon(\omega)}{\varepsilon_0} = 1 + \frac{P(\omega)}{\varepsilon_0 E(\omega)}. \tag{4.103}$$

Inserting the above polarization, we find the permittivity of the free electron gas

$$\frac{\varepsilon(\omega)}{\varepsilon_0} = 1 - \frac{ne^2}{\varepsilon_0 m\omega^2} = 1 - \frac{\omega_p^2}{\omega^2}. \tag{4.104}$$

We see that exactly the same plasma frequency as previously derived appears in this expression. To see the physical interpretation of the system's behavior when $\omega = \omega_p$, we first note that $\varepsilon(\omega_p) = 0$. In a non-magnetic isotropic medium, the electromagnetic wave equation reads

$$\mu_0 \frac{\partial^2 \boldsymbol{D}}{\partial t^2} = \nabla^2 \boldsymbol{E} \tag{4.105}$$

Looking for a plane-wave solution $\boldsymbol{E} = \boldsymbol{E}_0 \mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{r} - i\omega t}$, we obtain the following dispersion relation for the electromagnetic waves

$$\varepsilon(\boldsymbol{q}, \omega)\mu_0\omega^2 = q^2 \tag{4.106}$$

by using that $\boldsymbol{D} = \varepsilon(\boldsymbol{q}, \omega)\boldsymbol{E}$. We see that:

- If ε is real and > 0: for a real ω, q is real and the EM wave propagates.
- If ε is real and < 0: for a real ω, q is imaginary so that the wave is damped with a characteristic length $1/|q|$.

4.2.1 Experimental measurement of plasmon dispersion

The plasma frequency ω_p has special meaning since it marks the separation between positive and negative values of ε. For a metal, if the frequency is lower than ω_p, the real part is negative meaning that the light is completely reflected. The electrons near the surface can screen the electric fields of the light before it gets into the bulk. But if the frequency is higher than the plasmon frequency, the real part is positive and the metal behaves like a dielectric medium.

In general, the plasmon dispersion $\omega(\boldsymbol{q})$ is not simply equal to ω_p, but is momentum-dependent. Then, one can map out experimentally the plasmon energy-momentum relation by measuring how light with various frequencies (energies) and different angles of incidence is reflected, in effect via optical reflectivity measurements.

4.3 Screening

4.3.1 Thomas-Fermi screening

In metals, it is often stated that the Coulomb interaction can be neglected due screening by the free electrons. We now this show mathematically. Consider the potential V experienced by electrons exposed to a static field $\omega \to 0$. Considering the long wavelength limit $\boldsymbol{q} \to 0$, and using the expansion Eq. (4.92), we obtain

$$\chi_0(\boldsymbol{q}, 0) = -\frac{1}{\Omega}\sum_{\boldsymbol{k},s} \delta(\epsilon_{\boldsymbol{k}} - \epsilon_F) = -\frac{1}{\pi^2}\frac{k_F^2}{\hbar v_F} = -\frac{3n_0}{2\epsilon_F}. \tag{4.107}$$

This gives

$$\varepsilon_r(\boldsymbol{q}, 0) = 1 + \frac{k_{\mathrm{TF}}^2}{q^2} \tag{4.108}$$

with the so-called Thomas-Fermi wavevector $k_{\mathrm{TF}}^2 = 6\pi e^2 n_0/\epsilon_F$. The effect caused by the acquired $\boldsymbol{q}$-dependence of the renormalized dielectric function is best understood by considering a bare point charge $V_a(\boldsymbol{r}) = \frac{e^2}{4\pi\varepsilon_0 r}$ which in momentum space reads $V_a(\boldsymbol{q}) = e^2/(4\pi\varepsilon_0 q^2)$. The renormalized version of this in momentum space is according to our previous treatment given by

$$V(\boldsymbol{q}) = \frac{V_a(\boldsymbol{q})}{\varepsilon_r(\boldsymbol{q}, 0)} = \frac{4\pi e^2}{q^2 + k_{\mathrm{TF}}^2}. \tag{4.109}$$

Transforming back to real space, we obtained the desired result: a screened Coulomb potential

$$V(\boldsymbol{r}) = \frac{e^2}{4\pi\varepsilon_0}\mathrm{e}^{-k_{\mathrm{TF}}r}. \tag{4.110}$$

This screening can be visualized as a rearrangement of the electrons, turning the long-ranged Coulomb potential into a Yukawa potential with exponential decay. The length scale of the screening is k_{TF}^{-1}, the so-called Thomas-Fermi screening length. In ordinary metals, k_{TF} is of the same order of magnitude as k_F, meaning it is of order 0.1-1 nm, which is comparable to the distance between neighboring atoms. As a consequence, external electric fields cannot penetrate a metal, but are screened over a distance $1/k_{\mathrm{TF}}$. This legitimates one of the basic assumptions used in electrostatics with metals.

This result can also be derived from Gauss law when placing an external "test" charge inside an electron gas residing on an ionic background. Locally, electrons form a Fermi gas with Fermi energy ϵ_F and electron density $n_e(\epsilon_F)$ which neutralizes the ion charge. Assume that the external charge distribution $\rho_{\mathrm{ex}}(\boldsymbol{r})$ creates an electrostatic potential $\Phi(\boldsymbol{r})$. This potential, in turn, induces a redistribution of the electron charge relative $n_e(\epsilon_F)$. We call this redistribution of charge $\rho_{\mathrm{ind}}(\boldsymbol{r})$. Within the above approximation of how the electrons behave locally, we can then write

$$\rho_{\mathrm{ind}}(\boldsymbol{r}) = -e[n_e(\epsilon_F + e\Phi(\boldsymbol{r})) - n_e(\epsilon_F)] \tag{4.111}$$

where

$$n(\epsilon_F) = \frac{k_F^3}{(3\pi^2)} \tag{4.112}$$

and $\epsilon_F = \hbar^2 k_F^2/2m$. This approach should be reasonable so long that the spatial variation of the potential $\Phi(\boldsymbol{r})$ is slow compared to k_F^{-1}, so that we can indeed locally describe the electron gas as a filled Fermi sphere of corresponding electron density. The Poisson equation now follows from Gauss law $\nabla \cdot \boldsymbol{E} = \rho/\epsilon_0$ with $\boldsymbol{E} = -\nabla\Phi$, so that

$$\begin{aligned}\nabla^2\Phi(\boldsymbol{r}) = [\rho_{\text{ind}} + \rho_{\text{ex}}]/\epsilon_0 &\simeq \Big[e^2\Phi(\boldsymbol{r})\frac{\partial n_e(\epsilon)}{\partial\epsilon}\Big|_{\epsilon=\epsilon_F} - \rho_{\text{ex}}(\boldsymbol{r})\Big]/\epsilon_0 \\ &\equiv k_{\text{TF}}^2\Phi(\boldsymbol{r}) - \rho_{\text{ex}}(\boldsymbol{r})/\epsilon_0. \end{aligned} \tag{4.113}$$

Here, we defined

$$k_{\text{TF}}^2 = 4\pi e^2 \frac{\partial n_e(\epsilon)}{\partial\epsilon}\Big|_{\epsilon=\epsilon_F} = \frac{6\pi e^2 n_e}{\epsilon_F} \tag{4.114}$$

with $n_e = n_e(\epsilon_F)$. For a point charge Q located at the origin as our external charge distribution, $\rho_{\text{ex}} = Q\delta(\boldsymbol{r})$, we obtain

$$\Phi(\boldsymbol{r}) = Q\frac{\mathrm{e}^{-k_{\text{TF}}r}}{4\pi\epsilon_0 r}. \tag{4.115}$$

This corresponds precisely to the Yukawa potential (energy) obtained above.

4.3.2 Friedel oscillations

The static dielectric function above was obtained by considering the long wavelength limit $\boldsymbol{q} \to 0$ of the response function $\chi_0(\boldsymbol{q}, 0)$. However, it is in fact possible to find an exact expression for $\chi_0(\boldsymbol{q}, 0)$, and thus $\varepsilon_r(\boldsymbol{q}, 0)$, for a free electron system (see exercise for derivation):

$$\varepsilon_r(\boldsymbol{q}, 0) = 1 + \frac{4\pi e^2}{q^2}\frac{m}{\pi^2\hbar^2 q}\Big[\frac{qk_F}{2} + \Big(\frac{k_F^2}{2} - \frac{q^2}{8}\Big)\ln\Big|\frac{q/(2k_F)+1}{q/(2k_F)-1}\Big|\Big]. \tag{4.116}$$

The permittivity (also known as dielectric function) does not vary much at small $q \ll k_F$. At $q = \pm 2k_F$ there is, however, a logarithmic singularity, corresponding to a momentum transfer from one side of the Fermi sea to the exact opposite side.

Let us examine the effect that the dielectric function has on the induced charge that arises when we introduce a point charge at the origin. This point charge has a charge density $en_a(r) = en_{a0}\delta(\boldsymbol{r})$ which Fourier transformed gives $n_a(\boldsymbol{q}) = n_{a0}$. Generally, a function which is strongly localized in real space is nearly constant in reciprocal space and vice versa. We can now determine the spatial variation of the induced charge density by first Fourier-transforming Eq. (4.48) to obtain:

$$e\delta n(\boldsymbol{q}) = \frac{q^2\epsilon_0}{e^2}V_{\delta n}(\boldsymbol{q}) = \chi_0(\boldsymbol{q}, 0)V(\boldsymbol{q}) + \chi_0(\boldsymbol{q}, 0)\frac{V_a(\boldsymbol{q})}{\varepsilon(\boldsymbol{q}, 0)} = \frac{1 - \varepsilon(\boldsymbol{q}, 0)}{\varepsilon(\boldsymbol{q}, 0)}n_a(\boldsymbol{q}, 0). \tag{4.117}$$

Fourier-transforming back gives the real space distribution

$$\delta n(\boldsymbol{r}) = \int \frac{d^3q}{(2\pi)^3}\Big[\frac{1}{\varepsilon(\boldsymbol{q})} - 1\Big]n_a(\boldsymbol{q})\mathrm{e}^{i\boldsymbol{q}\cdot\boldsymbol{r}} = -\frac{1}{r}\int_0^\infty g(q)n_a(\boldsymbol{q})\sin(qr)\,dq \tag{4.118}$$

with

$$g(q) = \frac{q}{2\pi^2}\frac{\varepsilon_r(q) - 1}{\varepsilon_r(q)}. \tag{4.119}$$

Note that $g(q)$ vanishes for $q \to \infty$. After lengthy calculations, one arrives at

$$\delta n(r) \propto n_{a0} \frac{\cos(2k_F r)}{r^3} \tag{4.120}$$

in the limit $k_F r \gg 1$. We will give a simplified calculation below which shows the same qualitative result. The appearance of oscillations can be understood from the fact that electrons behave as waves quantum mechanically, so the interference between many such electron waves in the total distribution of electrons at different distances from the inserted static charge interfere destructively at some points (lowering the density) and constructively at other points (increasing the density) since the electrons all have different wavevectors.

Similar physics takes place not only in the presence of an inserted external charge, but whenever there is an abrupt change in the charge density, such as at the edge of a material (vacuum interface). Let us give a simplified mathematical model to describe the essence of the physics. Consider first the Friedel oscillations of the electron (charge) density induced by a hard wall. Consider an incoming electron with energy E and momentum $k = \sqrt{2mE}$ which scatters on the hard wall at $x = 0$:

$$\psi = \mathrm{e}^{\mathrm{i}kx} + r\mathrm{e}^{-\mathrm{i}kx}. \tag{4.121}$$

The reflection probability $R = |r|^2$ is unity and r is in general complex since the energy is smaller than the potential barrier. The density of the electron wavefunction is

$$\rho = |\psi|^2 = 1 + |r|^2 + 2\mathrm{Re}\{r\mathrm{e}^{-2\mathrm{i}kx}\}. \tag{4.122}$$

The reflection coefficient is generally:

$$r = \frac{k-q}{k+q}, \tag{4.123}$$

where $q = \sqrt{2m(E-V)}$. For $V \gg E$, as appropriate for a hard wall, the reflection coefficient is $r \simeq -1$. Therefore, we get

$$\rho = 2 + 2\cos(2kx). \tag{4.124}$$

To find the total electron density at a point x, we have to take into account the contribution from all possible wavevectors. At zero temperature, all wavevectors from 0 to k_F will contribute:

$$\rho_{\mathrm{tot}}(x) = \int_0^{k_F} dk\rho(x) = 2k_F + \frac{\sin(2k_F x)}{x}. \tag{4.125}$$

It is clear that an oscillating component appears which decays away from the interface. The physics behind this is that different wavelength electrons will be in phase close to the interface, but eventually start to interfere destructively since they oscillate over different length scales.

4.3.3 Dielectric function and susceptibility in various dimensions

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

$$\chi_0(q,\omega=0) = \begin{cases} -\frac{m}{\pi\hbar^2}\frac{2}{q}\ln\left|\frac{1+2/s}{1-2/s}\right| \text{ in 1D} \\ -\frac{m}{\pi\hbar^2}[1-\theta(s-2)\sqrt{1-(2/s)^2}] \text{ in 2D} \\ -\frac{m}{\pi^2\hbar^2}\frac{1}{q}\left[\left(\frac{k_F^2}{2}-\frac{q^2}{8}\right)\ln\left|\frac{s+2}{s-2}\right| + \frac{qk_F}{2}\right] \text{ in 3D} \end{cases}$$

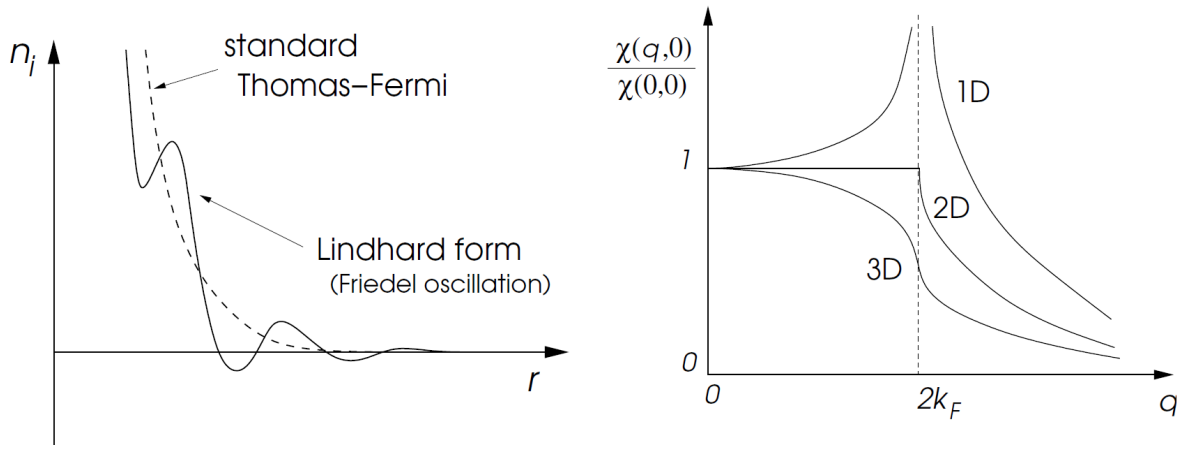

Figure 4.4: **Left:** Friedel oscillations of the charge distribution. **Right:** Lindhard functions for different dimensions. Figure taken from [3].

where $s = q/k_F$. We see that there is a singularity [3] at $q = 2k_F$ in all dimensions. In one dimension, there is a logarithmic divergence, in two dimensions there is a kink, and in three dimensions only the derivative diverges. Later we will see that this behavior may lead to instabilities of the metallic state, in particular for the one-dimensional case due to the strong singularity at $q = 2k_F$.

4.4 Phonons

The atoms in a lattice of a solid are not immobile but vibrate around their equilibrium positions. We will describe this new degree of freedom by treating the lattice as a continuous elastic medium (Jellium with elastic modulus λ). This approximation is suffcient to obtain some essential features of the interaction between lattice vibrations and electrons. In particular, renormalized screening effects will be found.

Phonons can be excited in different ways. One type of excitation is the acoustic phonon, which is characterized by a dispersion $E = ck$ for small k where c is the speed of sound. This type of phonon correspond to a coherent movement of atoms in the lattice displaced out of their equilibrium position. The displacement can be both in the direction of propagation of the phonon (longitudinal acoustic or LA phonon) or transverse to the propagation (transverse acoustic or TA phonon), the latter being similar to waves travelling on a string. When the wavelength (wavevector) of an acoustic phonon goes to infinity (zero), it corresponds to a simple translation of the entire crystal, which thus has zero energy as nothing is deformed. The name acoustic derives from the similar behavior of these excitations to sound waves in air.

Another type is an optical phonon, which has a gap in its spectrum: its energy is not zero for $k \to 0$. The optical phonon mode is excited by neighboring atoms moving in opposite directions. This can happen if the lattice basis consists of two or more atoms. If the crystal is ionic, so that neighboring atoms have opposite charge, this creates an oscillating dipole which interacts with electromagnetic

[3] Recall that a mathematical singularity is a point where the function is either not defined, diverges, or is not differentiable.

radiation (hence the name *optical*). For instance, optical phonons can absorb light and they can be excited by light because their energies are comparable to light in the *e.g.* infrared regime. In contrast, acoustic phonons typically have very small energies at small k compared to light. Optical phonons can also be longitudinal (LO) or transverse (LO).

Our approach will here be limited to mono-atomic unit cells for simplicity.

4.4.1 Vibration of a isotropic continuous medium

The deformation of an elastic medium can be described by the displacement of the infinitesimal volume element d^3r around a point $\boldsymbol{r}$ to a different point $\boldsymbol{r}'(\boldsymbol{r})$. We may then introduce the so-called displacement field $\boldsymbol{u}(\boldsymbol{r}) = \boldsymbol{r}'(\boldsymbol{r}) - \boldsymbol{r}$ as a function of $\boldsymbol{r}$. In general, $\boldsymbol{u}$ can also depend on time. To lowest order, the elastic energy cost associated with small deformations of an isotropic medium is

$$E_{\text{el}} = \frac{\lambda}{2} \int d^3r [\nabla \cdot \boldsymbol{u}(\boldsymbol{r}, t)]^2 \tag{4.126}$$

where λ is the elastic modulus. We can understand this expression intuitively: there is no deformation energy if the medium is just shifted uniformly, in which case $\boldsymbol{u} \neq 0$ but $\nabla \cdot \boldsymbol{u} = 0$. Moreover, the energy depends only on the magnitude of the deformation (guaranteed by the square operation) in an isotropic medium and not its direction.

The energy term above produces a restoring force trying to bring the system back to an undeformed state. In the model above, we have neglected shear contributions [4] The continuum form above is valid for deformation wavelengths much longer than the lattice constant, so that details of the arrangement of atoms in the lattice can be neglected. The kinetic energy of the motion of the medium is given by

$$E_{\text{kin}} = \frac{\rho_0}{2} \int d^3r \Big(\frac{\partial \boldsymbol{u}(\boldsymbol{r}, t)}{\partial t} \Big)^2 \tag{4.128}$$

where $\rho_0 = M_i n_i$ is the mass density with the ionic mass M_i and the ionic density n_i. To find the physically realized displacement field $\boldsymbol{u}(\boldsymbol{r}, t)$, we need to solve its equation of motion that follows from variation of the Lagrangian functional $L[\boldsymbol{u}] = E_{\text{kin}} - E_{\text{el}}$ with respect to $\boldsymbol{u}(\boldsymbol{r}, t)$. This gives the Euler-Lagrange equation

$$\frac{1}{c_s^2} \frac{\partial^2}{\partial t^2} \boldsymbol{u}(\boldsymbol{r}, t) - \nabla [\nabla \cdot \boldsymbol{u}(\boldsymbol{r}, t)] = 0. \tag{4.129}$$

This is a wave equation with sound velocity $c_s^2 = \lambda / \rho_0$. An arbitrary vector field $\boldsymbol{u}$ can always be resolved into a longitudinal and transverse component $\boldsymbol{u} = \boldsymbol{u}_L + \boldsymbol{u}_T$ where $\boldsymbol{u}_L$ has no curl while $\boldsymbol{u}_T$

[4] The most general form of the elastic energy of an isotropic medium takes the form

$$E_{\text{el}} = \int d^3r \sum_{\alpha, \beta \in x, y, z} [\frac{\lambda}{2} (\partial_\alpha u_\alpha)(\partial_\beta u_\beta) + \mu (\partial_\alpha u_\beta)(\partial_\alpha u_\beta)]. \tag{4.127}$$

The so-called Lamé coefficients λ and μ characterize the elastic properties. The elastic constant λ describes density modulations of the underlying lattice leading to longitudinal elastic waves, whereas μ corresponds to so-called shear deformations which lead to transversely polarized elastic waves. The latter are in many cases less important for the coupling of electrons and lattice vibrations than the former.

has no divergence. Using that $\nabla \times (\nabla \times \boldsymbol{a}) = \nabla(\nabla \cdot \boldsymbol{a}) - \nabla^2 \boldsymbol{a}$, we obtain by insertion

$$\frac{1}{c_s^2}\frac{\partial^2}{\partial t^2}\boldsymbol{u}_L(\boldsymbol{r},t) - \nabla^2 \boldsymbol{u}_L(\boldsymbol{r},t) = 0, \ \frac{\partial^2}{\partial t^2}\boldsymbol{u}_T(\boldsymbol{r},t) = 0. \tag{4.130}$$

This is the standard wave equation and we can set $\boldsymbol{u}_T = 0$ to exclude any linearly increasing solution in time. That leaves us with only the longitudinal part $\boldsymbol{u}_L$ (drop the L subscript from now). The resulting displacement field can be expanded into a complete set of so-called normal modes,

$$\boldsymbol{u}(\boldsymbol{r},t) = \frac{1}{\sqrt{\Omega}}\sum_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}}\Big(q_{\boldsymbol{k}}(t)\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} + q_{\boldsymbol{k}}(t)^*\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\Big). \tag{4.131}$$

where $\boldsymbol{e}_{\boldsymbol{k}} \parallel \boldsymbol{k}$ to ensure that $\boldsymbol{u}$ is curl-free. Here, every $q_{\boldsymbol{k}}(t)$ satisfies the equation

$$\frac{d^2}{dt^2}q_{\boldsymbol{k}} + \omega_{\boldsymbol{k}}^2 q_{\boldsymbol{k}} = 0, \tag{4.132}$$

with the frequency $\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}| = c_s k$ and the polarization vector $\boldsymbol{e}_{\boldsymbol{k}}$ has unit length. The total energy expressed in terms fo the normal modes reads

$$E = \sum_{\boldsymbol{k}} \rho_0 \omega_{\boldsymbol{k}}^2 [q_{\boldsymbol{k}}(t)q_{\boldsymbol{k}}^*(t) + q_{\boldsymbol{k}}^*(t)q_{\boldsymbol{k}}(t)]. \tag{4.133}$$

Next, we switch from a Lagrangian to Hamiltonian description by defining new canonical variables (as can be verified from Hamilton's equations):

$$\begin{aligned} Q_{\boldsymbol{k}} &= \sqrt{\rho_0}(q_{\boldsymbol{k}} + q_{\boldsymbol{k}}^*), \\ P_{\boldsymbol{k}} &= \frac{d}{dt}Q_{\boldsymbol{k}} = -i\omega_{\boldsymbol{k}}\sqrt{\rho_0}(q_{\boldsymbol{k}} - q_{\boldsymbol{k}}^*). \end{aligned} \tag{4.134}$$

The energy is then given by

$$E = \frac{1}{2}\sum_{\boldsymbol{k}}(P_{\boldsymbol{k}}^2 + \omega_{\boldsymbol{k}}^2 Q_{\boldsymbol{k}}^2). \tag{4.135}$$

This shows that the system is equivalent to an ensemble of independent harmonic oscillators, one for each mode $\boldsymbol{k}$. Since the classical Hamiltonian is now in place, the procedure to go quantum mechanical is clear: we promote the canonical variables to operators, $P_{\boldsymbol{k}} \to \hat{P}_{\boldsymbol{k}}$ and $Q_{\boldsymbol{k}} \to \hat{Q}_{\boldsymbol{k}}$, which by definition then obey the commutation relation

$$[\hat{Q}_{\boldsymbol{k}}, \hat{P}_{\boldsymbol{k}'}] = i\hbar\delta_{\boldsymbol{k},\boldsymbol{k}'}. \tag{4.136}$$

We can then introduce lowering and raising operators

$$\begin{aligned} \hat{b}_{\boldsymbol{k}} &= \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}}(\omega_{\boldsymbol{k}}\hat{Q}_{\boldsymbol{k}} + i\hat{P}_{\boldsymbol{k}}), \\ \hat{b}_{\boldsymbol{k}}^\dagger &= \frac{1}{\sqrt{2\omega_{\boldsymbol{k}}}}(\omega_{\boldsymbol{k}}\hat{Q}_{\boldsymbol{k}} - i\hat{P}_{\boldsymbol{k}}). \end{aligned} \tag{4.137}$$

These satisfy the commutation relation

$$[\hat{b}_{\boldsymbol{k}}, \hat{b}_{\boldsymbol{k}'}^\dagger] = \delta_{\boldsymbol{k},\boldsymbol{k}'}, \tag{4.138}$$

and every other commutator equal to zero. The physical interpretation of these relations is that these operators create and annihilate quasiparticles following Bose-Einstein statistics. The QM Hamiltonian corresponding to the classical energy in Eq. (4.135) is then

$$H = \sum_{\boldsymbol{k}} \hbar\omega_{\boldsymbol{k}} (\hat{b}^\dagger_{\boldsymbol{k}} \hat{b}_{\boldsymbol{k}} + \frac{1}{2}). \tag{4.139}$$

These bosonic quasiparticles are called *phonons* and have a well-defined energy-momentum relation $\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}|$.
The displacement field $u(\boldsymbol{r})$ can now be promoted to a quantum mechanical operator $\hat{u}(\boldsymbol{r})$:

$$\hat{u}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\boldsymbol{k}} \boldsymbol{e}_{\boldsymbol{k}} \sqrt{\frac{\hbar}{2\rho_0 \omega_{\boldsymbol{k}}}} [\hat{b}_{\boldsymbol{k}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}} + \hat{b}^\dagger_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}}]. \tag{4.140}$$

As mentioned above, the continuum approximation is valid for long wavelengths only, corresponding to small $\boldsymbol{k}$. For wavevectors at the edge of the BZ, $\boldsymbol{k} \simeq \pi/a$, the discreteness of the lattice begins to make itself known in the form of corrections to the linear dispersion $\omega_{\boldsymbol{k}}$ above. Since the number of degrees of freedom for a system of N_i atoms is $3N_i$, there exists a maximal wave vector called the Debye wavevector k_D. This allows us to define the corresponding Debye frequency $\omega_D = c_s k_D$ and the Debye temperature $\Theta_D = \hbar\omega_D/k_B$. Finally, inclusion of optical phonons requires that the arrangement of atoms within a unit cell has to be considered, whereas in our simple picture of a continuous medium only acoustic phonons exist.

4.4.2 Phonons in metals

Our considerations so far apply well to semiconducting materials, where ionic interactions are mediated via short-ranged covalent bonds (neighboring ions sharing electrons) so that oscillations around the equilibrium position may be approximated by a harmonic potential, as above. The situation is however more subtle for metals, where ions interact through the Coulomb interaction and are held together through an interplay with the mobile conduction electrons.
Looking at the positively charged background first, let us approximate it as an ionic gas. Similarly to the electronic gas, this background then exhibits well-defined collective plasma excitations at the ionic plasma frequency

$$\Omega_p^2 = \frac{4\pi n_i (Z_i e)^2}{M_i} \tag{4.141}$$

which we obtained by taking Eq. (4.95) with $n_0 \to n_i = n_0/Z_i$, the density of ions with charge number $Z_i, e \to Z_i e$, and $m \to M_i$ the atomic mass. Apparently, the above excitation energy does not vanish as $\boldsymbol{k} \to 0$. Therefore, it looks like the background of a metallic system can not be described as an elastic medium where the excitation spectrum is expected to be linear in k.
The missing aspect is indeed the feedback effect of the electrons, which react nearly instantaneously to the slow ionic motion, due to their much smaller mass. The finite plasma frequency is a consequence of the long-range nature of the Coulomb potential (as mentioned when we discussed plasma oscillations of the electron gas). However, we have seen that the above electrons tend to screen these potentials, especially for small wavevectors $\boldsymbol{k}$. Therefore, the "bare" ionic plasma frequency Ω_p above is thus renormalized to

$$\omega_{\boldsymbol{k}}^2 = \frac{\Omega_p^2}{\varepsilon_r(\boldsymbol{k}, 0)} = \frac{k^2 \Omega_p^2}{k^2 + k_{\mathrm{TF}}^2} \simeq (c_s k)^2. \tag{4.142}$$

when the presence of the electrons is taken into account, due to the modified Coulomb potential by a factor $1/\varepsilon_r(\boldsymbol{k},\omega)$. Note that we have set $\omega = 0$ in $\varepsilon(\boldsymbol{k},\omega)$ since phonon energies are typically much smaller than the electronic plasma frequency, so that we may approximate $\varepsilon(\boldsymbol{k},\omega_k) \simeq \varepsilon(\boldsymbol{k},0)$. Having included the back-reaction of the electrons, we then recover a linear dispersion of a sound wave ($\omega_{\boldsymbol{k}} = c_s|\boldsymbol{k}|$) is recovered with a new velocity of sound which reads

$$c_s^2 \simeq \frac{\Omega_p^2}{k_{\rm TF}^2} = \frac{Zm\omega_p^2}{M_i k_{\rm TF}^2} = \frac{1}{3} Z v_F^2 \frac{m}{M_i}. \tag{4.143}$$

We see that the phonon velocity is much smaller than the Fermi velocity. We can also compare energy scales for the ions and electrons, and find

$$\frac{\Theta_D}{T_F} \simeq \frac{c_s}{v_F} \ll 1 \tag{4.144}$$

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

4.4.3 Kohn anomaly

Employing the general expression for $\varepsilon_r(\boldsymbol{k},0)$ given by the Lindhard form, rather than the Thomas-Fermi approximation, we find from the expression

$$\omega_k^2 = \frac{\Omega_p^2}{\varepsilon_r(\boldsymbol{k},0)} \tag{4.145}$$

that ω_k is singular when $|\boldsymbol{k}| = 2k_F$. Note that this relation is valid even for large wavevectors $\boldsymbol{k}$ since phonon frequencies are small compared to the electronic plasma frequency. In particular, it follows from the general expression we have provided of $\varepsilon_r(\boldsymbol{k},0)$ in terms of the susceptibility $\chi_0(\boldsymbol{k},0)$ that $\frac{\partial \omega_{\boldsymbol{k}}}{\partial \boldsymbol{k}}$ either diverges or is not well-defined in the limit $k \to 2k_F$ due to the singular behavior of $\chi_0(\boldsymbol{k} \to 2\boldsymbol{k}_F, 0)$. This behavior is called the *Kohn anomaly* and results from the interaction between electrons and phonons. In effect, it is not contained in the initial elastic medium model we did to begin with which neglects ion-electron interactions. Below, we shall show the anomalous behavior explicitly for a one dimensional system.

4.4.4 Peierls instability in one dimension

The Kohn anomaly has particularly dramatic effects in a one-dimensional electron system. Here, we show that the electron-phonon coupling leads to an instability of the metallic state. Consider a 1D jellium model where the ionic background is treated as an elastic medium with a displacement field u along the extended direction (x-axis). To treat this analytically, we neglect both electron-electron interactions and the slow time evolution of the background modulation, giving us a Hamiltonian

$$H = H_{\rm isol} + H_{\rm int} \tag{4.146}$$

where the isolated electronic and ionic systems are described by

$$H_{\rm isol} = \sum_{k,s} \frac{\hbar^2 k^2}{2m} c_{ks}^\dagger c_{ks} + \frac{\lambda}{2} \int dx \Big(\frac{du(x)}{dx}\Big)^2 \tag{4.147}$$

whereas the interactions between the systems comes in via the Coulomb interaction

$$H_{\rm int} = -n_0 \sum_s \int\int dx dx' V(x-x') \frac{du(x)}{dx} \hat{\psi}_s^\dagger(x') \hat{\psi}_s(x'). \tag{4.148}$$

In the general theory of elastic media $\nabla \cdot \boldsymbol{u} = -\delta n/n_0$ describes density modulations due to the deformation [5]. The interaction term above models the coupling of the electrons to charge density fluctuations of the positively charged background through the screened Coulomb interaction $V(x-x')$. Thus, we are here for simplicity only taking into account the ion-electron Coulomb interaction caused by a deformation of the ion lattice due to phonons. We are not including the Coulomb interaction between ions and electrons in the absence of phonons, which we know is responsible for setting up the periodic potential that electrons move around in Instead, here we want to study the coupling between electrons and phonons.

We note in passing that transverse phonon modes, defined by $\nabla \cdot \boldsymbol{u} = 0$, thus do not couple to electrons to lowest order. Consider the ground state of N electrons in a system of length L, leading to an electronic density $n_0 = N/L$. For a uniform background $u(x) =$ constant, the Fermi wavevector of the free electrons is determined by

$$N = \sum_s \frac{L}{2\pi} \int_{-k_F}^{k_F} dk\ 1 = 2\frac{L}{2\pi} 2k_F \tag{4.153}$$

so that $k_F = \pi n_0/2$.

Consider now the Kohn anomaly of this system. For a small background modulation $u(x) \neq$ constant, the interaction term $H_{\rm int}$ is treated perturbatively and leads to a renormalization of the elastic modulus λ, which we now show. It will be useful to express the full Hamiltonian in momentum space

$$\begin{aligned} H_{\rm isol} &= \sum_{k,s} \frac{\hbar^2 k^2}{2m} c_{ks}^\dagger c_{ks} + \frac{\Omega\rho_0}{2} \sum_q \omega_q^2 u_q u_{-q}, \\ H_{\rm int} &= i \sum_{k,q,s} q[\tilde{V}_{-q} u_q c_{k+q,s}^\dagger c_{ks} - \tilde{V}_q u_{-} q c_{ks}^\dagger c_{k+q,s}], \end{aligned} \tag{4.154}$$

[5] Consider a small three-dimensional volume V of a body. When the body is deformed, an infinitesimal element of area $d\boldsymbol{A}(\boldsymbol{r})$ on the surface of a small volume V is displaced a vectorial distance $\boldsymbol{u}(\boldsymbol{r})$. This means that it sweeps out a volume $\boldsymbol{u} \cdot d\boldsymbol{A}$. Therefore, the total change in volume produced by a small displacement field $\boldsymbol{u}$ is:

$$\delta V = \int d\boldsymbol{A} \cdot \boldsymbol{u} = \int dV \nabla \cdot \boldsymbol{u} \simeq \nabla \cdot \boldsymbol{u} \int dV = (\nabla \cdot \boldsymbol{u}) V. \tag{4.149}$$

Above, we used Gauss' theorem and in the third equality approximated that $\nabla \cdot \boldsymbol{u}$ is essentially constant throughout the volume due to the smallness of the volume. Thus, we have proven that the change in volume is given by

$$\frac{\delta V}{V} = \nabla \cdot \boldsymbol{u}. \tag{4.150}$$

As the volume is deformed, the mass m originally present in the volume nevertheless must remain the same. The original mass density is $n_0 = m/V$. Now, the change in mass density δn will be given by

$$\begin{aligned} \delta n &= \frac{m}{V + \delta V} - \frac{m}{V} \\ &\simeq \frac{m}{V}\Big(1 - \frac{\delta V}{V}\Big) - \frac{m}{V} \\ &= \frac{m}{V}\Big(-\frac{\delta V}{V}\Big). \end{aligned} \tag{4.151}$$

This means that

$$\frac{\delta n}{n_0} = -\nabla \cdot \boldsymbol{u}. \tag{4.152}$$

where we used that $\lambda q^2 = \rho_0 \omega_q^2$ which we found earlier. Moreover, we defined

$$u(x) = \frac{1}{\sqrt{L}} \sum_q u_q \mathrm{e}^{iqx},$$
$$V(x) = \frac{1}{\sqrt{L}} \sum_q \tilde{V}_q \mathrm{e}^{iqx}. \tag{4.155}$$

Here, $\tilde{V}_q = e^2/[q^2 \varepsilon(q,0)]$. Now, we compute the second order correction to the ground state energy using Rayleigh-Schrödinger perturbation theory (the linear energy shift vanishes):

$$\begin{aligned}\Delta E^{(2)} &= \sum_{k,q,s} q^2 |\tilde{V}_q|^2 u_q u_{-q} \sum_n \frac{|\langle \Psi_0 | c^\dagger_{k,s} c_{k+q,s} | n\rangle|^2 + |\langle \Psi_0 | c^\dagger_{k+q,s} c_{k,s} | n\rangle|^2}{E_0 - E_n} \\ &= \sum_q |\tilde{V}_q|^2 q^2 u_q u_{-q} \sum_k \frac{n_{k+q} - n_k}{\epsilon_{k+q} - \epsilon_k} \\ &= \Omega \sum_q |\tilde{V}_q|^2 q^2 \chi_0(q,0) u_q u_{-q}. \end{aligned} \tag{4.156}$$

Here, the states $|n\rangle$ are electron-hole excitations of the filled Fermi sea. This term gives a correction to the energy eigenvalue of H_{isol}. The phonon frequency is therefore renormalized from its value without coupling to electrons, $\omega_q^2 = (\lambda/\rho_0) q^2 = c_s^2 q^2$, to

$$(\omega_q^{\mathrm{ren}})^2 \simeq \omega_q^2 + \Omega |\tilde{V}_q|^2 u_q u_{-q} q^2 \chi_0(q,0) \tag{4.157}$$

where $\chi_0(q,0) = \ln|(q+2k_F)/(q-2k_F)|$. From the behavior for $q \to 0$, we infer that the velocity of sound is renormalized. The modification becomes drastic at $q = 2k_F$. The phonon spectrum gradually 'softens', meaning the frequency is reduced, and even becomes negative. The latter behavior is unphysical and is an artifact of the perturbation approximation we have made. This behavior nevertheless suggests an instability in the system which cannot be treated via perturbation theory around the uniform state: the system apparently takes on a completely different characteristic.

4.4.5 Peierls instability at $q = 2k_F$

The new behavior turns out to be a Bose-Einstein condensation of phonons with wavevector $q = 2k_F$. This means that a macroscopic number of phonons condense into the same quantum state $q = 2k_F$, and such a coherent superposition corresponds to a static periodic deformation of the ionic background with wavevector $2k_F$. To be specific, our hypothesis is thus that the phonons form a coherent state (analogous to the coherent state of a harmonic oscillator):

$$|\Phi_Q^{\mathrm{coh}}\rangle = \mathrm{e}^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{(b_Q^\dagger)^n}{n} \alpha^n |0\rangle. \tag{4.158}$$

As with every coherent state, the phonon number for a given quantum number Q is not sharply defined (as the state is instead phase-coherent). The occupation of a given coherent state Q is then

$$\langle \Phi_Q^{\mathrm{coh}} | b_Q^\dagger b_Q | \Phi_Q^{\mathrm{coh}}\rangle = |\alpha|^2 \tag{4.159}$$

and moreover, we find that the expectation value of the displacement operator is:

$$\langle \Phi_Q^{\mathrm{coh}} | \hat{u}(x) | \Phi_Q^{\mathrm{coh}}\rangle = \frac{\hbar}{2L\rho_0\omega_Q} [\alpha \mathrm{e}^{iQx} + \alpha^* \mathrm{e}^{-iQx}] = u_0 \cos Qx \tag{4.160}$$

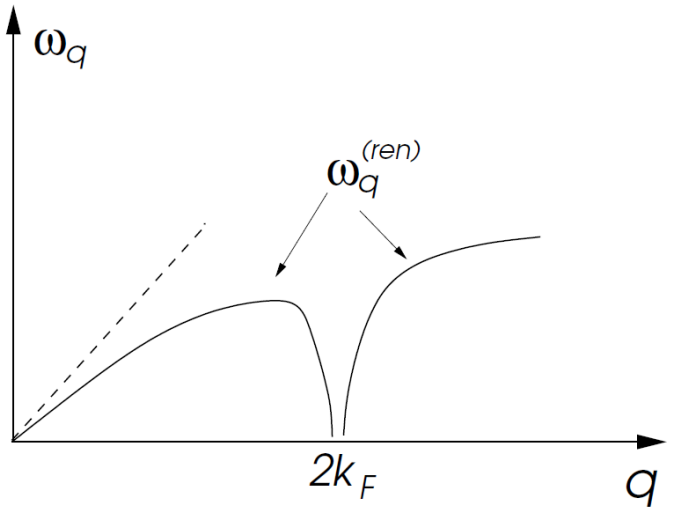

Figure 4.5: Kohn anomaly for the one-dimensional system with electron-phonon coupling. The renormalization of the phonon frequencty is divergent at $q = 2k_F$. Figure taken from Ref. [3].

where $u_0 = \hbar\alpha/\rho_0 L\omega_Q$ assuming that α is real. Motivated by this, we therefore investigate our hypothesis by assuming that the ionic background at $Q = 2k_F$ shows a periodic density modulation corresponding to the coherent phonon state and use a variational approach rather than a perturbative approach. In effect, we fix

$$u(x) = u_0 \cos Qx \tag{4.161}$$

with $Q = 2k_F$ and aim to identify the energetically most favorable value of u_0 variationally. We now that show that this modulation actually lowers the energy of the electrons. Assuming that u_0 is small, we can evaluate the electronic energy using the approximation of nearly free electrons where Q appears as a reciprocal lattice vector. According to our previous treatment of the band structure for nearly free electrons, the electronic spectrum for $0 \leq k \leq Q$ is then approximately determined by the secular equation:

$$\begin{vmatrix} \frac{\hbar^2 k^2}{2m} - E & \Delta \\ \Delta^* & \frac{\hbar^2 (k-Q)^2}{2m} - E \end{vmatrix} = 0 \tag{4.162}$$

where Δ follows from the Fourier transform of the potential $V(x)$:

$$\Delta = -iQu_0 n\tilde{V}_Q \tag{4.163}$$

with

$$\tilde{V}_Q = \int dx \mathrm{e}^{iQx} V(x). \tag{4.164}$$

From the above equation, we obtain the energy eigenstates

$$E_k^{\pm} \frac{\hbar^2}{4m}\left[(k-Q)^2 + k^2 \pm \sqrt{[(k-Q)^2 - k^2]^2 + 16m^2|\Delta|^2/\hbar^4}\right]. \tag{4.165}$$

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

$$E_{\rm tot}(u_0) = 2\sum_{0\le k\le Q} E_{k-} + \frac{\lambda L Q^2}{4}u_0^2 \tag{4.166}$$

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

$$\begin{aligned} 0 &= \frac{1}{L}\frac{dE_{\rm tot}}{du_0} \\ &= -u_0\frac{8Qmn^2\tilde{V}_Q^2}{\hbar^2\pi}\cdot\operatorname{arcsinh}\Big(\frac{\hbar^2 k_F}{2mn\tilde{V}_Q u_0}\Big) + \frac{\lambda}{2}Q^2u_0. \end{aligned} \tag{4.167}$$

Since u_0 is assumed small, we solve this equation by using the approximation $\operatorname{arsinh}(x) \simeq \ln(2x)$ for $x \gg 1$ and find

$$u_0 = \frac{2}{k_F}\frac{\epsilon_F}{n\tilde{V}_Q}{\rm e}^{-1/N(0)g} \tag{4.168}$$

where $\epsilon_F = \hbar^2k_F^2/2m$ is the Fermi energy and $N(0) = 2m/(\pi\hbar^2k_F)$ is the density of states at the Fermi energy. We introduced the coupling constant $g = 4n^2\tilde{V}_Q^2/\lambda$: the coupling is stronger the more polarizable (softer) the ionic background is, in effect when the elastic modulus λ is small. Note that the static displacement u_0 depends exponentially on the coupling and on the density of states.
The opening of an energy gap ΔE is a key reason for this so-called Peierls instability to occur:

$$\Delta E = E^+_{k_F} - E^-_{k_F} = 2|\Delta| = 8\epsilon_F\exp[-1/N(0)g], \tag{4.169}$$

at $k = \pm k_F$. The gap results in a lowering of the energy of the electron states in the lower band in the vicinity of the Fermi energy. For this reason this kind of instability is called a Fermi surface instability. Due to the gap, the metal has now turned into a semiconductor with a finite energy gap for all electron-hole excitations.

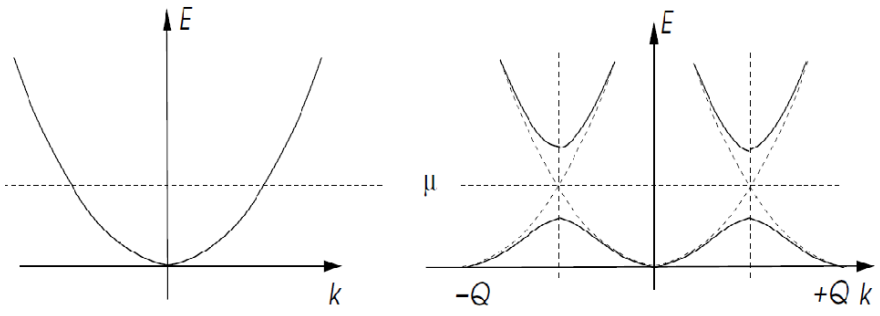

Figure 4.6: Change of the electron spectrum when the ionic background has a Peierls instability. The modulation of the ionic background yields gaps at the Fermi points and the system becomes an insulator. Figure taken from Ref. [3].

The modulation of the electron density follows the charge modulation due to the ionic lattice deformation, which can be seen by expressing the wavefunction of the electronic states

$$\psi_k(x) = \frac{1}{\sqrt{\Omega}}\frac{\Delta{\rm e}^{ikx} + (E_k - \epsilon_k){\rm e}^{i(k-Q)x}}{\sqrt{(E_k-\epsilon_k)^2 + |\Delta|^2}} \tag{4.170}$$

(which can be found using the treatment of nearly free electrons we looked at earlier). This is a superposition of two plane waves with wavevectors k and $k-Q$, respectively. Hence, the charge density from a given mode k reads

$$\rho_k(x) = -e|\psi_k(x)|^2 \tag{4.171}$$

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

$$\begin{aligned}\delta\rho(x) &= \sum_k \rho_k(x) - (-en) = \frac{e}{2}\int_0^{k_F} \frac{dk'}{2\pi} \frac{m|\Delta|\sin(Qx)}{\sqrt{\hbar^2 k_F^2 k'^2 + m^2|\Delta|^2}} \\ &= \frac{en|\Delta|}{16\epsilon_F}\ln\Big(\frac{2\epsilon_F}{|\Delta|}\Big)\sin(2k_F x).\end{aligned} \tag{4.172}$$

Such a state with a spatially modulated electronic charge density is called a *charge density wave* (CDW). This instability is important in quasi-1D metals (weakly coupled 1D chains) which are realized in for instance organic conductors. In higher dimensions, the effect of the Kohn anomaly is generally less pronounced, so that spontaneous deformations are less likely to occur. However, CDW instabilities can also be observed in two dimensions or more in systems with so-called nested Fermi surfaces. The nesting means that large parts of the Fermi surface are connected to each other via a well-defined wavevector. These systems resemble in some respects 1D systems. Finally, we note that the electron-phonon interaction is also crucial for another kind of Fermi surface instability: when metals exhibit superconductivity (more on this later).

4.4.6 Dynamics of phonons and the dielectric function

We know by now that an external potential V_a is screened by the polarization of the electrons. As the positive charged ionic background is also polarizable, its effect should also be included int he renormalization of the external potential. In general, the fully renormalized potential V_{ren} should therefore be expressed via

$$V_{\text{ren}} = V_a/\varepsilon_r \tag{4.173}$$

with a relative permittivity that includes both the contribution from the ions and electrons. In order to determine V_{ren} and ε, we first define the 'bare' (not renormalized) electronic and ionic dielectric functions $\varepsilon_r^{\text{el}}$ and $\varepsilon_r^{\text{ion}}$. The renormalized potential in Eq. (4.173) acn be expressed by considering different viewpoints. Firstly, if the ionic potential V_{ion} is added to the external potential V_a, the remaining screening is due to the electrons only:

$$\varepsilon_r^{\text{el}} V_{\text{ren}} = V_a + V_{\text{ion}}. \tag{4.174}$$

Secondly, the electronic potential V_{el} may be added to the external potential V_a so that the ions exclusively renormalize the new potential:

$$\varepsilon_r^{\text{ion}} V_{\text{ren}} = V_a + V_{\text{el}}. \tag{4.175}$$

Note that in the above equation, all effects of the electron polarization are included in V_{el} so that the dielectric function results from the 'bare' ions. Finally, we use the fact that the renormalized potential must be the sum of all the potentials:

$$V_{\text{ren}} = V_a + V_{\text{el}} + V_{\text{ion}}. \tag{4.176}$$

Combining the above equations provide us with

$$(\varepsilon_r^{\text{el}} + \varepsilon_r^{\text{ion}} - \varepsilon)V_{\text{ren}} = V_a + V_{\text{el}} + V_{\text{ion}} \tag{4.177}$$

so that we may identify

$$\varepsilon_r = \varepsilon_r^{\text{el}} + \varepsilon_r^{\text{ion}} - 1. \tag{4.178}$$

Taking now into account our previous results discussions of the plasma excitation of the bare ions and electrons, we found for small wavevectors $\boldsymbol{k} \to 0$:

$$\begin{aligned} \varepsilon_r^{\text{ion}} &= 1 - \frac{\Omega_p^2}{\omega^2}, \\ \varepsilon_r^{\text{el}} &= 1 + \frac{k_{\text{TF}}^2}{k^2}. \end{aligned} \tag{4.179}$$

In effect, $\varepsilon_r^{\text{el}}$ expresses the effect of plasma oscillations of the ionic lattice on the potential, whereas $\varepsilon_r^{\text{ion}}$ expresses the effect of plasma oscillations of the free electron gas on the potential.

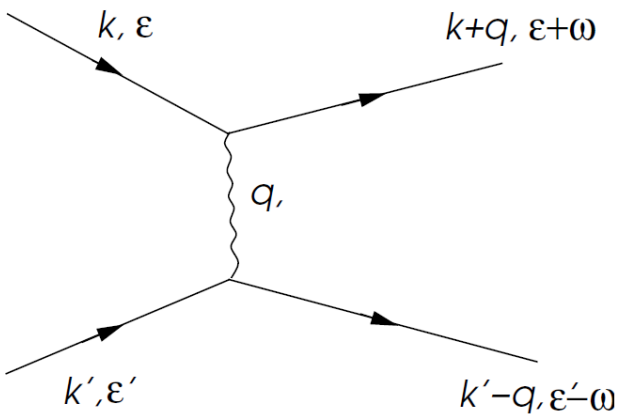

Figure 4.7: Diagram for the electron-electron interaction involving also electron-phonon coupling. Figure taken from Ref. [3].

For the electrons we used the result from the quasi-static limit Eq. (4.108) since the electrons respond very quickly to a perturbation due to their low mass. The full dielectric function can now be identified:

$$\varepsilon_r = 1 + \frac{k_{\text{TF}}^2}{k^2} - \frac{\Omega_p^2}{\omega^2} = \left(1 + \frac{k_{\text{TF}}^2}{k^2}\right)\left(1 - \frac{\omega_{\boldsymbol{k}}^2}{\omega^2}\right). \tag{4.180}$$

Here, we used our previously derived expression for the renormalized phonon dispersion relation $\omega_{\boldsymbol{k}}$. With this full dielectric function, the time-independent Coulomb interaction

$$V_a = \frac{e^2}{4\pi q^2 \varepsilon_0} \tag{4.181}$$

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

$$V_{\text{ren}} = \frac{e^2}{4\pi q^2 \varepsilon(\boldsymbol{q},\omega)} = \frac{e^2}{4\pi\varepsilon_0}\frac{1}{k_{\text{TF}}^2+q^2}\frac{\omega^2}{\omega^2-\omega_q^2}. \tag{4.182}$$

This reveals a very interesting and surprising feature for the effective electron-electron interaction where the effect of phonons have been taken into account. The above interaction corresponds to the matrix element for a scattering process of two electrons with momentum exchange $\boldsymbol{q}$ and energy exchange ω. The phonon frequency $\omega_{\boldsymbol{q}}$ is by definition always less than the Debye frequency ω_D. Hence, the effect of the phonons is almost irrelevant for energy exchanges ω that are much larger than ω_D. For such large frequencies, the time scale is too short for the slow ions to move and influence the interactions. However, the repulsive bare Coulomb potential is renormalized to an interaction which can actually be *attractive* for $\omega < \omega_D$. This can be understood physically as an electron temporarily polarizing the ion lattice, leaving an effective positive charge in its wake which attracts a second electron. This aspect of the electron-phonon interaction is crucial for superconductivity, as we will see in a later chapter.

4.4.7 Experimental measurement of phonon dispersion

Two common ways to measure the phonon dispersion (energy vs. momentum) are inelastic neutron scattering and X-ray scattering. The idea is that the neutrons [6] or X-ray photons will change their energy and momentum as they interact with the crystal lattice, and thus the phonons. By measuring the change in the energy and momentum of the neutrons and X-ray photons that emerge from the material after having interacted with it, one can map out the phonon dispersion relation.

4.5 Polarons

When an electron moves through a lattice, it will temporarily polarize the ion lattice, as discussed above. This causes a spatial deformation of the lattice and thus excites phonons. We now show that the electron moving through the lattice, inducing a phonon cloud in its wake, can be thought of as a new type of quasiparticle called a *polaron*. This polaron acts like an electron, but is heavier since it has to drag along with it a phonon cloud (which costs energy).

We begin by taking our Hamiltonian describing the interaction between phonons and electrons and inserting the quantized version of the displacements u_q. This is the Fourier-transform of $\hat{u}(x)$, omitting here the hat index for the operator:

$$u_q = \frac{1}{\sqrt{\Omega}}\sum_x \mathrm{e}^{-iqx}u(x) \tag{4.183}$$

for our simplified 1D model with longitudinal phonons. Straightforward insertion of this into Eq. (4.154) produces the so-called Frölich Hamiltonian:

$$H = \sum_k \epsilon_k c_k^\dagger c_k + \sum_q \hbar\omega_q b_q^\dagger b_q + \sum_{kq} M_q(b_{-q}^\dagger + b_q)c_{k+q}^\dagger c_k \tag{4.184}$$

[6] You may wonder how neutrons, void of any net charge, interact with the crystal lattice. This is via nucleon-nucleus scattering, which is primarily governed by the strong interaction (with some contribution from the electromagnetic and weak interaction, depending on the energy of the incident neutrons).

where the amplitude M_q describing the electron-phonon coupling is

$$M_q = iq\tilde{V}_{-q}\sqrt{2\frac{\hbar}{\rho_0\omega_k}}. \tag{4.185}$$

Note that we have omitted spin indices for simplicity, since spin is conserved during these electron-phonon interactions. Longitudinal phonons carry no spin, whereas transverse phonons (similarly to transversely polarized photons) carry spin $S = 1$.

The above Hamiltonian thus contains an unperturbed part, describing free electrons with dispersion ϵ_k and free phonons with dispersion ω_q, and a coupling term. In one case, $b_q c^\dagger_{k+q} c_k$ scatters an electron by the absorption of a phonon with momentum q. In the other case, $b^\dagger_{-q} c^\dagger_{k+q} c_k$ scatters an electron by emission of a phonon with momentum $-q$. The total momentum of the creation operators equal that of the annihilation operators. These momenta could *in principle* differ by a reciprocal lattice vector, corresponding to so-called Umklapp scattering, but we will neglect this. We now want to consider how the excitation spectrum of the electrons changing due to the interaction term.

Treating the interaction term as a perturbation, we can compute the change in total energy compared to the ground-state $|\Phi\rangle$ which has energy E_0, where $|\Phi\rangle$ is a state in Fock space comprised of a product of single-particle electron and phonon states. Thus, the total energy is $E_0 + E_2$ since the first order perturbation correction vanishes:

$$E_0 = \sum_k \epsilon_k \langle n_k\rangle + \sum_q \hbar\omega_q \langle N_q\rangle, \tag{4.186}$$

where $n_k = c^\dagger_k c_k$ and $N_q = b^\dagger_k b_k$, whereas the second order contribution is obtained as (see one of the exercises)

$$E_2 = \sum_{kq} |M_q|^2 \langle n_k(1-n_{k-q})\rangle \Big[\frac{\langle N_{-q}\rangle}{\epsilon_k - \epsilon_{k-q} + \hbar\omega_{-q}} + \frac{\langle N_q + 1\rangle}{\epsilon_k - \epsilon_{k-q} - \hbar\omega_q}\Big] \tag{4.187}$$

where $\langle \ldots \rangle$ denotes the expectation value of the operator in the unperturbed state $|\Phi_0\rangle$. For instance, $\langle n_k\rangle = f_k$ with f_k being the Fermi-Dirac distribution function.

Now, the electron part of the ground state energy E_0 is given by $\sum_q \epsilon_k \langle n_k\rangle$, so that the electron excitation energy is given by $\epsilon_k = \partial E_0/\partial\langle n_k\rangle$. This expresses the change in energy due to a change in electron occupation of the electron state with momentum $\boldsymbol{k}$. To find the effective energy ϵ'_k of an electron excitation when the interaction with phonons is taken into account, we can thus analogously compute

$$\begin{aligned}\epsilon'_k &= \frac{\partial(E_0 + E_2)}{\partial\langle n_k\rangle} \\ &= \epsilon_k + \sum_{k'} |M_{k'-k}|^2 \Big[\frac{1-\langle n_{k'}\rangle}{\epsilon_k - \epsilon_{k'} - \hbar\omega_q} - \frac{\langle n_{k'}\rangle}{\epsilon_{k'} - \epsilon_k - \hbar\omega_q}\Big]\end{aligned} \tag{4.188}$$

where we assumed low temperatures so that $\langle N_q\rangle \ll 1$. The electron excitation energy is modified by the presence of scattering on phonons. This changes the group velocity of the electrons from its

bare value $\hbar v_k^0 = \partial\epsilon_k/\partial k = \hbar^2 k/m$. Instead, one finds after some algebra and approximations (not curriculum):

$$\hbar v_k = \partial\epsilon_k'/\partial k = v_k^0(1-\alpha). \tag{4.189}$$

where α is a positive parameter proportional to the momentum-averaged value of the scattering amplitude M_q and the electron density of states at the Fermi level. This is equivalent to an effective electron mass of

$$m' = \frac{m}{1-\alpha} > m. \tag{4.190}$$

The polaron particle this appears heavier than the bare electron. The extra mass comes about due to the interaction between the electron and the ions. The electron induces a change in the spatial position of the ions in its vicinity. This change in position corresponds to an excitation of phonon modes. Namely, quantum mechanically the amplitude of the ion displacements are not allowed to take continuous values, but only discrete amounts. When the electron moves through the lattice with a given momentum, it thus drags this ionic deformation along with it, which costs energy. The drag effect is what causes the mass to increase of the effective quasiparticle. The polaron can thus be thought of as an electron dressed with lattice polarization (a cloud of virtual phonons).

4.5.1 Experimental measurement of polaron dispersion

Since polarons are electrons "dressed" with phonons, one can use the previously mentioned ARPES to measure the influence of phonons on the electron energies. The emission of photoelectrons is very fast compared to the sluggish motion of atoms, meaning that the lattice does not have time to relax while the electron is being removed. Therefore, the energy measured by the photoemission corresponds to the polaronic energy. More specifically, one has $\hbar\omega - E_k - \Phi = E_{\text{pol}}$ where ω is the photon energy, E_k is the kinetic energy of the emitted electron, and Φ is the work function of the material.

4.6 Polaritons

Polaritons is a type of excitation which results from the coupling of electromagnetic waves with an electric or magnetic-dipole carrying excitation, such as phonons or excitons. Polaritons are collective excitations (bosonic quasiparticle) and is thus different from the polaron (fermionic quasiparticle), which we have seen is an electron plus an attached phonon cloud.

Whenever the electromagnetic field couples sufficiently strongly to *e.g.* the phonon, plasmon, or exciton, the model of photons propagating freely in the crystals is inadequate. An important characteristic of polaritons is the strong dependency of their propagation speed (i.e. velocity of light) through the crystal on the frequency of the photon.

Here, we consider in more detail the phonon-polariton which describes excitations due to the coupling of phonons and photons (light). Other types of polaritons exist as well, such as exciton-polaritons.

4.6.1 Phonon polariton

These are quasiparticles generated in solids when photons couple to transverse optical phonons with the same frequency and wavevector. The presence of an optical phonon branch requires that there

is at least two atoms in the unit cell basis. Let us first show this before moving on to the polaritons.

Consider a linear chain of atoms with distance a between each atom. The chain consists of alternating atoms of different types: one with mass m and one with mass M. Thus, we can in the end recover the case of a monoatomic chain if we set $m = M$. For now, keep the masses different. Let the atoms on even sites have mass m and the ones on odd sites have M. Assume that we have a harmonic oscillator type restoring potential between each pair of atoms and set the spring constant κ in this potential to equal for all pairs for simplicity. Let $u_n(t) = x_n(t) - na$ be the deviation from equilibrium for the n-th atom. The Hamiltonian for this system is then

$$H = \sum_j \frac{p_j^2}{2m_j} + \frac{\kappa}{2}\sum_j (u_j - u_{j-1})^2 \tag{4.191}$$

where m_j is the mass of the atom on site j and $p = m\dot{u}_n$. From Hamilton's equation of motion, we have $\dot{p}_n = -\frac{\partial H}{\partial u_n}$. This gives us the equations of motion

$$\begin{aligned} m\ddot{u}_{2n} &= -\kappa(2u_{2n} - u_{2n-1} - u_{2n+1}), \\ M\ddot{u}_{2n+1} &= -\kappa(2u_{2n+1} - u_{2n} - u_{2n+2}). \end{aligned} \tag{4.192}$$

We impose periodic boundary conditions $u_{n+N} = u_n$ for a chain of total length of N sites. For a large value of N, the precise choice of boundary conditions becomes insignificant. Using the a plane-wave ansatz for the solutions:

$$u_{2n} = A\mathrm{e}^{2ikna - i\omega t},\ u_{2n+1} = B\mathrm{e}^{2ikna - i\omega t}, \tag{4.193}$$

we see that the solutions are invariant under $k \to k + \pi/a$. This means that the 1st Brillouin zone corresponds to $k = \in \left[-\frac{\pi}{2a}, \frac{\pi}{2a}\right)$ which is half the size of a monoatomic chain. This is reasonable since the periodicity of the lattice has been doubled due to the diatomic basis. Plugging the ansatz into the equations of motion gives us a relation between the amplitudes A, B:

$$\omega^2 \begin{pmatrix} m & 0 \\ 0 & M \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \kappa \begin{pmatrix} 2 & -(1 + \mathrm{e}^{2ika}) \\ -(1 + \mathrm{e}^{-2ika}) & 2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}. \tag{4.194}$$

This is an eigenvalue equation for the phonon dispersion ω. It is solved in the standard way of rewriting the above equation to the form $Px = 0$ where the determinant of the matrix P has to be zero, providing

$$\omega^2 = \omega_\pm^2 = \frac{\kappa}{mM}\Big[M + m \pm \sqrt{(M-m)^2 + 4Mm\cos^2(ka)}\Big]. \tag{4.195}$$

Thus, we obtain two phonon branches: $\omega_\pm$. The key thing to note here is that there exists an energy gap between them. At the edge of the BZ $k = \pm\pi/2a$, we obtain a gap

$$\hbar(\omega_+ - \omega_-)|_{k=\pm\pi/2a} = \hbar\sqrt{2\kappa}\left|\frac{1}{\sqrt{m}} - \frac{1}{\sqrt{M}}\right|. \tag{4.196}$$

Thus, the upper branch ω_+ is the optical phonon whereas ω_- is the acoustic phonon. Note how the gap vanishes when $m = M$, and we only obtain an acoustic phonon ω_- which has a vanishing dispersion $\omega_- \to 0$ as $k \to 0$. But what about ω_+? It seemingly does not vanish at $k \to 0$ even if $M = m$. The explanation is that for $M = m$, ω_+ is just the part of ω_- that would belong to the

regimes $k \in \left[-\frac{\pi}{a}, \frac{\pi}{2a}\right), \left[\frac{\pi}{2a}, \frac{\pi}{a}\right)$ if we had done the analysis for a monoatomic chain, in which case the allowed k-values would have been $\left[-\frac{\pi}{a}, \frac{\pi}{a}\right)$. So for $M = m$, ω_+ is just the acoustic phonon dispersion folded back into the BZ corresponding to the diatomic case.

We can also see from the above analysis what kind of motion the acoustic and optical branches of the phonons correspond to in the case $m \neq M$. This is seen most easily by considering $k = 0$. In this case, the acoustic branch $\omega_- = 0$ and has eigenvector $(A, B) = (1, 1)$. The atoms thus move in phase in the acoustic branch. However, in the optical branch, we have $\omega_+ = \sqrt{2\kappa(m^{-1} + M^{-1})}$ with eigenvector $(A, B) = (M, -m)$. In the optical branch, the atoms therefore move out of phase.

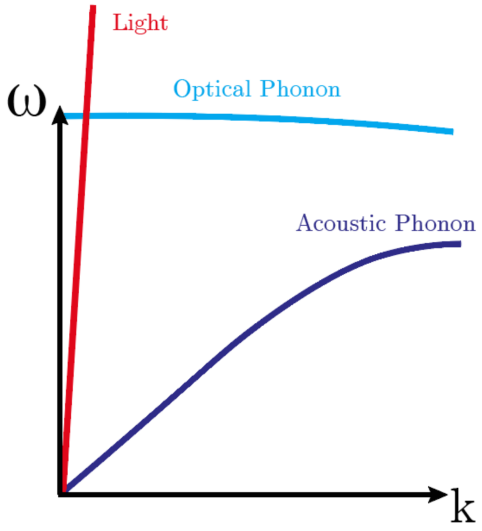

Figure 4.8: Qualitative sketch of phonon and photon dispersion relation. Figure taken from Physics Stack Exchange.

Let us now return to the phonon polaritons. Fig. 4.8 shows a qualitative sketch of a photon dispersion relation and both acoustical and optical phonon excitations. The steep photon dispersion comes from the fact that light travels much faster than sound. At the point where the photon dispersion crosses the optical phonon branch, polaritons can be formed. This typically occurs at very small k-values due to the large speed of light.

Let us derive the essential features of the dispersion relation of polaritons, which are thus hybridized phonon-photon excitations, from a simple model. To do so, consider first the optical phonon which at $k \to 0$ (meaning a small energy and a resulting small displacement, causing a restoring force which is linear in the displacement) can be modelled as a charge on a spring, assuming the atoms in the basis are charged. This charge satisfies the equation

$$m\frac{d^2x}{dt^2} = -eE - Cx \tag{4.197}$$

where x is the displacement of the charge, E is the electric field, and C is the spring constant of the system. This causes an atom oscillation with the phonon frequency $\omega_T = \sqrt{C/m}$. The oscillation

of charges also causes an electric polarization

$$P = -Nex \tag{4.198}$$

where N is the number of displaced atoms. Assuming a plane-wave trial solution for x, E, P, so that $x = x_0 \mathrm{e}^{i(ky-\omega t)}$ + h.c. and similarly for the other quantities, we obtain

$$\frac{Ne^2}{m}E + (\omega^2 - \omega_T^2)P = 0. \tag{4.199}$$

Now we consider the equation obeyed by the photon in matter. Maxwell's equation reads (in the absence of free charges and currents):

$$\nabla^2 E = \mu_r \mu_0 \frac{\partial^2 D}{\partial t^2} \tag{4.200}$$

where μ_r is the relative permeability, which we set to unity for simplicity. A few comments are in order here. First, we wrote the displacement of the atoms (x) above as a plane-wave propagating in the y-direction. This corresponds to transverse optical phonons. For such phonons, $\nabla \cdot \boldsymbol{E} = 0$, as explained in Fig. 4.9. The form of Maxwell's equation in Eq. (4.200) is only valid precisely for $\nabla \cdot \boldsymbol{E} = 0$, so our treatment is only valid precisely for transverse optical phonons.

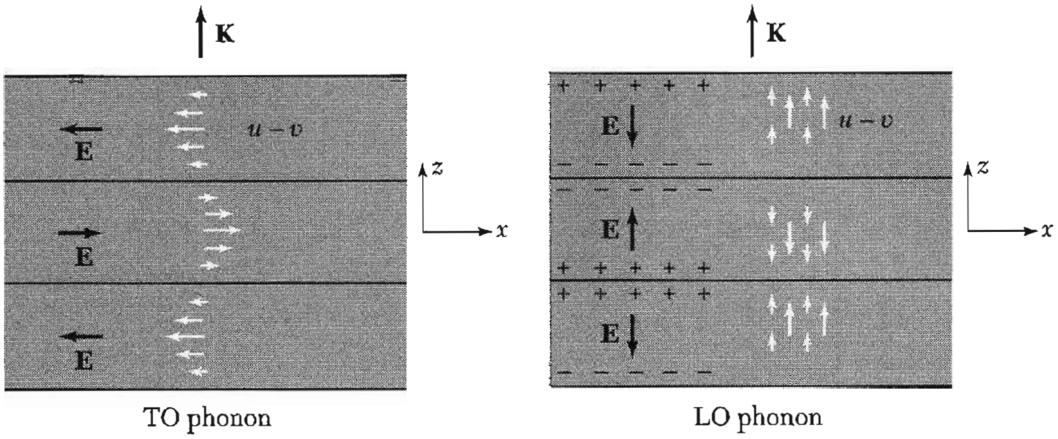

Figure 4.9: Relative displacements of the positive and negative ions at a given instant of time for an optical phonon travelling along the z-axis. The planes of nodes (zero displacement) are shown. For long wavelength phonons, the nodal planes are separated by many planes of atoms. In the transverse optical phonon mode, the particle displacement is perpendicular to the wavevector $\boldsymbol{K}$. The macroscopic electric field in an infinite medium will lie only in the $\pm x$ direction for the mode shown, and it follows that $\nabla \cdot \boldsymbol{E} = 0$. In the longitudinal optical phonon mode, the particle displacements and hence dielectric polarization $\boldsymbol{P}$ are parallell to the wavevector, so that $\nabla \cdot \boldsymbol{E} \neq 0$. Figure taken from Ref. [7].

Now use the relation between the electric displacement field and the electric field

$$D = \epsilon_0 E + P. \tag{4.201}$$

Employing again our plane-wave trial solution, we obtain:

$$(\mu_0\epsilon_0\omega^2 - k^2)E + \mu_0\omega^2 P = 0. \tag{4.202}$$

Now we have two coupled equations for E and P. A non-trivial solution exists for the dispersion relation $\omega(k)$ if the following determinant is zero (corresponding to the linear equation system):

$$\begin{vmatrix} \mu_0\epsilon_0\omega^2 - k^2 & \mu_0\omega^2 \\ \frac{Ne^2}{m} & \omega^2 - \omega_T^2 \end{vmatrix} \tag{4.203}$$

Two solutions exist for ω for each k-value, and upper and a lower branch. Fig. 4.10 shows the dispersion of light, the optical phonon branch, and the lower and upper branch of the new dispersion relation we have obtained through the coupling of light and sound. The acoustical branch is not shown, since it does not participate in the polariton formation.

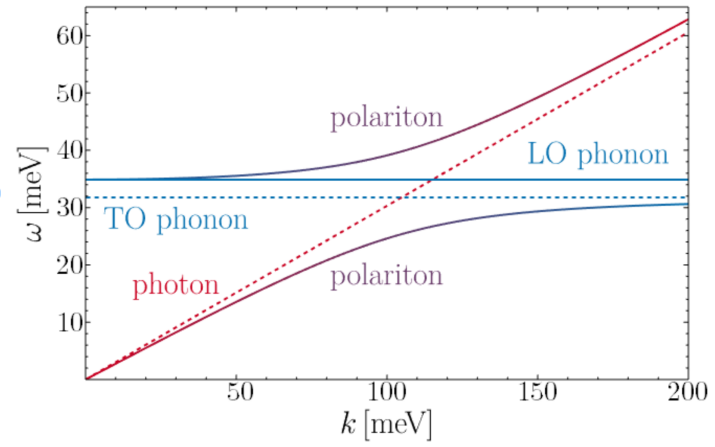

Figure 4.10: Photon, optical phonon, and polariton branches in GaAs. Figure taken from A. Mitridate *et al.* , Phys. Rev. D **102**, 095005 (2020).

The lower branch of the new dispersion is first photon-like, since it follows the photon-dispersion, until it comes close to the region where the photon and phonon branches would normally cross. The lower branch then starts to progressively look more and more similar to the phonon dispersion. The opposite happens for the upper branch. Using a second quantized formalism, both phonons and photons are bosonic quasiparticles, and so polaritons are also bosonic.

We see that the upper branch begins at the longitudinal optical phonon dispersion. To see why this is so, recall from Fig. 4.9 that longitudinal phonons have $\nabla \cdot \boldsymbol{E} \neq 0$, but $\nabla \cdot \boldsymbol{D} = 0$ (no free charges). Since $\boldsymbol{D} = \varepsilon(\omega)\boldsymbol{E}$ in general, the only way to satisfy this equation is to have $\varepsilon(\omega) = 0$. The zeros of the effective permittivity thus determine the longitudinal optical phonon energy ω_{LO} at $k \to 0$. We can obtain an analytical expression for $\varepsilon(\omega)$, and thus identify ω_{LO}, as follows.

Since $D = \varepsilon(\omega)E$ and $D = \varepsilon_0 E + P$, we have that

$$\varepsilon(\omega) = \varepsilon_0 + \frac{P}{E}. \tag{4.204}$$

From Eq. (4.199), we find P/E and insert it into the above equation:

$$\varepsilon(\omega) = \varepsilon(\infty) + \frac{Ne^2/m}{\omega_T^2 - \omega^2} \tag{4.205}$$

where we used that $\varepsilon_0 = \varepsilon(\infty)$. We can rewrite this as

$$\varepsilon(\omega) = \varepsilon(\infty) + [\varepsilon(0) - \varepsilon(\infty)]\frac{\omega_T^2}{\omega_T^2 - \omega^2}. \tag{4.206}$$

Since we know that the longitudinal optical phonon energy is defined by $\varepsilon(\omega = \omega_L) = 0$, it follows that

$$\varepsilon(\infty)\omega_L^2 = \varepsilon(0)\omega_T^2. \tag{4.207}$$

This value for ω_L can be verified to be exactly where the upper polariton branch begins at $k = 0$ from our determinant expression above, where one obtains the solutions

$$\omega^2 = \pm\Big(\omega_T^2 + \frac{Ne^2}{m\epsilon_0}\Big). \tag{4.208}$$

Thus, at k-values near the avoided crossing, the dispersion neither looks like photons or phonons - here, we talk about polaritons. A gap in the dispersion relation arises due to the avoided crossing. As a consequence, light will reflect (via Bragg scattering) more effectively for frequencies inside the gap, since there are no available modes present in the material. For instance, in the material AlAs light frequencies between about 11 THz and 12.3 THz will be reflected.

4.6.2 Experimental measurement of polariton dispersion

The way to observe polaritons is therefore via optical measurements, i.e. to look at the reflectance of the material under investigation. One of the reflectance peaks correspond to the band gap which arises due to the formation of polaritons. One can also use Raman spectroscopy to map out a more complete energy-momentum relation for the polaritons. In such a scenario, a laser is pointed at the material being studied. Choosing a suitable laser wavelength, this laser can induce the formation of a polariton in the sample. Looking at the Stokes shifted [7] emitted radiation and by using the conservation of energy and the known laser energy, one can calculate the polariton energy, from which one can construct the dispersion relation.

[7] Stokes shift is the difference in energy between positions of the maxima of the absorption and emission spectra of the same electronic transition. When a system, such as an atom, absorbs a photon, it gains energy and enters an excited state. The system can relax by emitting a photon. The Stokes shift occurs when the energy of the emitted photon is lower than that of the absorbed photon, representing the difference in energy of the two photons.

Chapter 5

Dielectrics and ferroelectrics

To understand dielectric and ferroelectric materials, we must first understand the relation between an applied electric field and the internal electric field in a crystal. In particular, what is the relation in the material between the polarization $\boldsymbol{P}$ and the macroscopic electric field $\boldsymbol{E}$ in Maxwell's equations? And what is the relation between the polarization and the local electric field which acts at the site of an attom in the lattice?

We will begin this chapter with a few sections briefly reviewing basic concepts that you partially may have seen in other courses, and then move on to more advanced topics.

5.1 Polarization

The polarization $\boldsymbol{P}$ is defined as the electric dipole moment per unit volume. The total dipole moment is defined as

$$\boldsymbol{p} = \sum_n q_n \boldsymbol{r}_n \tag{5.1}$$

where r_n is the position vector of the charge q_n and the summation goes over all charges n. The value of the sum is independent of the origin chosen for the position vectors, provided that the system is neutral: $\sum_n q_n = 0$. This can be seen by letting $\boldsymbol{r}'_n = \boldsymbol{r}_n + \boldsymbol{R}$ where $\boldsymbol{R}$ is some arbitrary vector and computing $\boldsymbol{p} = \sum_n q_n \boldsymbol{r}'_n$.

The electric field at a point $\boldsymbol{r}$ from a dipole moment $\boldsymbol{p}$ is given by a standard result of elementary electrostatics (see for instance the Griffiths book):

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{3(\boldsymbol{p}\cdot\boldsymbol{r})\boldsymbol{r} - r^2\boldsymbol{p}}{4\pi\varepsilon_0 r^5}. \tag{5.2}$$

The forcelines of a dipole pointing along the z-axis are shown in Fig. 5.1.

5.2 Electric field properties

The total electric field inside a body has two contributions: the applied electric field $\boldsymbol{E}_0$, defined as the field produced by fixed charges external to the body. The other contribution is the sum of the electric fields of all the charges that constitute the body itself. If the body is neutral, the

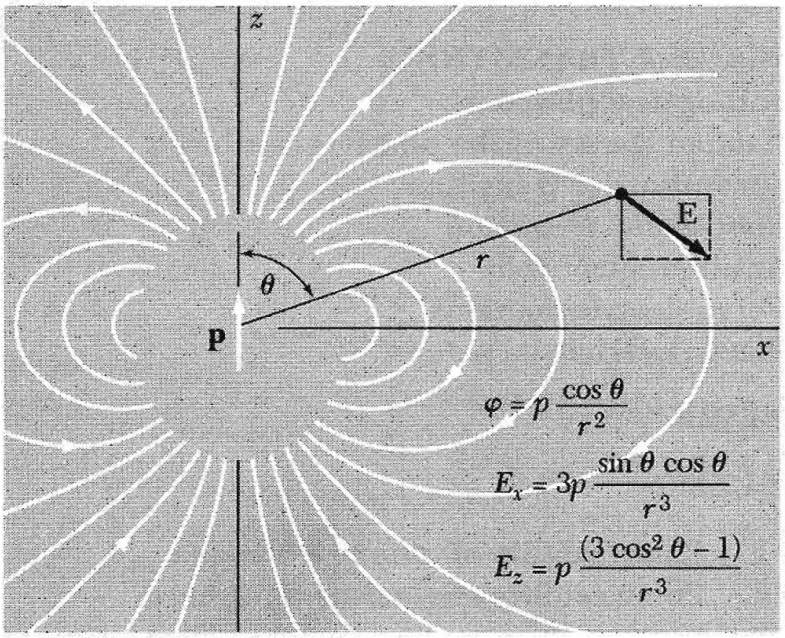

Figure 5.1: Electrostatic potential and field components in CGS unit at position r, θ for a dipole $\boldsymbol{p}$ directed along the z-axis. To get SI units, replace p by $p/4\pi\varepsilon_0$. Figure taken from [7].

contribution to the average field may be expressed in terms of the sum of the fields of atomic dipoles.

We define the average electric field $\boldsymbol{E}(\boldsymbol{r}_0)$ as the average field over the volume of the crystal cell that contains the lattice point $\boldsymbol{r}_0$:

$$\boldsymbol{E}(\boldsymbol{r}_0) = \frac{1}{V_c}\int dV\, \boldsymbol{E}_{\text{local}}(\boldsymbol{r}). \tag{5.3}$$

Here, $\boldsymbol{E}_{\text{local}}(\boldsymbol{r})$ is the microscopic electric field at the point $\boldsymbol{r}$. The field $\boldsymbol{E}$ is a much smoother quantity than $\boldsymbol{E}_{\text{local}}$, since it is averaged. The dipole field given in the previous section is precisely $\boldsymbol{E}_{\text{local}}(\boldsymbol{r})$ since it is a microscopic unsmoothed field.

We refer to $\boldsymbol{E}$ as the macroscopic electric field. It is the adequate quantity to consider for all problems in the electrodynamics of the crystals provided that we know the connection between $\boldsymbol{E}$, the polarization $\boldsymbol{P}$, and the current density $\boldsymbol{j}$, and provided that the wavelengths of interest are long in comparison with the lattice spacing.

To find the contribution of the polarization to the macroscopic electric field, one can simplify the sum over all the dipoles in the body. By a famous theorem in electrostatics, the macroscopic electric field caused by a uniform polarization is equal to the electric field in vacuum of a fictitious surface charge density $\sigma = \hat{\boldsymbol{n}} \cdot \boldsymbol{P}$ on the surface of the body [1] . Here, $\hat{\boldsymbol{n}}$ is the unit normal to the surface, drawn outward from the polarized matter. To see how this works, we can apply the result to a

[1]For a single dipole $\boldsymbol{p}$, the electric potential is $V(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\frac{\hat{\boldsymbol{r}}\cdot\boldsymbol{p}}{r^2}$ assuming that the separation distance d between the opposite charges in the dipole is small compared to r. This is seen by Taylor expanding the combined potential of two opposite charges separated by d. For an infinitesimal dipole moment $d\boldsymbol{p} = \boldsymbol{P}d\Omega$, where $d\Omega$ is an infinitesimal

thin dielectric slab as shown in Fig. 5.2(a) with a uniform polariation $\boldsymbol{P}$. The electric field $\boldsymbol{E}_1(\boldsymbol{r})$ produced by the polarization is equal to the field produced by the fictitious surface charge density σ on the surface of the slab.

The electric field $\boldsymbol{E}_1$ due to these charges has a simple form at any point between the plates, bu comfortably removed from their edges. Namely, using Gauss' law $E_1 = -|\sigma|/\varepsilon_0 = -P/\varepsilon_0$. Thus, the total macroscopic field inside the slab with $\hat{\boldsymbol{z}}$ as the unit vector normal to the plane is:

$$\boldsymbol{E} = \boldsymbol{E}_0 + \boldsymbol{E}_1 = \boldsymbol{E}_0 - \frac{P}{\varepsilon_0}\hat{\boldsymbol{z}}. \tag{5.8}$$

Therefore, $\boldsymbol{E}_1$ is the field of the surface charge densities and is smoothly varying in space inside and outside of the body. It is smooth indeed when viewed even on an atomic scale since we have replaced the discrete lattice dipoles $\boldsymbol{p}_j$ in the material with the smoothed polarization $\boldsymbol{P}$.
The field $\boldsymbol{E}_1$ is known as the *depolarization field*, since it tends to oppose the applied field $\boldsymbol{E}_0$ inside the body, shown in Fig. 5.3. Specimens in the shape of ellipsoids, a class that includes spheres, cylinders, discs, has the advntageous property that a uniform polarization produces a uniform depolarization field inside the body.

Generally, if P_j $(j = x, y, z)$ are the components of $\boldsymbol{P}$ relative the principal axes of an ellipsoid, the components of the depolarization field are

$$E_{1j} = -\frac{N_j P_j}{\varepsilon_0}. \tag{5.9}$$

Here, N_j are depolarization factors: their specific values depend on the ratios of the principal axes. They are positive and satisfy $\sum_j N_j = 1$. For a sphere, $N_j = 1/3$. Values of N parallell to the figure axis of ellipsoids of revolution are plotted in Fig. 5.3. The depolarization field can be suppressed to zero in two ways: either by working parallel to the axis of a long specimen or by making an electrical connection between electrodes deposited on the opposite surfaces of a thin slab.

---

volume element, the total potential is therefore

$$V(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\hat{\boldsymbol{\rho}}\cdot d\boldsymbol{p}}{\rho^2} = \frac{1}{4\pi\epsilon_0}\int \frac{\hat{\boldsymbol{\rho}}\cdot \boldsymbol{P}(\boldsymbol{r}')}{\rho^2} d\Omega'. \tag{5.4}$$

Here, we defined the coordinate $\boldsymbol{\rho} = \boldsymbol{r} - \boldsymbol{r}'$ where the source coordinate of the dipole is $\boldsymbol{r}'$. The integral thus goes over the entire volume of the material, adding all contributions from infinitesimal dipole moments at various $\boldsymbol{r}'$ to produce the total potential at $\boldsymbol{r}$. Now, we note that $\nabla'(1/\rho) = \hat{\boldsymbol{\rho}}/\rho^2$, where we have differentiated with respect to the source coordinate. Therefore, we may write

$$V = \frac{1}{4\pi\epsilon_0}\int \boldsymbol{P}\cdot\nabla'\Big(\frac{1}{\rho}\Big) d\Omega'. \tag{5.5}$$

Using the product rule $\boldsymbol{P}\cdot\nabla'(1/\rho) = \nabla'\cdot(\boldsymbol{P}/\rho) - (1/\rho)\nabla'\cdot\boldsymbol{P}$ and Gauss law, we obtain

$$V = \frac{1}{4\pi\epsilon_0}\oint \frac{\boldsymbol{P}}{\rho} d\boldsymbol{A}' + \frac{1}{4\pi\epsilon_0}\int \frac{(-\nabla'\cdot\boldsymbol{P})}{\rho} d\Omega'. \tag{5.6}$$

The first integral goes over the surface of the sample $d\boldsymbol{A}' = \hat{\boldsymbol{n}} dA'$ and the second over the volume. For a uniform polarization $\boldsymbol{P}$, the second term vanishes. We can then write the above in terms of an effective surface charge $\sigma_b \equiv \boldsymbol{P}\cdot\hat{\boldsymbol{n}}$:

$$V = \frac{1}{4\pi\epsilon_0}\oint \frac{\sigma_b}{\rho} dA'. \tag{5.7}$$

which completes the proof.

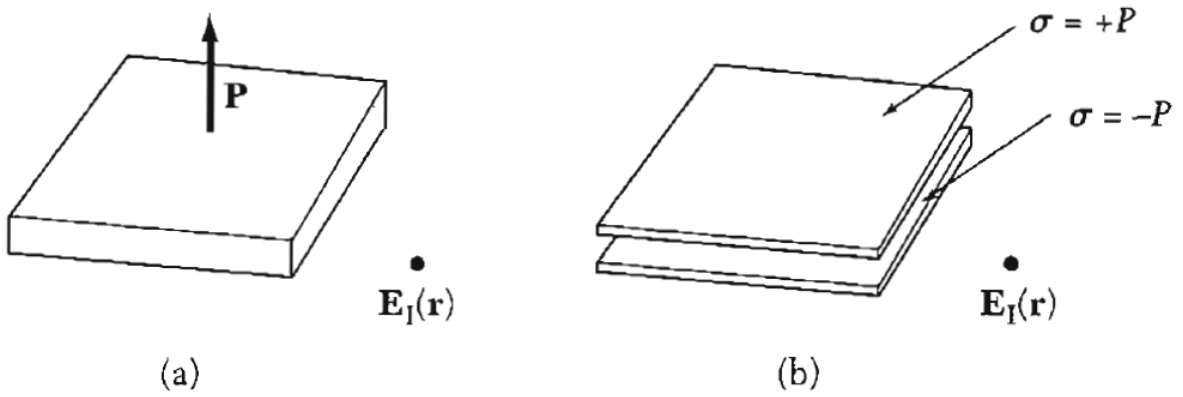

Figure 5.2: **(a)** A uniformly polarized dielectric slab, with the polarization vector $\boldsymbol{P}$ normal to the plane of the slab. **(b)** A pair of uniformly charged parallel plates which give rise to the identical electric field $\boldsymbol{E}_1$ as in (a). The upper plate has surface charge density $\sigma = +P$ and the lower $\sigma = -P$. Figure taken from [7].

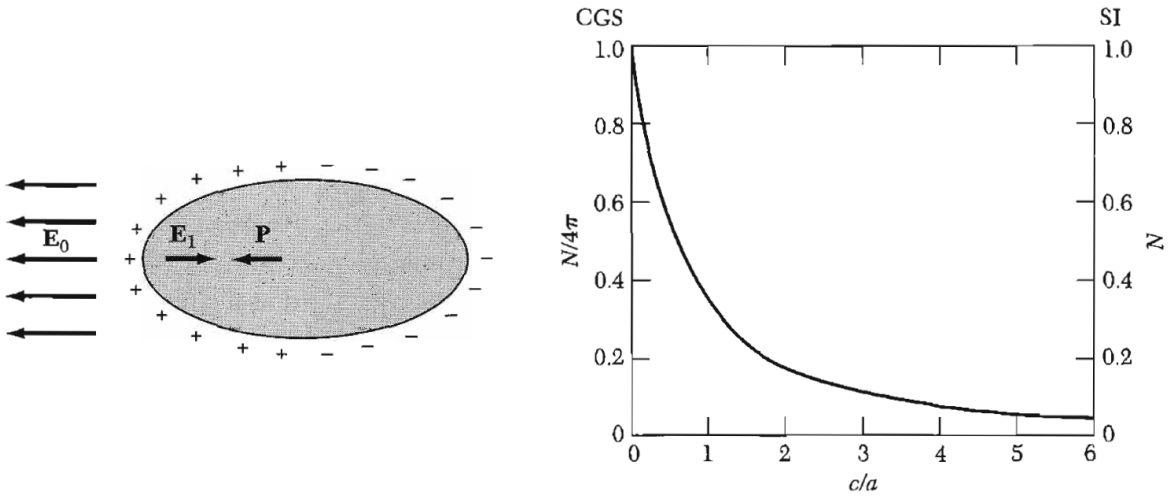

Figure 5.3: **Left:** the depolarization field $\boldsymbol{E}_1$ is opposite to $\boldsymbol{P}$. The fictious surface charges are indicated. The field of these charges is $\boldsymbol{E}_1$ within the ellipsoid. **Right:** depolarization factor N parallel to the figure axis of ellipsoids of revolution, as a function of the axial ratio c/a. Figure taken from [7].

A uniform electric field $\boldsymbol{E}_0$ will induce a uniform polarization in an ellipsoid. The dielectric susceptibility χ is defined via

$$\boldsymbol{P} = \varepsilon_0 \chi \boldsymbol{E} \tag{5.10}$$

and can in general be a tensor. It connects the macroscopic field $\boldsymbol{E}$ inside the ellipsoid with the

polarization $\boldsymbol{P}$. In the special case where $\boldsymbol{E}_0$ is uniform and parallel to a principal axis of the ellipsoid, then $E = E_0 - NP/\varepsilon_0$, as we have seen, and thus

$$P = \frac{\chi\varepsilon_0}{1 + N\chi} E_0 \tag{5.11}$$

so that the value of the polarization depends on the depolarization factor N.

Before proceeding to discuss the polarizibility of a material, we need to talk about the *local electric field of an atom*. This is generally different from the macroscopic electric field $\boldsymbol{E}$. As an example, consider the local field at a site with a cubir arrangement of neighbors in a crystal, and the crystal itself has a spherical shape. The macroscopic electric field in a sphere is

$$\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2 = \boldsymbol{E}_0 - \frac{1}{3\varepsilon_0}\boldsymbol{P}. \tag{5.12}$$

But consider now instead the field that acts on the atom at the center ofthe sphere. If all dipoles are parallell to the z-axis and have magnitude p, the z-component of the field at the center due to all other dipoles is according to Eq. (refeq:Eduetodipole):

$$E_{\text{dipole}} = \frac{p}{4\pi\varepsilon_0} \sum \frac{3z_i^2 - r_i^2}{r_i^5}. \tag{5.13}$$

The summation over the x, y, z directions are equivalent because of the symmetry of the lattice and of the sphere:

$$\sum_i \frac{z_i^2}{r_i^5} = \sum_i \frac{y_i^2}{r_i^5} = \sum_i \frac{x_i^2}{r_i^5} \tag{5.14}$$

from which we may conclude that $E_{\text{dipole}} = 0$. The correct local field is there just the applied field $\boldsymbol{E}_{\text{local}} = \boldsymbol{E}_0$ for an atomic site with a cubic environment in a spherical specimen. This is different from the macroscopic average field $\boldsymbol{E}$ above, which includes the depolarizing field.

5.3 Dielectric constant and polarizibility

The (relative) dielectric constant ε_r of an isotropic or cubic medium relative vacuum is defined in terms of the macroscopic field E

$$\varepsilon_r = \frac{\varepsilon_0 E + P}{\varepsilon_0 E} = 1 + \chi. \tag{5.15}$$

The susceptibility χ is related to the dielectric constant via

$$\chi = \frac{P}{\varepsilon_0 E} = \varepsilon - 1. \tag{5.16}$$

In a noncubic crystal, the dielectric response is instead described by the components of the susceptibility tensor or of the dielectric constant tensor

$$P_\mu = \chi_{\mu\nu}\varepsilon_0 E_\nu,\ \varepsilon_{\mu\nu} = \delta_{\mu\nu} = \delta_{\mu\nu} + \chi_{\mu\nu}. \tag{5.17}$$

The polarizibility α of an atom can now be defined in terms of the local electric field of the atom:

$$p = \alpha E_{\text{local}} \tag{5.18}$$

where p is the dipole moment of the atom. This polarizibility is thus an atomic property, but the dielectric constant will depend on the manner in which the atoms are assembled to form a crystal. For non-spherical atom, α will be a tensor. Polarizability thus to the tendency of matter, when subject to an electric field, to acquire an electric dipole moment in proportion to that applied field: the polarizibility is responsible for a material's dielectric constant.

The polarization of a crystal may be expressed approximately as the product of the polarizibilities of the atoms times the local electric field:

$$P = \sum_j n_j p_j = \sum_j n_j \alpha_j E_{\text{loc}}(j) \tag{5.19}$$

where n_j is the concentration of atoms (number per volume) and α_j the polarizibility of atoms j, while $E_{\text{loc}}(j)$ is the local field at atomic site j.
To relate the dielectric constant to the polarizibility, the result will depend on the relation that holds between the macroscopic electric field and the local electric field. We use the previous relation found for a sphere, Eq. (5.12), to express

$$P = \sum_j n_j \alpha_j (E + P/3\varepsilon_0). \tag{5.20}$$

Then, we may solve for P to find the electric susceptibility

$$\chi = \frac{P}{\varepsilon_0 E} = \frac{\sum_j n_j \alpha_j}{\varepsilon_0 - \frac{1}{3}\sum_j n_j \alpha_j}. \tag{5.21}$$

Using the definition $\varepsilon_r = 1 + \chi$, we finally obtain the so-called Clausius-Mossotti relation

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{1}{3\varepsilon_0} \sum_j n_j \alpha_j. \tag{5.22}$$

It relates the dielectric constant to the electronic polarizibility for crystals for which the Lorentz local field relation Eq. (5.12) holds. It can also be viewed as connecting the bulk behaviour (polarization density due to an external electric field according to the electric susceptibility $\chi = \varepsilon_R - 1$) with the molecular polarizability α due to the local field.

The total polarizibility α generally depends on the frequency of the local field ω and may usually be separated into three parts: electronic, ionic, and dipolar. These contributions dictate to what extent a dipole moment p is formed due to the local electric field. The electronic contribution arises from the displacement of the electron shell relative to a nucleus due to the field; the ionic contribution comes from the displacement of a charged ion with respect to other ions; the dipolar polarizibility arises from molecules with a permanent electric dipole moment that can change orientation in an applied electric field. In general, the polarizibility is a tensor.

We can give an approximate (classical) estimate of the electronic contribution to the polarizibility. An electron bound harmonically to an atomic will show a resonance absorption at a frequency $\omega_0 = \sqrt{\beta/m}$. The displacement x of the electron caused by the application of the field E_{loc} is:

$$-eE_{\text{loc}} = \beta x \tag{5.23}$$

so that the static electronic polarizibility is

$$\alpha_{\rm el} = p/E_{\rm loc} = -ex/E_{.\rm oc} = e^2/m\omega_0^2. \tag{5.24}$$

As mentioned above, the electronic polarizibility will depend on frequency. Consider a time-dependent local electric field (sinusoidal) so that the equation of motion becomes

$$m\frac{d^2x}{dt^2} + m\omega_0^2 x = -eE_{\rm loc}\sin(\omega t). \tag{5.25}$$

For $x = x_0\sin(\omega t)$, we thus obtain

$$m(-\omega^2 + \omega_0^2)x_0 = -eE_{\rm loc}. \tag{5.26}$$

The dipole moment has the amplitude

$$p_0 = -ex_0 = \frac{e^2E_{\rm loc}}{m(\omega_0^2 - \omega^2)}, \tag{5.27}$$

which gives the frequency dependent polarizibility

$$\alpha_{\rm el}(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2}. \tag{5.28}$$

5.4 Structural phase transitions

It is common for crystals to transform from one crystal structure to another as the temperature or pressure is varied. A stable structure, call it A, at absolute zero generally has the lowest accessible internal energy of all possible structures. However, the structure A can be varied with application of pressure. This is because a low atomic volume will favor closer-packed structures, turning even hydrogen into a metal under extreme pressure.

Another structure, call it B, may have a softer or lower frequency phonon spectrum than A. As temperature is increased, the phonons in B will be more highly excited (higher thermal average occupancy) than the phonons in A. Since entropy increases with the occupancy, the entropy of B will become higher than the entropy of A as T increases. Then it is possible for the stable structure to transform from A to B with increasing temperature. This is because the most stable structure at any given, constant temperature T is determined by the minimum of the Helmholtz free energy $F = U - TS$. Thus, a transition from structure A to B will take if a temperature T_c exists (below the melting point of the material) such that $F_A(T_c) = F_B(T_c)$.

Often times, several structures have nearly the same internal energy at absolute zero. The phonon dispersion relations for the structures may nevertheless be rather different. The phonon energies are sensitive to the number and arrangement of the nearby atoms, and these quantities change as the structure is changed.

Some structural phase transitions only have a small effect on the macroscopic physical properties of the material. Other structural phase transitions can have a spectacular effect on the macroscopic electrical properties. Ferroelectric transitions are precisely such a subgroup of structural phase transitions: they are marked by the appearance of a spontaneous dielectric polarization in

the crystal. Ferroelectrics are of high interest, both theoretically and practically, as they display electro-optical effects, piezoelectric effect, pyroelectric effect, and often have unusually high and strongly temperature-dependent values of the dielectric constant. We now look deeper into this class of materials.

5.5 Ferroelectric crystals

A ferroelectric crystal exhibits an electric dipole moment even in the absence of an external electric field. Let us first examine this in terms of symmetry which is a very important concept in physics and is a fundamental way of expressing its laws since it is related to the conservation of quantities such as energy, momentum and charge. Symmetry breaking is equally as important though and can make a system more complex. In a ferroelectric crystals, inversion symmetry must be broken in the ferroelectric state. This means that the unit cell of the material does not have parity symmetry: there exists no point in the unit cell around which you can perform the transformation $\boldsymbol{r} \rightarrow -\boldsymbol{r}$ and obtain the same arrangement of atoms that you had in the first place. In the ferroelectric state of a crystal, the center of positive charge does not coincide with the center of negative charge. Ferroelectricity is in many ways analogous to ferromagnetism, with electric polarization playing the role of magnetization.

However, noncentrosymmetricity is not the only requirement for ferroelectricity. In addition, the lattice structure must be such that it hosts a net electric dipole moment. It is possible to have a lattice structure that is noncentrosymmetric but without any net dipole moment. This is called a non-polar lattice structure (shown in A in Fig. 5.4). A polar lattice structure shown in B, in which case a net electric dipole moment does exist.

Moreover, ferroelectric materials must be insulators in order to host a spontaneous electric field inside the material. Were they metals, the itinerant electrons could simply have moved to rearrange as to cancel the total electric field.

A plot of the polarization versus applied electric field for a ferroelectric material shows a hysteresis loop meaning that the polarization depends on the history of the sample for a given value of the applied electric field. Ferroelectricity disappears above a certain temperature called the critical temperature T_c (or transition temperature). Above the transition, the crystal is said to be in a *paraelectric state*. The dielectric constant can vary strongly with temperature, as shown in Fig. 5.5. In the paraelectric state, the dielectric constant drops rapidly as temperature increases since the material is then in a non-polar phase without any electric dipole moment present and due to the increasing thermal energy of the system which makes it harder to create order. However, in the ferroelectric phase, the dielectric constant can also drop after T_c. The reason for this can also be understood in terms of the thermal energy of the dipole moments created by the atoms in the unit cells. Namely, in order for the dipole moments to orient along an applied electric field, a certain energy barrier must be overcome due to crystal field anisotropies in the unit cell. As temperature is reduced, there is less and less thermal energy available for atoms to overcome this energy barrier and for their dipole moment to point in the favorable direction parallel to the electric field. This "orientational" mode of the dipole moment thus freezes out as temperature is reduced, so that fewer and fewer atoms can contribute to the electric susceptibility ($\chi = \varepsilon_r - 1$).

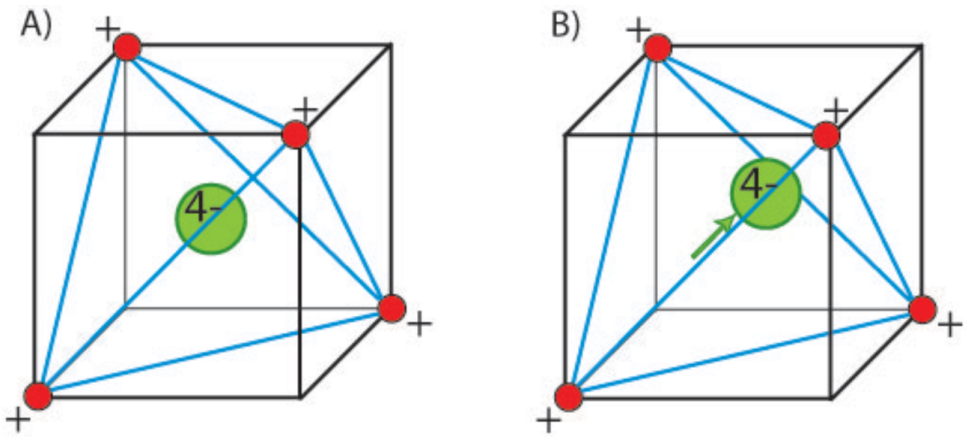

Figure 5.4: **A)** The structure is non-polar, as there is no displacement of the central atom, an no net dipole moment. Note that the unit cell is still noncentrosymmetric: applying $\boldsymbol{r} \to -\boldsymbol{r}$ does not leave the structure invariant. **B)** The central atom is displaced an the structure is polar. Now, there exists an inherent dipole moment in the structure. Figure taken from `https://www.doitpoms.ac.uk/tlplib/ferroelectrics/printall.php`.

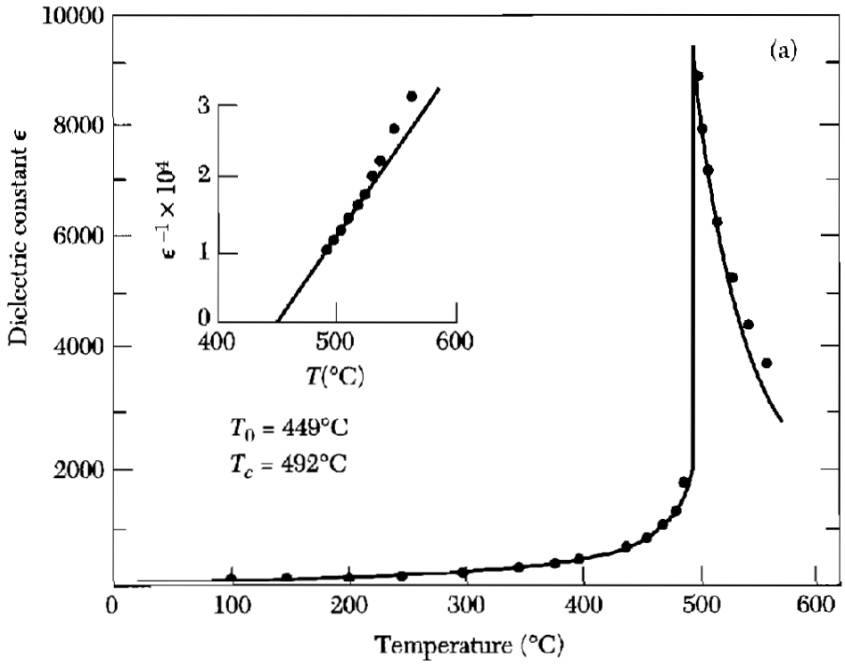

Figure 5.5: The temperature variation of the dielectric constant ε_r of $PbTiO_3$. Figure taken from [7].

Classification of ferroelectric crystals

We list in Fig. 5.6 below some of the crystals commonly considered to be ferroelectric, along with their Curie point T_c at which the crystal changes from the low-temperature polarized state to the

high-temperature unpolarized state. Thermal motion tends to destroy the ferroelectric order, and some ferroelectric crystals in fact have no Curie point because the physically melt before leaving the ferroelectric phase. The figure also includes values of the spontaneous polarization P_s.

	T_c, in K	P_s, in μC cm $^{-2}$, at T K	
KH_2PO_4	123	4.75	[96]
KD_2PO_4	213	4.83	[180]
RbH_2PO_4	147	5.6	[90]
KH_2AsO_4	97	5.0	[78]
GeTe	670	—	—
Tri-glycine sulfate	322	2.8	[29]
Tri-glycine selenate	295	3.2	[283]
$BaTiO_3$	408	26.0	[296]
$KNbO_3$	708	30.0	[523]
$PbTiO_3$	765	>50	[296]
$LiTaO_3$	938	50	
$LiNbO_3$	1480	71	[296]

Figure 5.6: Properties of ferroelectric cyrstals. The last five belong to the class of materials called perovskites. Figure taken from [7].

The ferroelectric transition in crystals may be classified into two main groups: order-disorder or displacive. In an order-disorder ferroelectric, there is a dipole moment in each unit cell, but at high temperatures they are pointing in random directions. Upon lowering the temperature and going through the phase transition, the dipoles order, all pointing in the same direction within a domain. This is similar to how magnetization arises in most ferromagnets, with electron spins playing the role of dipole moments.

The character of the transition can also be defined in terms of the behavior of the lowest frequency ('soft') optical phonon modes. If a soft mode can propagate through the crystal even at the transition from the ferroelectric to the paraelectric state, then the transition is displacive. We will focus below on this type of transition into the ferroelectric state, which typically taktes place in crystal structures closely related to the perovskites. An example of a material in this class is barium titanate $BaTiO_3$, which has an observed saturation polarization P_s at room temperature of approximately 3×10^{-1} C·m^{-2}. The volume of a cell is $(4 \times 10^{-8})^3$ cm^3, so the dipole moment of a cell is

$$p \simeq 2 \times 10^{-29} \text{ C·m} \tag{5.29}$$

5.6 Displacive transitions

The ferroelectric displace transition can be understood in terms of a so-called polarization catastrophe. This catastrophe occurs when some critical condition for the polarization is satisfied, causing the polarization to become very large. When this happens, the local electric field caused by the ionic displacement causes a larger force than the elastic restoring force, thus giving an asymmetrical shift

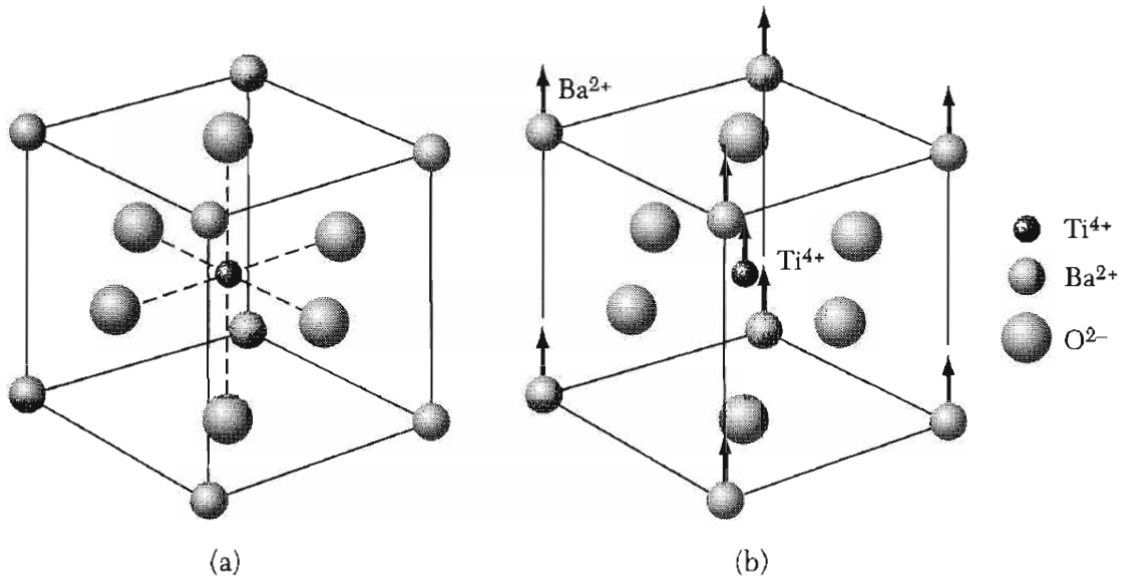

Figure 5.7: **(a)** The crystal structure of barium titanate. The structure is cubic, with Ba^{2+} ions at the cube corners, O^{2-} ions at the face centers, and a Ti^{4+} ion at the body center. **(b)** Below the Curie temperature, the structure is slightly deformed with Ba and Ti ions displaced relative to the O ions, thereby developing a dipole moment. The upper and lower oxygen ions may move downward slightly. Figure taken from [7].

in the position of the ions. Higher order restoring forces will limit the shift to a finite displacement.

The occurrence of ferroelectricity (and antiferroelectricity) in many perovskite-structure crystals suggests that this structure is favorably disposed to a displacive transition. Typically, the O^{2-} ions do not have a cubic surrounding and the local field factors turn out to be unusually large. Let us provide a simple analysis of the catastrophe theory. The local field at all atoms is equal to $\boldsymbol{E} + \boldsymbol{P}/3\varepsilon_0$. We previously found for the dielectric constant that

$$\varepsilon = \frac{1 + 2/3\varepsilon_0 \sum_i n_i \alpha_i}{1 - 1/3\varepsilon_0 \sum_i N_i \alpha_i}. \tag{5.30}$$

Here, α_i is the electronic plus ionic polarizibility of an ion of type i and n_i is the number of ions i per unit volume. We see that the dielectric constant becomes infinite, permitting a finite polarization in zero applied field, when

$$\sum_i n_i \alpha_i = 3\varepsilon_0. \tag{5.31}$$

This is the abovementioned condition for the polarization catastrophe. The condition can be triggered and ferroelectricity correspondingly ensue due to a structural phase transition as temperature is varied, as described in the previous section.

Let us examine how the permittivity changes when we have a small deviation from the condition

triggering the polarization catastrophe. Let us define $s \ll 1$ according to

$$\frac{1}{3\varepsilon_0}\sum_i n_i\alpha_i = 1 - 3s, \tag{5.32}$$

so that $3s$ is the deviation from the critical value, the dielectric constant takes the simple form

$$\varepsilon = 1/s. \tag{5.33}$$

Measurements of the electric permittivity and its temperature-dependence reveals that it varies linearly according to

$$s \simeq (T - T_c)/\xi \tag{5.34}$$

where ξ is a material-dependent constant. Such a variation of s, or correspondingly α_i, can come from the thermal expansion of the lattice as temperature varied: when temperature increases, the number of phonons increase and the amplitude of the lattice vibrations grows, causing the atoms to have a larger average distance between them. Measurements of ε versus temperature show excellent agreement with the above formula for s, see Fig. 5.8.

5.7 Modern theory of polarizibility

We know that an electrically neutral object with some separation between its positive and negative charge constitutes an electric dipole. A great example is a water molecule, which has a little bit of excess negative charge on the oxygen atom, and a little deficit of electrons on the hydrogen atoms.

Once we pick an origin for our coordinate system, we can define the electric dipole moment of some charge distribution as $\boldsymbol{p} = \int \boldsymbol{r}\rho(\boldsymbol{r})d^3\boldsymbol{r}$, , where ρ is the local charge density. Earlier in these notes, we have discussed an induced dipole: the dipole moment that is produced when some object like a molecule has its charges rearrange due to an applied electric field. In that case, $\boldsymbol{p}_{\text{ind}} = \alpha\boldsymbol{E}$, where α is the polarizability. In general α is a tensor, since $\boldsymbol{p}$ and $\boldsymbol{E}$ don't have to point in the same direction.

If we stick a slab of some insulator between metal plates and apply a voltage across the plates to generate an electric field, we know from elementary electromagnetism that the charges inside the insulator slightly redistribute themselves - the material polarizes. If we imagine dividing the material into little chunks, we can define the polarization $\boldsymbol{P}$ as the electric dipole moment per unit volume. For a solid, we can pick some volume and define $\boldsymbol{P} = \boldsymbol{p}/V$, where V is the volume over which the integral is done for calculating $\boldsymbol{p}$. .
We can go farther than that. Say that the insulator is built up out of a bunch of little polarizable objects each with polarization α , where we let each polarizable object see both the externally applied electric field and the electric field from its neighboring dipoles. If we know the total electric field around the polarizable object, we can solve for $\boldsymbol{P}$ and therefore the relative dielectric constant in terms of α . This the essence of the Clausius-Mossotti relation we derived earlier.

For finite and neutral systems, such as molecules or atoms, the above treatment poses no conceptual or practical problems. The definitions $\boldsymbol{p} = \sum_i q_i\boldsymbol{r}_i$ and the generalized $\boldsymbol{d} = \int \rho(\boldsymbol{r})\boldsymbol{r}d\boldsymbol{r}$ are well-defined and can be evaluated to yield results which are consistent with experiments.

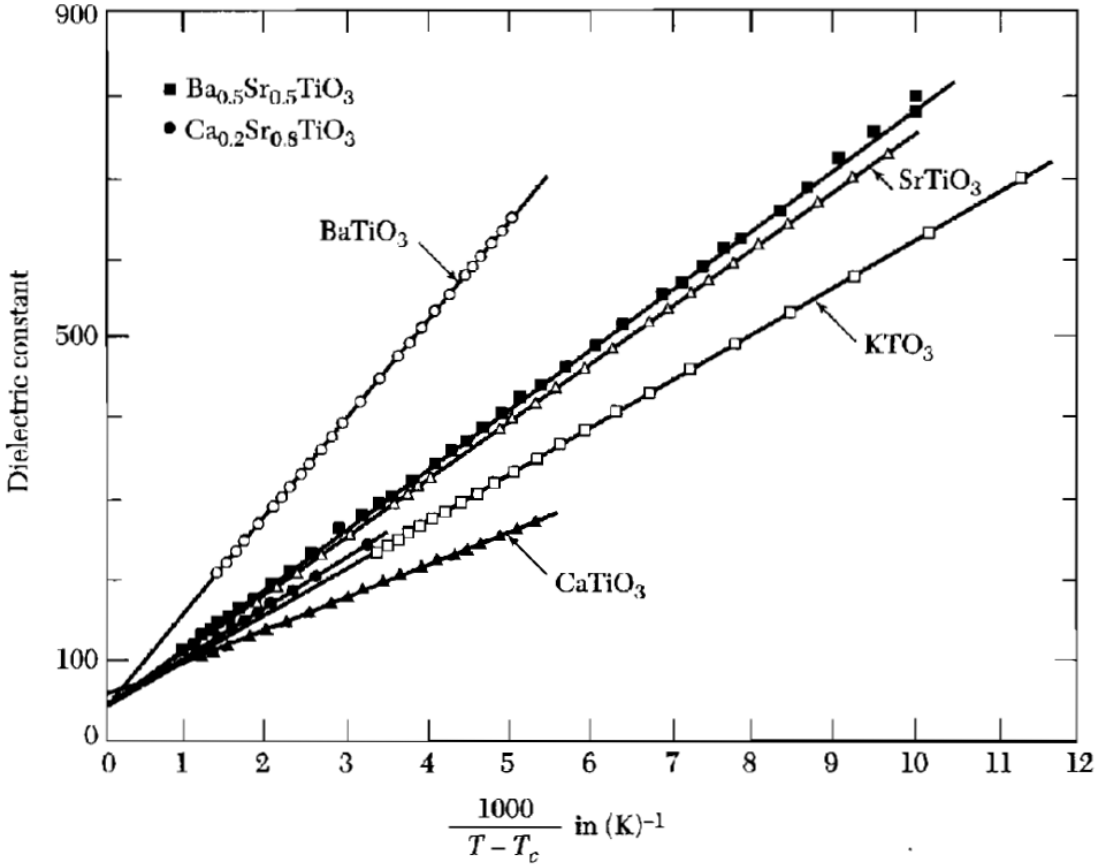

Figure 5.8: Dielectric constant versus $1/(T - T_c)$ in the paraelectric state $T > T_c$ of perovskites, after G. Rupprecht and R. O. Bell. Figure taken from [7].

In crystalline solids, however, it turns out that there is a serious problem that we have not considered yet. Because the charge in a crystal is distributed periodically in space, the definition of $\boldsymbol{P}$ given above is ambiguous because there are many ways to define the 'unit cell' over which the integral is performed. This is a big deal, as we will now show.

The usual way to define intrinsic quantities in macroscopic systems is to introduce the property per unit volume or mass. For example the magnetization is the magnetic moment per unit volume, and the bulk analog to the electric dipole moment, the electric polarization, should be represented by the electric dipole moment per unit volume. The relevant quantity is then evaluated within a small repeat unit – the unit cell – of the solid, and normalized with the volume of the chosen unit cell. The problem with this simple method in the case of electric polarization can be understood in the simple one-dimensional cartoon of Fig. 1: without performing any calculations, we can see that the two equally valid unit cells shown with dashed lines have completely opposite orientations of the polarization!

In a periodic system, there thus appears to be a fundamental problem with defining uniquely the

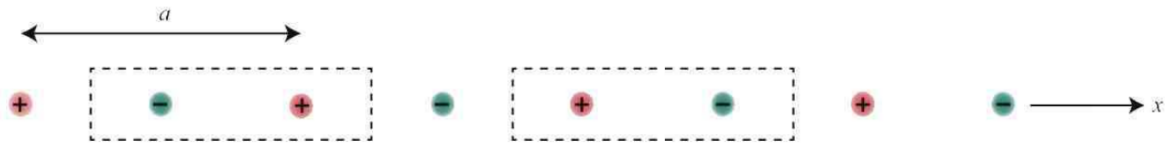

Figure 5.9: One-dimensional chain of alternating anions and cations, spaced a distance $a/2$ apart, where a is the lattice constant. The dashed lines indicate two representative unit cells which are used in the text for calculation of the polarization. Figure taken from [9].

polarization, meaning the dipole moment per volume. This problem can be resolved by looking at polarization differences rather than absolute polarizations, which is typically what is measured in experiments anyway: for more details on this, see one of the exercises. In the modern theory of polarizibility, one therefore instead operates with polarization differences which are expressed using quantum mechanical Berry-phases obtained from the band structure of the material. For instance, to compute the spontaneous polarization of a ferroelectric material, one evaluates the difference in polarization between the polar phase of the material and its centrosymmetric phase (final, polarized state and initial, unpolarized state). This is beyond the scope of this course. It is interesting to note that this modern theory of polarizbility is rather new: it was developed in the 1990s.

We end this section by noting that the modern theory of polarizibility is a necessity when working with truly translationally invariant systems, meaning periodic boundary conditions for a lattice. As we stated in the beginning of this section, there is no ambiguity in calculating the polarization of a finite-sized object. If we can compute the dipole moment of a neutral atom, then we can also do it for two atoms. If we can do it for two atoms, we can in principle do it for an entire material since it is built up by many atoms. In practice, this may however not be so easily done, and could require a lot of computational power to truly model a finite-size macroscopic material. It is analytically easier to compute a quantity for a translationally invariant system, since we can then work in $\boldsymbol{k}$-space, and this is where the modern theory of polarizability shines.

5.8 Landau theory

We start by developing the general framework for Landau theory and thereafter apply it specifically to ferroelectrics.

5.8.1 General theory

Phase transitions are a core topic in condensed matter physics. In a large class of phase transitions, the system undergoes a symmetry change. This is the domain of Landau theory. There exists another type of phase transition called topological phase transitions which cannot be described within the Landau theory framework, for instance in the two-dimensional XY spin model, but this will not be covered here.

Here, we will learn about Landau theory. Let us commence with a few generic facts about phase transitions.

- A phase transition can be driven by many parameters - temperature, pressure, magnetic or electric field to mention a few. If the driving parameter is temperature, the high-temperature phase is usually more disordered, meaning it has a higher symmetry than the low-temperature phase.

- Phase transitions signal a change in the entropy of the system. This change can be discontinuous or continuous.

    1. If it is discontinuous, the phase transition is accompanied by a release of heat (called latent heat) and all the other thermodynamics quantities (internal energy, entropy, volume, and so on) are discontinuous as well. Such a phase transition is called a first-order transition.

    2. If it is continuous, the thermodynamic quantities are continuous across the transition and there is no release of heat. However, the first derivatives of thermodynamic quantities are discontinuous. In particular,a the specific heat has a pronounced anomaly which we will look at below.

The order parameter

To characterize the change in symmetry as one moves from one phase to another in a system. To do so, an immensely useful concept is the order parameter. This measures the degree of departure from the high symmetry, disordered phase. The order parameter should have the property that it is finite in the ordered (low-symmetry) phase and zero in the disordered (high-symmetry) phase.

There are many different order parameters, depending on the system at hand and the type of order we are dealing with. The order parameter can be a vector, namely the magnetization, as in the case of ferromagnets. It can also be a scalar in the case of superconductivity (see later chapter in these notes). It all depends on what type of ordering we are facing in the material: order parameters can also be exotic objects like pseudoscalars which transform differently than scalars under certain operations, and they may even be complex quantities.

The Neumann principle

When a phase transition occurs, the symmetry of the ordered phase is usually a subgroup of that of the disordered phase. Symmetry may be lowered because of loss of translational and/or rotational invariance. Changes in point group symmetry (i.e., loss of rotation/mirror symmetries) are extremely important, since they allow new macroscopic physical phenomena. This is expressed in the famous Neumann Principle (from Franz Ernst Neumann 1798-1895):

The symmetry elements of any macroscopic physical property of a crystal must include the symmetry elements of the point group of the crystal.

When a phase transition results in the development of a new macroscopic property that couples to an external field, it is said to be a ferroic transition. All ferroic transitions reduce the point-group symmetry. The two most famous examples of ferroic transitions are treated in this course: ferroelectricity and ferromagnetism.

Phenomenological considerations

One of the most significant contributions of Lev Davidovich Landau (1908-1968) — one of the great physicists of the 20th century — has been the theory of phase transitions bearing his name. Landau theory is of central importance in many fields of condensed matter physics, including structural phase transitions, magnetism and superconductivity (the latter through a modification of the original theory known as Ginzburg-Landau theory.

The essential feature of Landau theory is that it is a phenomenological theory. This means that, unlike a microscopic theory, it is not concerned with the details of the interactions at the atomic level that ultimately should govern the behaviour of any system. For a structural phase transitions, microscopic interactions would be ionic and covalent bonding, Coulomb interactions, Van der Waals interactions etc.; for a magnetic system, exchange and dipole interactions; for a superconductor, pairing interactions, and so on. Instead, Landau theory is chiefly concerned with symmetry — in fact, it only applies to phase transitions entailing a change in symmetry. One of the upshots of this is that systems with similar symmetries — even very different systems, which we might expect to have very different microscopic theories — would look very similar within Landau theory. This connection between very distant branches of physics might be thought of as the origin of the idea of universality, which was to play a fundamental role in further developments of Landau theory. Despite not saying anything about the microscopic origin of the phase transition, Landau theory is therefore immensely useful across a broad range of material systems.

The Landau free energy

The central idea of Landau theory is the construction of a quantity, known as Landau free energy or F, which describes the energetics of the system in the vicinity of a phase transition. The free energy F, which can be usually thought of as an approximation to the Helmholtz or Gibbs free energy per unit volume, is a real quantity. It depends on temperature, pressure and any other relevant external parameter (such as electric or magnetic field, stress, and so on). The Helmholtz free energy is used when temperature is fixed, whereas the Gibbs free energy is used when pressure is fixed. The latter is the relevant free energy to use in the presence of external fields that apply a force (i.e. pressure) on the system. Crucially, the Landau free energy also depends on the order parameters of the system.

The free energy should obey two conditions in Landau's prescription:

- Be analytic in the order parameter(s) and its gradients.
- Obey the symmetry of the Hamiltonian.

For a given set of external parameters, the stable state of the system is the one for which the Landau free energy is minimal as a function of all internal degrees of freedom.

As we just mentioned, F depends on the internal variables of the system through the order parameter. In the Landau construction, one thus implicitly assumes the existence of a high-symmetry phase somewhere in the phase diagram, most likely at high temperatures. In this state, all the order parameters are zero. One can therefore naturally decompose F as:

$$F = F_0 + \Delta F(\eta^i). \tag{5.35}$$

Here, F_0 does not depend on the order parameter(s) and has no influence on the phase transition. The part ΔF of the free energy depends on the order parameter(s) and is small in the vicinity of the

phase transition, since it is zero once the transition is crossed into the high-symmetry disordered phase. We consider here for simplicity only real order parameters $\eta^i \in \mathbb{R}$.

The following statement is the point of departure for the Landau analysis: for any value of the order parameters, ΔF is invariant by any element g of the high-symmetry group G_0. In addition, ΔF may possess the additional symmetries of free space (such as parity and time-reversal) except if external fields are present which break any of these symmetries.

Since ΔF is assumed small near the phase transition, the next natural step is to perform a Taylor expansion of $\Delta F(\eta^i)$ in powers of η^i. Consider a simple real, one-dimensional order parameter. The expansion is then

$$\Delta F = -\eta H + \frac{a}{2}\eta^2 + \frac{c}{3}\eta^3 + \frac{b}{4}\eta^4 + \mathcal{O}(\eta^5). \tag{5.36}$$

The odd-power terms are restricted by symmetry: they are not present if $\eta \to -\eta$ is an allowed symmetry in the system. For instance, if the energy of a magnetic system is equal regardless of whether a magnetization m points up or down, then $\pm m$ are energetically equivalent solutions and the free energy must be invariant under $m \to -m$.

In any case, finding the stable states of the system entails minimizing the free energy ΔF as a function of the order parameters. We now continue with this simple free energy example and simply remark that order parameters will in general be more complicated, for instance having vectorial nature or being complex fields.

In Eq. (2), we assume that the order parameter η has been normalized in such a way that the coefficient of the coupling term $-\eta H$ which describes how η couples to the generalized external field H is -1. We now proceed to discuss the physical meaning of each term in the above expansion.

The linear term in η

In the Landau free energy, there cannot exist a linear term in the order parameters that does not couple to external fields. The reason is that if such a term was present, then the conditions providing the minimum free energy configuration [$\partial F/\partial \eta = 0$ and $\partial^2 \eta/\partial \eta^2 > 0$] would yield a solution $\eta \neq 0$ at any value of the expansion coefficients, i.e. any temperature. Then η would not be a suitable order parameter since it is supposed to be zero above the critical temperature. Another reason is purely mathematical: if a linear term exists, it can simply be removed by shifting η by a constant value, $\eta \to \eta + C = \tilde{\eta}$, and the new expansion in terms $\tilde{\eta}$ would not have a linear term.

The coupling term to the external field $-\eta H$ can be present, but not always (for ferroic transitions, it is by definition, as mentioned earlier). Introducing a generalized polarization as

$$P = -\frac{\partial F}{\partial H}, \tag{5.37}$$

it follows that for ferroic transitions the order parameter η is the generalized polarization (for instance the magnetization of a ferromagnet).

The quadratic term in η

The second term is generally allowed in the expansion, and its sign has an important consequence. Namely, a phase transition occurs when the coefficient of the quadratic term in the order parameter

expansion (a above) changes sign from positive to negative. If the parameter driving the phase transition is temperature, the coefficient of the quadratic term is usually written $a = a'(T - T_c)$ where T_c is the transition/critical temperature.

The cubic term in η

Landau free energies would not be well-defined if the Taylor expansion stopped at an odd-order term, because the free energy would in that case be unbound from below. Nevertheless, cubic terms are very important in the Landau expansion, because if present they force the transition to be first-order (see one of the exercises). However, symmetry places restrictions on whether cubic terms are allowed. For instance, if the transformation $\eta \to -\eta$ is a symmetry of the system, cubic terms are disallowed. Thus, in order for the phase transition to be continuous (second order), the cubic term in the Landau expansion must be forbidden by symmetry.

The quartic term and higher-order terms in η

The quartic term is essential in producing a well-behaved free energy that is bound in energy from below, the requirement being that $F \to \infty$ when $|\eta| \to \infty$. In our simple example aboave, this is satisfied when $b > 0$. If higher-order terms are present, such as 6th order terms which are generally also present and allowed by symmetry), the quartic term can be negative. In such a case, it is possible to chose the value of the expansion coefficients so that a Landau free energy with a negative quartic term can produce a first-order phase transition. In fact, a change in sign of the quartic term is probably the easiest mechanism to produce a change in character from second to first order of the phase transition.

Analysis of a free energy

Let us now analyze a simple type of Landau free energy, namely our above expansion but without the odd-order terms:

$$\Delta F = -\eta H + \frac{a}{2}\eta^2 + \frac{b}{4}\eta^4. \tag{5.38}$$

We now know that this free energy describes a continuous phase transition. Our purpose is now to extract a few relevant thermodynamic parameters both above and below the phase transition.

Let us start with the order parameter, which is the generalized polarization due to the existence of the coupling to an external field. We minimize ΔF with respect to η:

$$-H + a'(T - T_c)\eta + b\eta^3 = 0. \tag{5.39}$$

In zero field, any non-trivial solution for η will be a sopntaneous polarization occuring in the material. The above equation has two simple solutions for $H = 0$. One is $\eta = 0$, which is allowed for any temperature. The other one is

$$\eta = \pm\sqrt{\frac{a'}{b}}(T_c - T)^1 2, \tag{5.40}$$

and this solution is only allowed below T_c since η is assumed real. Analyzing the second derivative, it follows straightforwardly that $\eta = 0$ is the global minimum for $T > T_c$, whereas for $T < T_c$ it is a local minimum. The situation is shown in Fig. 5.10(a). Had we included a cubic term in the expansion, the transition would have changed to first order. This is seen from the free energy curve

in Fig. 5.10(b) by noting that the value of η at which the global maximum of the curve occurs changes abruptly from finite to zero at a certain temperature.

Had we included $H \neq 0$, the a solution $\eta \neq 0$ exists both above and below T_c. In other words, the external field breaks the symmetry of the disordered, high-symmetry phase above T_c, and there no longer exists a true phase transition.

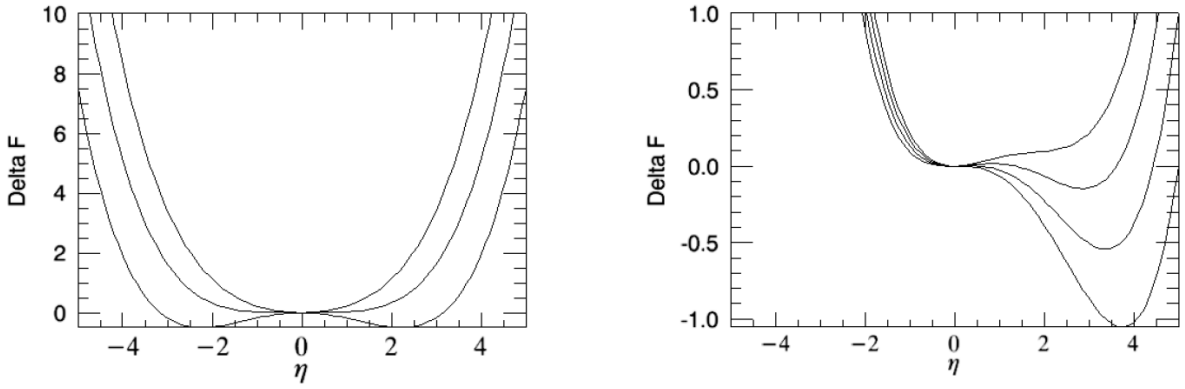

Figure 5.10: Two examples of the temperature dependence of Landau free energy. The different lines correspond to different temperatures T. **Left:** Simple form with quadratic and quartic terms produces a 2nd order phase transition. **Right:** Adding a cubic term produces a 1st order phase transition. Figure taken from [12].

The generalized susceptibility

The generalised susceptibility (magnetic susceptibility for a ferromagnetic transition, dielectric constant for a ferroelectric transition, etc) can also be calculated from the Landau free energy as

$$\chi = \frac{\partial \eta}{\partial H} \tag{5.41}$$

meaning that the susceptibility expresses how much the order parameter changes when changing the external field H. Since we generally have $F = -\eta H + \Delta F'(\eta)$, where $\Delta F'$ contains all terms depending on η except its coupling to the external field, we can write

$$\frac{\partial F}{\partial \eta} = -H + \frac{\partial \Delta F'}{\partial \eta} = 0. \tag{5.42}$$

This is true for all values of the field since the value of the order parameter which is physically realized is the one that minimizes the free energy, as we have seen above. Taking now the *total* derivative of Eq. (5.42) with respect to H, we obtain

$$-1 + \frac{\partial^2 \Delta F'}{\partial \eta^2} \frac{\partial \eta}{\partial H} = 0. \tag{5.43}$$

This gives the formula for the susceptibility expressed in terms of the Landau free energy:

$$\chi^{-1} = \frac{\partial^2 \Delta F'}{\partial \eta^2}. \tag{5.44}$$

With our simple free energy considered above, we obtain

$$\chi^{-1} = a'(T - T_c) + 3\eta^2 b. \tag{5.45}$$

At zero external field, one therefore finds (using the value of η that minimizes the free energy) different temperature dependence above and below the critical temperature

$$\chi^{-1} = \begin{cases} a'(T - T_c) \text{ for } T > T_c \\ 2a'(T_c - T) \text{ for } T < T_c \end{cases} . \tag{5.46}$$

Note how the zero-field susceptibility diverges at the critical temperature. The negative power-law behavior of the generalized polarization (order parameter) and the divergence of the susceptibility near the transition are essentially universal properties for continuous (2nd order) phase transitions. However, the critical exponents (often referred to as β for the generalized polarization and γ/γ' for the susceptibility above/below T_c) will in practice be quite different from the Landau predictions above of $\beta = 1/2$ and $\gamma = \gamma' = 1$. More accurate predictions for these exponents which match experimental results better can be recovered by using a more complex theory that takes into account the effect of fluctuations of the order parameter.

5.8.2 Landau theory of the ferroelectric phase transition

From the above, we know that a ferroelectric with a first order phase transition between the ferro- and paraelectric state has a discontinuous change of the polarization at the transition temperature. For a second order phase transition, the polarization and degree of order goes to zero in a continuous manner, so that the polarization can become arbitrarily small as one approaches T_c. Recall that in a second order transition, there is no latent heat: the order parameter (the polarization in the case of ferroelectrics and the magnetization in the case of ferromagnets) is not discontinuous at the transition temperature T_c. In a first order transition, there is a latent heat as the order parameter changes discontinuously at the transition temperature.

Following the procedure explained in the general case for a Landau theory, we expand the free energy in the polarization P as the order parameter. Generally, we should be able to expand it as

$$F(P, T, E) = -EP + g_0 + \frac{1}{2} g_s P^2 + \frac{1}{4} g_4 P^4 + \dots \tag{5.47}$$

where all coefficients g_n may depend on temperature. Here, P is the polarization, E the external electric field, and T temperature. Note that the free energy does not contain any terms which are odd powers of P in the absence of an external field. This is the case when the unpolarized crystal has a center of inversion symmetry: if the crystal is noncentrosymmetric, odd terms in P are allowed by symmetry. This is because the free energy must respect the symmetries obeyed by the physical system itself. If the crystal in the paraelectric state has a center of inversion, then the spontaneous polarization appearing below T_c is a spontaneous symmetry breaking: this means that both P and $-P$ are energetically equivalent solutions. Instead, if the crystal in the paralectric state does not have a center of inversion, the system has already explicitly broken inversion symmetry: this means that P and $-P$ in the polarized state are not energetically equivalent.

The value of P that is realized in the actual material in thermal equilibrium at a fixed temperature is given by the minimum of F. In effect, one differentiates the Helmholtz free energy (isothermal)

to obtain the equilibrium polarization value:

$$\frac{\partial F}{\partial P} = 0 = -E + g_2 P + g_4 P^3 + \dots \tag{5.48}$$

and solves for P. We disregard here any depolarization field effects (covered previously in this chapter), and thus effectively assume that the material under consideration is a long rod with the external applied field E parallell to the long axis.

To obtain a ferroelectric state, the coefficient g_2 must pass through zero at some temperature T_0:

$$g_2 = \gamma(T - T_0) \tag{5.49}$$

where γ is taken as a positive constant and T_0 may be equal to or different from the transition temperature T_c. A small positive g_2 means that the lattice is "soft" and close to an instability. A negative value of g_2 means that the unpolarized lattice ($P = 0$) is unstable, so that the ground state of the system must have $P \neq 0$. The temperature-dependence of g_n comes about due to thermal expansion and other effects from anharmonic lattice interactions (in effect, going beyond quadratic displacements of the lattice from its equilibrium position, resulting in anharmonic vibrations and phonon behavior).

5.8.3 Second-order transition

If $g_4 > 0$ in our expansion, no qualitatively new physics comes about due to g_6, so let us truncate the expansion after P^4. The polarization for zero applied electric field is found from $\partial F/\partial P = 0$:

$$\gamma(T - T_0)P + g_4 P^3 = 0. \tag{5.50}$$

This has two solutions. One is $P = 0$, which corresponds to an unstable point. The other is $P^2 = (\gamma/g_4)(T_0 - T)$, which is a minimum of the free energy. For $T \geq T_0$, the only real root of this expression is $P = 0$ since γ and g_4 are both positive. Thus, T_0 is the Curie temperature in our model. For $T < T_0$, the minimum of the Landau free energy in zero applied electric field occurs at

$$|P| = (\gamma/g_4)^{1/2}(T_0 - T)^{1/2}. \tag{5.51}$$

This is a monotonically decreasing function as T approaches T_0 from below. The phase transition is a second order transition since the polarization goes continuously to zero as T_0 is approached. This type of ferroelectric transition takes place in *e.g.* $LiTaO_3$.

5.8.4 First order transition

The transition is first order if g_4 is negative. However, F can then grow indefinitely in the negative direction unless we keep g_6 and keep it positive in order to put a lower bound on F. The equilibrium condition at $E = 0$ is then given by

$$\gamma(T - T_0)P - |g_4|P^3 + g_6 P^5 = 0. \tag{5.52}$$

As before, there are two solutions. Either $P = 0$ or

$$\gamma(T - T_0) - |g_4|P^2 + g_6 P^4 = 0. \tag{5.53}$$

Comparing the values of the free energies in the paraelectric ($P = 0$) and ferroelectric phase ($P \neq 0$) at the transition temperature $T = T_c$, where the minimum in the ferroelectric phase becomes lower than in the paraelectric phase, one finds that they are equal. The free energy versus polarization as well as the characteristic variation of P with temperature for this first-order phase transition is shown in Fig. 5.11. Such a transition occurs in $BaTiO_3$.

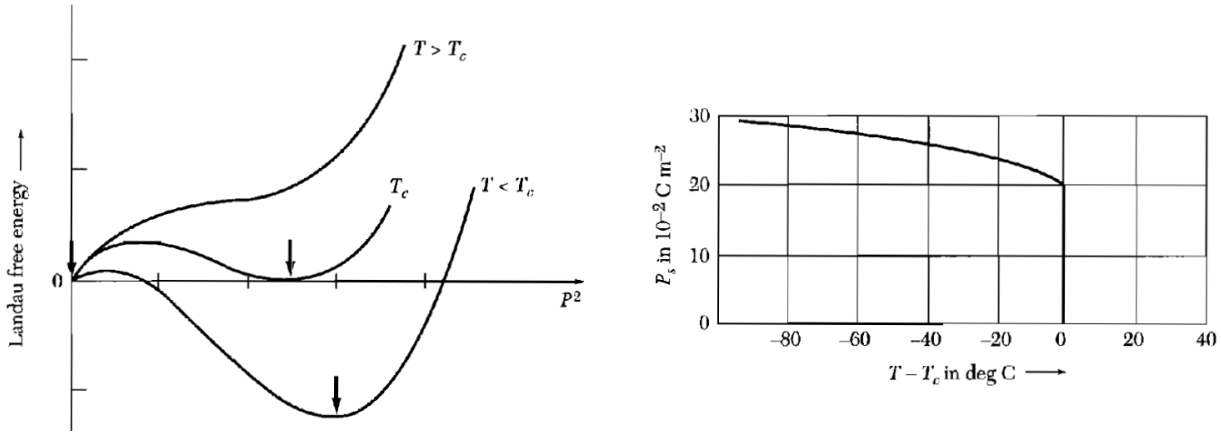

Figure 5.11: **Left:** Landau free energy function versus square of the polarization in a first-order phase transition at representative temperatures. At T_c, the Landau function has equal minima at $P = 0$ and at a finite P, as shown. For $T < T_c$, the absolute minimum is at larger values of P. As T increases above T_c, there is a discontinuous change in the position of the absolute minimum. The arrows mark the minima. **Right:** Calculated values of the spontaneous polarization as a function of temperature, with parameters as for barium titanate. Figure taken from [7].

5.9 Antiferroelectricity and ferroelectric domains

A ferroelectric displacement is not the only type of instability that may develop in a dielectric crystal. Other deformations occur which, even if they do not give a spontaneous polarization, may be accompanied by changes in the dielectric constant. One type of deformation is called antiferroelectric and has neighboring lines of ions displaced in opposite senses. The perovskite structure appears to be susceptible to many types of deformation, often with little difference in energy between them. The phase diagrams of mixed perovskite systems, such as the $PbZr0_3$–$PbTiO_3$ system, show transitions between para-, ferro-, and antiferroelectric states.

Ferroelectric materials generally also feature so-called domains. Consider a ferroelectric crystal (such as barium titanate in the tetragonal phase) in which the spontaneous polarization may be either up or down the c-axis of the crystal. A ferroelectric crystal generally consists or regions called domains within each of which the polarization is in the same direction, but in adjacent domains the polarization is in different directions. In Fig. 5.12 the polarization is in opposite directions. The net polarization depends on the difference in the volumes of the upward- and downward-directed domains. The crystal as a whole will appear to be unpolarized, as measured by the charge on electrodes covering the ends, when the volumes of domains in opposite senses are equal. The total dipole moment of the crystal may be changed by the movement of the walls between domains or by the nucleation of new domains. The domain boundaries change size and shape when the intensity of an applied electric field is altered.

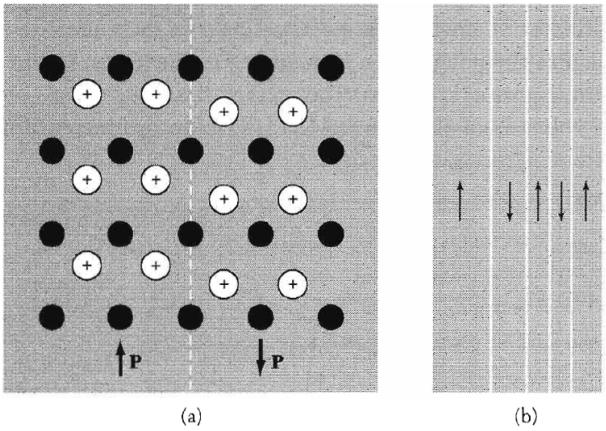

Figure 5.12: **(a)** Schematic drawing of atomic displacements on either side of a boundary between domains polarized in opposite directions in a ferroelectric cystal. **(b)** View of a domain structure, showing 180° boundaries between domains polarized in opposite direction. Figure taken from [7].

5.10 Pyro- and piezoelectricity

Pyro- and piezoelectric materials are closely related to ferroelectric materials. They also require a noncentrosymmetric lattice structure, and all ferroelectric materials are in fact pyro- and piezoelectric. The reverse is not true, as seen from Fig. 5.13.

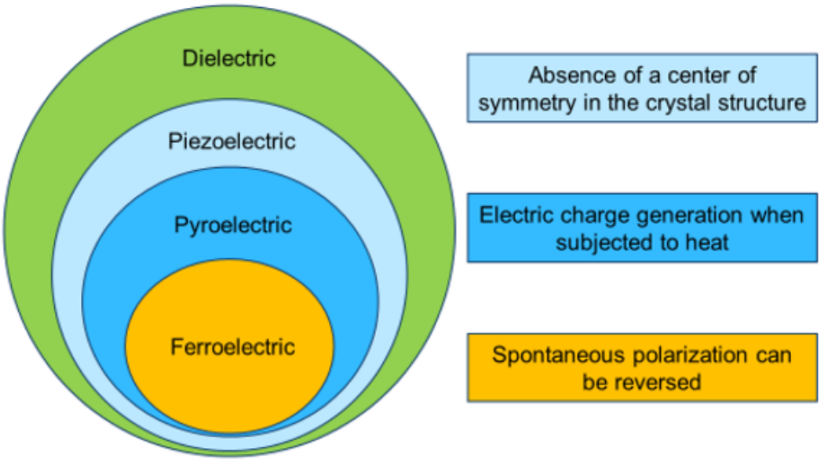

Figure 5.13: Relation between dielectric, piezoelectric, pyroelectric, and ferroelectric material. Taken from kemet.com

Pyroelectricity is a property of certain crystals which are naturally electrically polarized. It can be described as the ability of certain materials to generate a temporary voltage when they are heated or cooled. The change in temperature modifies the positions of the atoms slightly within the crystal structure, so that the polarization of the material changes. This polarization change gives rise to

a voltage across the crystal. If the temperature stays constant at its new value, the pyroelectric voltage gradually disappears due to leakage current. The leakage can be due to electrons moving through the crystal, ions moving through the air, or current leaking through a voltmeter attached across the crystal.

All crystals in a ferroelectric state are also piezoelectric. Applying a stress Z to the crystal changes the electric polarization, as shown in Fig. 5.14. This is because the stress alters the distance between the atoms, which in turn changes the electric dipole moment in the ferroelectric state. The relation between these quantities may be written phenomenologically

$$P = Zd + \varepsilon_0 E\chi_e, \; e = Zs + Ed. \tag{5.54}$$

Here, P is the polarization, Z quantitifies the applied stress, d is the piezoelectric strain constant, E is the electric field, χ_e the dielectric susceptibility, e the elastic strain, and s the elastic compliance constant. This relation displays the evolution of the polarization by an applied stress and the development of elastic strain by an applied electric field. Note that a crystal may be piezoelectric without being ferroelectric. This is the case for *e.g.* quartz, whereas barium titanate is both. Piezoelectric materials are used for instance in medicine to monitor blood pressure and respiration.

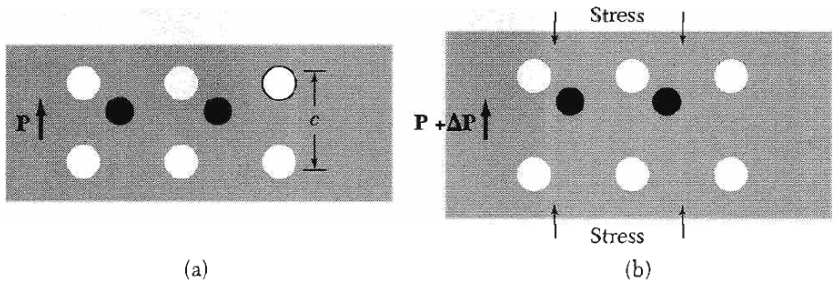

Figure 5.14: **(a)** Unstressed ferroelectric crystal and **(b)** stressed ferroelectric crystal. The stress changes the polarization by ΔP, the induced piezoelectric polarization. Figure taken from [7].

Chapter 6

Superconductivity

6.1 Fundamental properties

Superconductivity is characterized by a vanishing static electrical resistivity and an expulsion of the magnetic field from the interior of a sample. Superconductivity is somewhat related to the phenomena of superfluidity (in He-3 and He-4) and Bose-Einstein condensation (in weakly interacting boson systems), but we will not go into details on this phenomena here.

After H. Kamerlingh Onnes had managed to liquify Helium, it became possible to reach temperatures low enough to achieve superconductivity in some chemical elements. In 1911, he found that the static (dc) resistivity of mercury abruptly fell to zero at a critical temperature T_c of about 4.1 K. In a normal metal, the resistivity decreases with decreasing temperature but saturates at a finite value for $T \to 0$. The most stringent bounds on the resistivity can be obtained not from direct measurement but from the decay of persistent currents, or rather from the lack thereof. A current set up (by induction) in a superconducting ring is found to persist without measurable decay after the electromotive force driving the current has been switched off.

Assuming exponential decay, $I(t) = I(0)\mathrm{e}^{-t/\tau}$, a lower bound on the decay time τ is found. From this, an upper bound of $\rho \leq 10^{-26}$ Ωm has been extracted for the resistivity by File and Mills (1963). For comparison, the resistivity of copper at room temperature is $\rho \simeq 1.7 \times 10^{-8}$ Ωm.

The second essential observation was that superconductors not only prevent a magnetic field from entering, but actively expel the magnetic field from their interior. This was observed by W. Meissner and R. Ochsenfeld in 1933 and is now called the Meissner or Meissner-Ochsenfeld effect. Thus, superconductor have a negative susceptibility $\chi < 0$ and in fact zero magnetic permeability: they are perfect diamagnets. It costs energy to make the magnetic field nonuniform although the externally applied field is uniform. It is plausible that at some externally applied magnetic field $B_c(T)$ this cost will be so high that there is no advantage in forming a superconducting state.

6.1.1 Conventional vs unconventional superconductivity

To specify which superconductors discovered after Hg are conventional, we need a definition of what conventional means in this context. There are at least two inequivalent but often coinciding definitions: Conventional superconductors

- Show a superconducting state of trivial symmetry: the superconductor does not break any

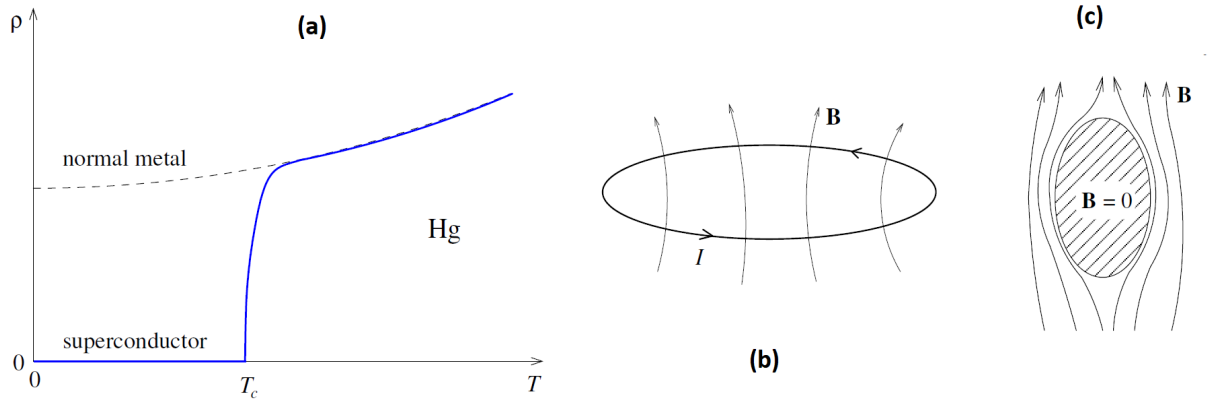

Figure 6.1: **(a)** Resitivity vs. temperature for Hg showing an abrupt drop to zero resistivity at the critical temperature T_c when the material enters a superconducting state. **(b)** An electric current can be induced in a ring by magnetic induction, forming a test for the decay of superconducting currents. **(c)** The Meissner effect consists of superconductors shielding external magnetic fields from entering their interior. This occurs via shielding supercurrents circulating at the surface of the superconductor. Figure taken from [2].

additional symmetry beyond a global U(1) phase symmetry, which we will discuss in more detail later.

- Result from an attractive interaction between electrons mediated by phonons. We have already alluded to how this can occur in a previous chapter in these notes.

Conventional superconductivity was observed in quite a lot of elements at low temperatures. The record critical temperature for elements are $T_c = 9.3$ K for Nb under ambient pressure and $T_c = 29$ K for Ca under high pressure. Superconductivity is in fact rather common in the periodic table, about 53 pure elements show it under some conditions. Many alloys and intermetallic compounds were also found to show conventional superconductivity according to the above criteria. Of these, for a long time Nb3Ge had the highest known T_c of 23.2 K. But it is now thought that MgB_2 ($T_c = 39$ K, discovered to be superconducting in 2001) and a few related compounds are also conventional superconductors in the above sense. They nevertheless show some interesting properties. The rather high $T_c = 39$ K of MgB_2 is interesting since it is on the order of the maximum $T_{\max} \simeq 30$ K that was originally expected for phonon-driven superconductivity. To increase T_c further, the interaction between electrons and phonons would have to be stronger. However, it was thought that stronger interactions would make the material unstable towards a charge density wave. MgB_2 would thus be an "optimal" conventional superconductor.

Superconductivity with rather high T_c has also been found in fullerites, i.e., compounds containing fullerene anions. The record Tc in this class is at present $T_c = 38$ K for body-centered cubic Cs_3C_{60} under pressure. Superconductivity in fullerites was originally thought to be driven by phonons with strong molecular vibration character but there is recent evidence that it might be unconventional (not phonon driven). The received wisdom was challenged by the discovery of superconductivity in hydrogen sulfide at up to 203 K under a high pressure of above 150 GPa (corresponding to about 1.5 million standard atmospheres), see Drozdov et al., Nature **525**, 73 (2015). The responsible com-

pound is thought to be H_3S forming a body-centered cubic lattice. Amazingly, superconductivity in this system appears to be phonon driven and in this sense conventional. The presence of the light hydrogen atoms is crucial for this. That H atoms could favor superconductivity at high temperatures was already proposed by N. W. Ashcroft in 1968. There is an ongoing search for high-Tc superconductivity in other hydrogen compounds. Recently, this effort has led to the discovery of superconductivity with $T_c \simeq 250$ K in LaH_{10} at high pressure, see Drozkov et al., Nature **569**, 528 (2019). The material is a clathrate, in which the metal atom is surrounded by a cage of hydrogen atoms.

By the late 1970s, superconductivity seemed to be a more or less closed subject. It was well understood based on the BCS theory and extensions thereof that dealt with strong interactions. It only occured at temperatures up to 23.2K (Nb_3Ge) and thus did not promise widespread technological application. It was restricted to non-magnetic metallic elements and simple compounds. This situation started to change dramatically in 1979. Since then, superconductivity has been observed in various materials classes that are very different from each other and from the typical low-T_c superconductors known previously. In many cases, the superconductivity is unconventional and often T_c is rather high.

The most dramatic development occurred in 1986, when J. G. Bednorz and K. A. Müller observed superconductivity in $La_{2-x}Ba_xCuO_4$ (the layered perovskite cuprate La_2CuO_4 with some Ba substituted for La) with Tc on the order of 35 K. In the following years, many other superconductors based on the same type of nearly flat CuO_2 planes sketched below were discovered. The record transition temperatures for cuprates are $T_c = 138$ K for $Hg_{0.8}Tl_{0.2}Ba_2Ca_2Cu_3O_{8+\delta}$ at ambient pressure and $T_c = 164$ K for $HgBa_2Ca_2Cu_3O_{8+\delta}$ under high pressure. Many experimental probes show that the cuprates are unconventional superconductors. High-T_c superconductivity in the cuprates was historically important since many advanced methods of many-body theory have been developed motivated by the desire to understand this phenomenon. Superconductivity is believed to be unconventional in the cuprates, since the order parameter describing the superconducting phase breaks additional symmetries than in conventional superconductor and because it is heavily debated if electron-phonon coupling is the glue binding the electrons together.

6.2 London equation

In 1935, F. and H. London proposed a phenomenological theory for the electrodynamic properties of superconductors. It is based on a two-fluid picture: the basic assumption was that there exists electrons behaving as a normal fluid of concentration n_n and a superfluid of concentration n_s, where $n_n + n_s = n = N/V$. Such a picture seemed quite plausible given the properties observed in superconducting materials, although nobody yet understood how the fermionic electrons could microscopically form a superfluid. The normal fluid is postulated to behave normally, i.e., to carry an ohmic current

$$\boldsymbol{j}_n = \sigma_n \boldsymbol{E} \tag{6.1}$$

governed by the Drude conductivity

$$\sigma_n = \frac{e^2 n_n \tau}{m}. \tag{6.2}$$

[1] The superfluid is assumed to be insensitive to scattering. This leads to free acceleration of the charges. The Londons have assumed that n_n and n_s are both uniform (constant in space) and stationary (constant in time). These are serious restrictions of London theory, which will be overcome by Ginzburg-Landau theory that will developed later. With the supercurrent

$$\boldsymbol{j}_s = -en_s\boldsymbol{v}_s \tag{6.6}$$

and Newton's equation of motion

$$\frac{d}{dt}\boldsymbol{v}_s = \boldsymbol{F}/m = -\frac{e\boldsymbol{E}}{m} \tag{6.7}$$

it follows that

$$\frac{\partial \boldsymbol{j}_s}{\partial t} = \frac{e^2 n_s}{m}\boldsymbol{E}. \tag{6.8}$$

This is called *the first London equation.*

Next, we note that the curl of the first London equation is

$$\frac{\partial}{\partial t}\nabla\cdot\boldsymbol{j}_s = \frac{e^2 n_s}{m}\nabla\times\boldsymbol{E} = -\frac{e^2 n_s}{m}\frac{\partial \boldsymbol{B}}{\partial t} \tag{6.9}$$

where the last equality is obtained via Faraday's law. The equation can now be integrated in time to give

$$\nabla\times\boldsymbol{j}_s = -\frac{e^2 n_s}{m}\boldsymbol{B} + \boldsymbol{C}(\boldsymbol{r}) \tag{6.10}$$

where the last term represents a constant of integration at each point $\boldsymbol{r}$ inside the superconductor. The constant $\boldsymbol{C}(\boldsymbol{r})$ should be determined from the initial conditions. If we start from a superconducting body in zero applied magnetic field, we have $\boldsymbol{j}_s = 0$ and $\boldsymbol{B} = 0$ initially so that $\boldsymbol{C}(\boldsymbol{r}) = 0$. To describe the Meissner-effect, we have to consider the case of a body becoming superconducting (by cooling) in a non-zero applied field. However, the outcome cannot be derived within London

[1] This expression can be derived intuitively as follows. Assume that particles with charge q and mass m in a material are exposed to an external electric field $\boldsymbol{E}$ that is uniform and constant in time. Electrons can scatter on impurities and phonons in the material, and thus can only accumulate a finite momentum between each collision. Assume that each collision occurs on average every τ seconds. Then, an electron will accumulate a momentum $q\boldsymbol{E}\tau$ (force times time) between two collisions. However, each collision may in the simplest approximation be equally likely to bounce the electron forward as backward. On average, all prior contributions to the electron's momentum can therefore be ignored, and the average momentum of the electron can then be taken as

$$\boldsymbol{p} = q\boldsymbol{E}\tau. \tag{6.3}$$

Substituting the relations

$$\begin{aligned}\boldsymbol{p} &= m\boldsymbol{v},\\ \boldsymbol{j} &= nq\boldsymbol{v},\end{aligned} \tag{6.4}$$

results in the formulation of Ohm's law with precisely

$$\boldsymbol{j} = \frac{nq^2\tau}{m}\boldsymbol{E} \tag{6.5}$$

This result can also be formally derived in a more accurate theory utilizing the so-called Kubo formula.

theory since we have assumed the superfluid density n_s to be constant in time.

To account for the flux expulsion, The Londons postulated that $\boldsymbol{C} \equiv 0$ regardless of the history of the system. This leads to

$$\nabla \boldsymbol{j}_s = -\frac{e^2 n}{m}\boldsymbol{B} \tag{6.11}$$

This is the *second London equation*.

To show how this gives rise to a decaying magnetic field inside the superconductor, precisely in line with experimental observations, we take the curl of Ampere's law:

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{j} = \mu_0(\boldsymbol{j}_n + \boldsymbol{j}_s) \tag{6.12}$$

and obtain

$$\nabla \times \nabla \times B = -\frac{e^2 n_s}{m}\boldsymbol{B} - \sigma_n \frac{\partial \boldsymbol{B}}{\partial t}. \tag{6.13}$$

We can drop the last term since we are considering a stationary state, and then use an identity from vector calculates for a double cross product:

$$-\nabla(\nabla \cdot \boldsymbol{B}) + \nabla^2 \boldsymbol{B} = \frac{e^2 n_s}{m}\boldsymbol{B}. \tag{6.14}$$

Introducing the London penetration depth which thus appears as a characteristic length scale for the magnetic field in the superconductor

$$\lambda_L \equiv \sqrt{\frac{m}{e^2 n_s}} \tag{6.15}$$

we have arrived at a simple equation

$$\nabla^2 \boldsymbol{B} = \frac{1}{\lambda_L^2}\boldsymbol{B}. \tag{6.16}$$

To solve this, consider a semi-infinite superconductor filling the half space $x > 0$. An external field $\boldsymbol{B} = B_{\text{ext}}\hat{y}$ is applied parallel to the surface. It follows that the equation above is solved by

$$\boldsymbol{B}(x) = B_{\text{ext}}\hat{y}\mathrm{e}^{-x/\lambda_L} \text{ for } x \geq 0. \tag{6.17}$$

The magnetic field thus decreases exponentially with the distance from the surface of the superconductor. In the bulk, we indeed find $\boldsymbol{B} \to 0$, which is the Meissner effect. Put differently, the photon has become massive inside the superconductor and decays as it propagates in it. This is a physical manifestation of the Higgs mechanism and spontaneous symmetry breaking for those of you familiar with that from *e.g.* particle physics: U(1) symmetry is spontaneously broken in the superconductor, and local gauge invariance renders the gauge field to become massive.

We can also obtain the supercurrent inside the superconductor from the second London equation, now that we have computed the magnetic field. From

$$\nabla \times \boldsymbol{j}_s = -\frac{1}{\lambda_L^2}\boldsymbol{B} \tag{6.18}$$

and the continuity equation for charge current

$$\nabla \cdot \boldsymbol{j}_s = 0 \tag{6.19}$$

we find that

$$\boldsymbol{j}_s(x) = -\frac{1}{\lambda_L} B_{\text{ext}} \hat{z} \mathrm{e}^{-x/\lambda_L} \text{ for x}\geq 0. \tag{6.20}$$

Thus, the supercurrent flows in the direction parallel to the surface and perpendicular to $\boldsymbol{B}$ and decreases into the bulk on the same scale as λ_L. The supercurrent can therefore be understood as a screening current required to keep the magnetic field out of the bulk of the superconductor.

The two London equations can be summarized using the vector potential $\boldsymbol{A}$ since $-\partial_t \boldsymbol{A} = \boldsymbol{E}$ and $\nabla \times \boldsymbol{A} = \boldsymbol{B}$:

$$\boldsymbol{j}_s = -\frac{e^2 n_s}{m} \boldsymbol{A}. \tag{6.21}$$

This equation at first glance seems highly problematic since it is not gauge-invariant: changing $\boldsymbol{A} \to \boldsymbol{A} + \nabla\chi$ changes the supercurrent. The two London equations are in contrast gauge invariant since they are expressed in terms of the physical and gauge invariant fields $\boldsymbol{E}$ and $\boldsymbol{B}$. The solution to this is that the above equation is only valid for a specific choice of gauge $\boldsymbol{A}$. Firstly, charge conservation requires $\nabla \cdot \boldsymbol{j}_s = 0$, so the vector potential satisfying the above equation must be chosen as transverse:

$$\nabla \cdot \boldsymbol{A} = 0. \tag{6.22}$$

This is called the London gauge in this context and the Coulomb gauge more generally. But even this equation is not sufficient to uniquely determine $\boldsymbol{A}$. In order for us to write that the supercurrent is proportional to the vector potential, the vector potential must fulfil the properties of a physical charge current both at the surface of the superconductor and in the bulk. In other words, we must also have $\boldsymbol{A} = 0$ in the bulk of the superconductor (which is consistent with the decay of the magnetic field) and we must have $\boldsymbol{A} \cdot \boldsymbol{n} = 0$ at the surfaces of the superconductor, where $\boldsymbol{n}$ is the normal vector at the surface of the superconductor.

6.3 Ginzburg-Landau theory

As we described in a previous chapter, Landau introduced the concept of the order parameter to describe phase transitions. In this context, an order parameter is a thermodynamic variable that is zero on one side of the transition and nonzero on the other. In ferromagnets, the magnetization $\boldsymbol{M}$ is the order parameter. The theory neglects fluctuations, which means that the order parameter is assumed to be constant in time and space (Ginzburg-Landau theory will go beyond this!). Landau theory is thus a mean-field theory. Now the appropriate thermodynamic potential can be written as a function of the order parameter, which we call Δ for superconductors, and certain other thermodynamic quantities such as pressure or volume, magnetic field, etc. We will always call the potential the free energy F, but whether it really is a free energy, a free enthalpy, or something else depends on which quantities are given (pressure vs. volume etc.). Hence, we write

$$F = F(\Delta, T) \tag{6.23}$$

where T is the temperature, and further variables have been suppressed. The equilibrium state at temperature T is the one that minimizes the free energy. Generally, we do not know $F = F(\Delta, T)$ explicitly. Landau's idea was to expand F in terms of Δ, including only those terms that are allowed by the symmetry of the system and keeping the minimum number of the simplest terms required to get nontrivial results.

Lacking a microscopic theory for superconductivity, which was developed in 1957 by Bardeen, Cooper, and Schrieffer, a reasonable attempt would be to state that the superfluid electrons responsible for the superconducting properties were described by a wavefunction ψ which in general could be complex. The magnitude of this wavefunction should be related to the density of superfluid electrons, so that $|\psi|^2 \propto n_s$. Thus, let us propose the complex wavefunction ψ as the order parameter for a superconductor. The phase of a wavefunction should not have physical consequence since the global phase of quantum states is generally not observable, providing the expansion

$$F = \alpha|\psi|^2 + \frac{\beta}{2}|\psi|^4 + \mathcal{O}(|\psi|^6). \tag{6.24}$$

We have not included any odd powers in $|\psi|$ since they are not differentiable at $\psi = 0$ (the value of the derivative is different depending on which limit you approach $\psi = 0$ from). If $\beta > 0$, $F(\psi)$ is bounded from below, as it should be physically. Now there are two different cases

- If $\alpha \geq 0$, $F(\psi)$ has a single minimum at $\psi = 0$. Thus, the equilibrium state has $n_s = 0$. This is clearly a normal metal/fluid ($n_n = n$ where n is the total electron density $n = n_s + n_n$).
- If $\alpha < 0$, $F(\psi)$ has a ring of minima with equal amplitude $|\psi|$ but arbitrary phase. We see that

  $$\frac{\partial F}{\partial|\psi|} = 2\alpha|\psi| + 2\beta|\psi|^3 = 0 \rightarrow |\psi| = 0 \text{ (this is a maximum) or } |\psi| = \sqrt{-\alpha/\beta}. \tag{6.25}$$

  Note that the radicand is positive.

The behavior of the free energy is shown in Fig. 6.2(a). For $\alpha < 0$, $F(\psi)$ is often called the Mexican-hat or wine-bottle potential (imagine this figure rotated around the vertical axis to find F as a function of the complex ψ). In Landau theory, the phase transition clearly occurs when $\alpha = 0$. Since $T = T_c$ then by definition, it is useful to expand α and β to leading order in T around t_c. Hence,

$$\begin{aligned} &\alpha \simeq \alpha'(T - T_c),\ \alpha' > 0 \\ &\beta \simeq \text{constant}. \end{aligned} \tag{6.26}$$

Then, the order parameter below T_c satisfies

$$|\psi| = \sqrt{\frac{\alpha'}{\beta}}\sqrt{T_c - T}. \tag{6.27}$$

The scaling $|\psi| \propto (T_c - T)^{1/2}$ is characteristic for so-called mean field theories, where fluctuations of the order parameter are neglected and the value of the order parameter is taken to be a fixed value.

All solutions of ψ which has the above value of $|\psi|$ minimize the free energy. However, for a given material in a given experiment, only one such ψ can be realized. The order parameter in the equilibrium state is thus $\psi = |\psi|e^{i\phi}$ with some fixed phase ϕ. This state is itself not invariant under

rotations of the phase, even though the free energy is. We then say that the global U(1) symmetry of the system is spontaneously broken since the free energy F has the symmetry, but the particular equilibrium ground-state realized in the material does not. It is called a U(1) symmetry since the group U(1) of unitary 1×1 matrices contains precisely phase factors $e^{i\phi}$.

Note also that the expansion of F up to fourth order is only justified as long as ψ is small. Its quantitative validity is thus limited to temperatures not too far below T_c. However, it is still possible to gain insight into the qualitative behavior of a system by applying Landau theory outside of this range.

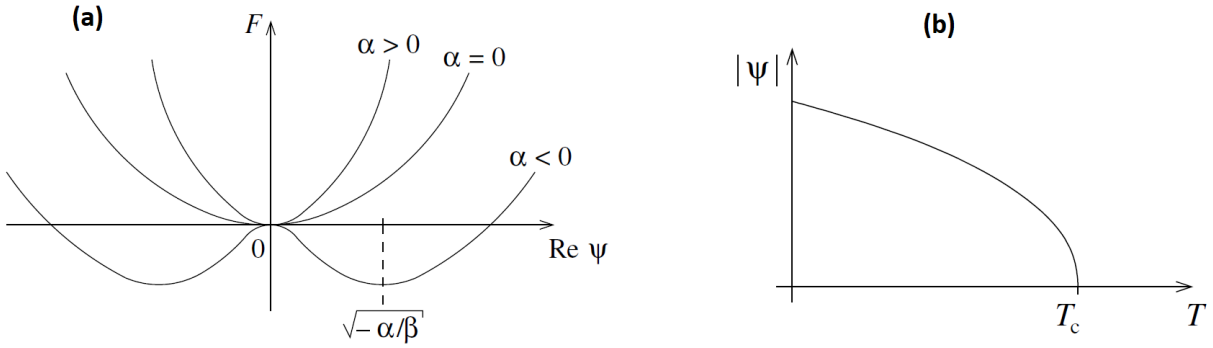

Figure 6.2: **(a)** Free energy as a function of the complex order parameter (along the real axis of the latter) for different choices of α. **(b)** Temperature-dependence of the order parameter in a superconductor. Figure taken from [2].

We now take the step to a more realistic and general situation where we allow the order parameter to be spatially nonuniform. To do this, we have to include terms in the free energy expansion that involve gradients $\nabla\psi$.

It is instructive to first disregard the charge of the electrons, so that they do not couple to any external fields. In effect, we are then considering a neutral superfluid where particles can move without loss, but without carrying charge. After doing this, we introduce charge to the particles so that we describe a superconductor.

We now allow our order parameter $\psi = \psi(\boldsymbol{r})$ to depend on position. Any spatial changes of $\psi(\boldsymbol{r})$ are expected to cost energy. The reason for this expectation is that we know that a quantum mechanical wavefunction increases its kinetic energy the more rapidly it varies in space, due to the kinetic energy term $-\nabla^2/2m$. We thus include the simplest terms containing gradients of ψ which are allowed by symmetry (invariant under a phase change $\phi \to \phi + \phi_0$ of the order parameter $\psi = |\psi| e^{\beta\phi}$):

$$F[\psi] = \int d^3r [\alpha|\psi|^2 + \frac{\beta}{2}|\psi|^4 + \gamma|\nabla\psi|^2] \tag{6.28}$$

Here, we have slightly changed the definition of α and β compared to the uniform case since there is an additional integral over space (which is simply the volume of the system in the uniform case). We must have $\gamma > 0$ in order to prevent the system from being spontaneously highly nonuniform:

a variation in ψ should cost energy. The above expression is a functional of ψ. Strictly speaking, calling it a free energy is an abuse of language since the proper free energy is only the value assumed by $F[\psi]$ at its minimum. Free energy functional would be more suitable, but in practice it is often referred to simply as the free energy. To put the gradient term into a more familiar form resembling kinetic energy, we define $\gamma = \hbar^2/2m^*$ where m^* is the effective mass of the particles that the superconducting wavefunction expresses.

From the above $F[\psi]$, we can now derive a differential equation for the function $\psi(\boldsymbol{r})$ that minimizes F. The derivation is similar to the derivation of the Lagrange-Eulerequation from Hamilton's principle $\delta S = 0$ in classical mechanics where S is the action of the system. Here, we similarly want to find the stationary point of the free energy: $\delta F = 0$. We write

$$\psi(\boldsymbol{r}) = \psi_0(\boldsymbol{r}) + \eta(\boldsymbol{r}) \tag{6.29}$$

where $\psi_0(\boldsymbol{r})$ is the as of yet unknown solution and $\eta(\boldsymbol{r})$ is a small deviation from it. Then:

$$\begin{aligned}
\delta F &= F[\psi_0 + \eta] - F[\psi_0] \\
&= \int d^3r \Big[\alpha\psi_0^*\eta + \alpha\eta^*\psi_0 + \beta\psi_0^*|\psi_0|^2\eta + \beta\eta^*|\psi_0|^2\psi_0 + \frac{\hbar^2}{2m^*}(\nabla\psi_0^*)\nabla\eta + \frac{\hbar^2}{2m^*}(\nabla\eta^*)\nabla\psi_0\Big] \\
&\quad + \mathcal{O}[\eta^2, (\eta^*)^2] \\
&= \int d^3r \Big[\alpha\psi_0^*\eta + \alpha\eta^*\psi_0 + \beta\psi_0^*|\psi_0|^2\eta + \beta\eta^*|\psi_0|^2\psi_0 - \frac{\hbar^2}{2m^*}(\nabla^2\psi_0^*)\eta - \frac{\hbar^2}{2m^*}(\eta^*)\nabla^2\psi_0\Big] \\
&\quad + \mathcal{O}[\eta^2, (\eta^*)^2]
\end{aligned} \tag{6.30}$$

where we obtained the last line by doing an integration by parts. If ψ_0 is to minimize F, the terms linear in η, η^* have to vanish. In effect, the prefactors of η, η^* have to vanish for all $\boldsymbol{r}$. We conclude that

$$\alpha\psi_0 + \beta|\psi_0|^2\psi_0 - \frac{\hbar^2}{2m^*}\nabla^2\psi_0 = 0. \tag{6.31}$$

Let us for brevity drop the subcript 0 from now on and rearrange terms slightly to reveal a striking similarity to the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi_0\alpha\psi_0 + \beta|\psi_0|^2\psi_0 = 0. \tag{6.32}$$

The first two terms look like kinetic energy and a potential, whereas the last term is nonlinear. We now use this to find the solution for $\psi(\boldsymbol{r})$ close to a surface. Imposing a boundary condition $\psi(0) = 0$ and considering 1D for simplicity, we also need a second boundary condition since it is a second order differential equation. But we have such a boundary condition since we've seen that deep inside the bulk of the superconductor the solution should read

$$\lim_{x\to\infty}\psi(x) = \sqrt{-\frac{\alpha}{\beta}}. \tag{6.33}$$

Writing ψ in terms of a dimensionless function, $\psi(x) = \sqrt{-\frac{\alpha}{\beta}}f(x)$, we thus obtain

$$-\frac{\hbar^2}{2m^*\alpha}f''(x) + f(x) - [f(x)]^3 = 0. \tag{6.34}$$

Note that we can choose $f(x)$ to be real since all coefficients in the equation for $\psi(x)$ are real. The above equation reveals that a new characteristic length scale has appeared, namely

$$\xi^2 = -\frac{\hbar^2}{2m^*\alpha} \simeq \frac{\hbar^2}{2m^*\alpha'(T_c - T)} > 0. \tag{6.35}$$

This is called the Ginzburg-Landau coherence length and has a strong temperature dependence, in particular close to T_c. One can solve this equation, and finds that

$$f(x) = \tanh\frac{x}{\sqrt{2}\xi} \tag{6.36}$$

satisfies the proper boundary conditions both at $x = 0$ and $x \to \infty$. The order parameter ψ in the superconductor thus monotonically recovers as one moves from the surface into the bulk. Thus, the coherence length ξ is the characteristic length for spatial variations in $\psi(x)$.

With the above analysis in hand, we now take into account the charge degree of freedom, as we should since the electrons are not neutral. However, to do so, we have to perform the standard procedure of replacing canonical momentum with the gauge-invariant kinetic momentum (also called minimal coupling):

$$\frac{\hbar}{i}\nabla \to \frac{\hbar}{i}\nabla - q\boldsymbol{A}, \tag{6.37}$$

where $\boldsymbol{A}$ is the vector potential. We are taking into account the possibility that q may not be $-e$ by using a general charge. The above substitution introduces a coupling between the electromagnetic field $\boldsymbol{A}$ and the matter described by ψ. We should also include the energy density $B^2/2\mu_0$ of the magnetic field, since we now allow $\boldsymbol{A}$ to be present. Finally, what about spin? Since superconductors experimentally are known to expel magnetic fields, it seems likely that the condensate of superconducting electrons [2] should not carry any net spin polarization. This is true for all conventional and most unconventional superconductors.

We then obtain the total free energy functional

$$\begin{aligned} F[\psi, \boldsymbol{A}] &= \int d^3r \Big[\alpha|\psi|^2 + \frac{\beta}{2}|\psi|^4 + \frac{1}{2m^*}\Big|\Big(\frac{\hbar}{i}\nabla - q\boldsymbol{A}\Big)\psi\Big|^2 + \frac{B^2}{2\mu_0}\Big] \\ &= \int d^3r \Big[\alpha|\psi|^2 + \frac{\beta}{2}|\psi|^4 + \frac{1}{2m^*}\psi^*\Big(\frac{\hbar}{i}\nabla - q\boldsymbol{A}\Big)^2\psi + \frac{B^2}{2\mu_0}\Big]. \end{aligned} \tag{6.38}$$

[2] Some useful definitions follow. *Condensed matter physics:* a field of physics dealing with the physical properties of matter in all its phases, in particular the solid and liquid phase (but also including gas and plasma). The largest subfield of physics. *Solid state physics:* a field of physics dealing with the physical properties of matter in a solid state. A subfield of condensed matter physics. *Soft matter or soft condensed matter*: a field of physics dealing with matter where the physical mechanisms dictating the behavior of the material (such as deformation of a foam or gel) occur at an energy scale of similar magnitude as room temperature thermal energy k_BT. Quantum mechanics plays no particular role in the description of such types of matter. A subfield of condensed matter physics. *Condensate:* here we must delineate between the meaning of this word in a classical and quantum physics context. In a classical physics context, a condensate is the liquid phase of matter produced when a gas cools down and condensates. In a quantum physics context, it has a different meaning. A condensate is a state of matter where a macroscopic number of particles occupy a quantum state which is either identical or share at least one key property. These particles can be composite. If the particles are bosons, they can all occupy exactly the same quantum state. If the particles are composite bosons, like electron pairs in superconductors, they can share one or more key properties, such as having the same phase. If the particles are fermions, they can share a common property, like a macroscopic number of electrons in a ferromagnet all having spin $\uparrow$ in a given direction. A magnet can thus be called a spin condensate.

Minimizing with respect to ψ, we can follow the same procedure as in the previous section:

$$\frac{1}{2m^*}\Big(\frac{\hbar}{i}\nabla - q\boldsymbol{A}\Big)^2\psi + \alpha\psi + \beta|\psi|^2\psi = 0. \tag{6.39}$$

To minimize F with respect to $\boldsymbol{A}$, we write $\boldsymbol{A}(\boldsymbol{r}) = \boldsymbol{A}_0(\boldsymbol{r}) + \boldsymbol{a}(\boldsymbol{r})$ and obtain

$$\begin{aligned} F[\psi, \boldsymbol{A}_0 + \boldsymbol{a}] &= F[\psi, \boldsymbol{A}_0] + \int d^3r\Big[- \frac{q}{2m^*}\Big(\Big[\frac{\hbar}{i}\nabla\Big]^*\psi + \psi^*\frac{\hbar}{i}\nabla\psi\Big)\cdot\boldsymbol{a} \\ &= \frac{q^2}{m^*}|\psi|^2\boldsymbol{A}_0\cdot\boldsymbol{a} + \frac{1}{\mu_0}(\nabla\times\boldsymbol{B}_0)\cdot\boldsymbol{a}\Big] + \mathcal{O}(a^2). \end{aligned} \tag{6.40}$$

Here, we used $\nabla\times\boldsymbol{A}_0 = \boldsymbol{B}_0$ and Gauss' theorem. Again, at the minimum the coefficient of the linear term in $\boldsymbol{a}$ must vanish. Dropping the subscript, we obtain

$$-i\frac{\hbar q}{2m^*}([\nabla\psi^*]\psi - \psi^*\nabla\psi) + \frac{q^2}{m^*}|\psi|^2\boldsymbol{A} + \frac{1}{\mu_0}\nabla\times\boldsymbol{B} = 0. \tag{6.41}$$

With Ampere's law $\boldsymbol{j} = \frac{1}{\mu_0}\nabla\times\boldsymbol{B}$, we thus arrive at

$$\boldsymbol{j} = i\frac{q\hbar}{2m^*}([\nabla\psi^*]\psi - \psi^*\nabla\psi) - \frac{q^2}{m^*}|\psi|^2\boldsymbol{A}, \tag{6.42}$$

where we dropped the subscript 's' on the current $\boldsymbol{j}$, since we assumed that the normal current is negligible. This equation shows that when a material can be described by a macroscopic order parameter $\psi = |\psi|\mathrm{e}^{i\phi}$, which spontaneously breaks U(1) symmetry, a current can flow without any electric field (and is instead driven by a gradient in the phase ϕ if we consider a situation without magnetic field and thus set $\boldsymbol{A} = 0$) and consequently without any voltage drop: this is a supercurrent. This does not happen in a normal metal, since the electrons cannot be described by a single coherent macroscopic wavefunction.

Equations (6.39) and (6.42) are called the Ginzburg-Landau equations. If we take the limit uniform $\psi(\boldsymbol{r})$, Eq. (6.42) simplifies to

$$\boldsymbol{j} = -\frac{q^2|\psi|^2}{m^*}\boldsymbol{A}. \tag{6.43}$$

In order to be consistent with the London equation we derived earlier,

$$\boldsymbol{j} = -\frac{e^2 n_s}{m}\boldsymbol{A} \tag{6.44}$$

we must have

$$\frac{q^2|\psi|^2}{m^*} = \frac{e^2 n_s}{m}. \tag{6.45}$$

In superconductors, experiments show that electrons form pairs by attracting each other via an exchange of phonons. Such pairs are called Cooper pairs and the complete microscopic theory describing how these pairs make up the superconducting state was developed by BCS. We should therefore set $q = -2e$ and $m^* = 2m_e$, which leaves us with a superfluid density

$$n_s = 2|\psi|^2. \tag{6.46}$$

London theory thus follows as a special case from Ginzburg-Landau theory.

6.4 Flux quantization and vortices

The two characteristic length scales we have found in our analysis are:

- λ: the magnetic field penetration depth into a superconductor.
- ξ: the coherence length providing the characteristic length for spatial variations of the order parameter.

Superconductors can now be divided into two types depending on the ratio of these parameters, called the Ginzburg-Landau parameter $\kappa = \lambda/\xi$. If $\kappa \ll 1$, the superconductor is said to be of type I. If $\kappa \gg 1$, it is said to be of type II. The separating value is $\kappa = 1/\sqrt{2}$, and we will justify this below. Since both λ and ξ in our treatment scale as $(T_c - T)^{-1/2}$, κ is roughly temperature-independent in the superconducting state.

Type I and II superconductors behave differently in an applied magnetic field. This is because the value of κ determines whether it is more energetically favorable for the magnetic field to be shut out from the superconductor all the way up to the normal phase transition at $T = T_c$, *or* if is actually energetically favorable to partially transmit the magnetic field through the superconductor through magnetic "channels" called vortices. We here give a simple argument for this distinction.

Consider a type I and type II superconductor near an interface ($r = 0$) where superconductivity is absent. The superconductivity then recovers over a distance ξ, whereas screening extends over a distance λ. The ratio between these length scales are different in type I and type II superconductors, as explained above. We now analyze the energy gain and loss in the system due to the expulsion of magnetic field and suppression of superconductivity.

To expel the magnetic field from the volume of the superconductor costs an energy equal to the magnetic field energy. This is $B_c^2/(2\mu_0)$ per volume. Since this happens over a volume λA, where A is the interface area, we are reducing the amount of energy in the system by $E_\lambda = \lambda A B_c^2/(2\mu_0)$.

At the same time, the formation of Cooper pairs is also suppressed in a volume ξA since superconductivity recovers over a distance ξ. Remarkably, it turns out that the energy gain (so-called condensation energy) from having a superconducting condensate is also equal to $B_c^2/(2\mu_0)$ per unit volume [3]. Therefore, by suppressing superconductivity close to the interface, we are increasing the

[3]The appropriate free energy describing a system at a fixed temperature T and with an applied external field $\mu_0 H$ is the Gibbs free energy G. Let $g = G/V$ denote the Gibbs free energy density. Upon changing the external field by dH, the change in energy density is $dg = -\mu_0 M dH$. This holds both in the normal and superconducting state. To find the total free energy change upon increasing the field from H to the critical field H_c at a constant temperature in the superconducting state, we obtain

$$\int_0^{H_c} dg_s = -\mu_0 \int_0^{H_c} M_s(T,H) dH. \tag{6.47}$$

Consider now a type I superconductor. The Meissner state is characterized by $B = 0$ inside the superconductor, in effect $M = -H$. Therefore,

$$\int_0^{H_c(T)} dg_s = -\mu_0 \int_0^{H_c(T)} (-H) dH = \frac{\mu_0}{2} H_c^2(T). \tag{6.48}$$

On the left hand-side, we obtain

$$\int_0^{H_c} dg_s = g_s(T,H_c) - g_s(T,0), \tag{6.49}$$

energy of the system by an amount $E_\xi = \xi A B_c^2/(2\mu_0)$.

The energy difference

$$\Delta E = E_\xi - E_\lambda = (\xi - \lambda) A B_c^2/(2\mu_0) \tag{6.53}$$

then determines whether or not the system wants to maximize the surface area between superconducting and normal regions. This shows that if $\xi > \lambda$, it is not beneficial energetically for the superconductor to set up interfaces to normal regions since $\Delta E > 0$. This is the type I case, up to a factor of $1/\sqrt{2}$. Instead, if $\xi < \lambda$, there is an energy gain when the superconductor maximizes the number of interface areas to non-superconducting regions. This is precisely what happens when the superconductor allows the formation of the aforementioned vortices, which are tubes of magnetic flux where superconductivity is suppressed that penetrate through the superconductor. This corresponds to type II superconductivity.

We note in passing that when taking into account specific sample geometries, there are so-called demagnetization effects taking place. This means that the induced magnetic field from the supercurrent alters the net magnetic field at the edges of the superconductor so that it differs from the external field. In effect, due to the magnetic response of the superconductor, the field right outside of it is not in general simply equal to the external field. It can for instance focus the magnetic field lines in certain parts of the sample, depending on its shape. This also causes the net magnetic field inside the material to deviate slightly from the applied magnetic field. As a result, it is possible to reach a magnetic field value inside the material exceeding the critical strength B_c even when applying an external field that is smaller than B_c.

But this seems to give rise to a paradox. If superconductivity is destroyed by a field smaller than B_c, then upon entering the normal state there are no induced supercurrents which try to screen the external field. This means that the field inside the material is equal to the external field, which is smaller than B_c. Then, it looks like we have destroyed superconductivity with a field that is smaller than B_c. The solution to this is that the superconductor compromises. It allows thin normal laminae to form, so that part of the material can still benefit from the condensation energy. For a small demagnetization factor, the superconductor does not enter this so-called intermediate state until the field becomes very close to the critical field.

---

so that

$$g_s(T, H_c) - g_s(T, 0) = \frac{\mu_0}{2} H_c^2(T). \tag{6.50}$$

We now want to calculate the energy difference between the normal and the superconducting state. This will tell us which of the two states that is thermodynamically favored as the ground-state. To do so, recall that the magnetic susceptibility χ in a normal metal is usually small, $\chi \ll 1$. This means that the Gibbs free energy in a normal metal at a field H_c is practically the same as the Gibbs free energy in a normal metal at zero field. Thus, we write $g_n(T, H_c) = g_n(T, 0)$. At the critical field H_c, the free energies of the superconducting and normal states are equal:

$$g_n(T, H_c) = g_s(T, H_c). \tag{6.51}$$

Inserting now $g_n(T, H_c) = g_n(T, 0)$ and Eq. (6.50) brings us to

$$g_n(T, 0) - g_s(T, 0) = \frac{1}{2\mu_0} B_c^2(T). \tag{6.52}$$

This is an interesting result: the difference in the free energy density between a normal metal and a superconductor can be expressed in terms of the critical field B_c of the superconductor. It shows that the normal-state has a higher energy than the superconductor at zero field, and thus the superconducting state is the preferred ground-state.

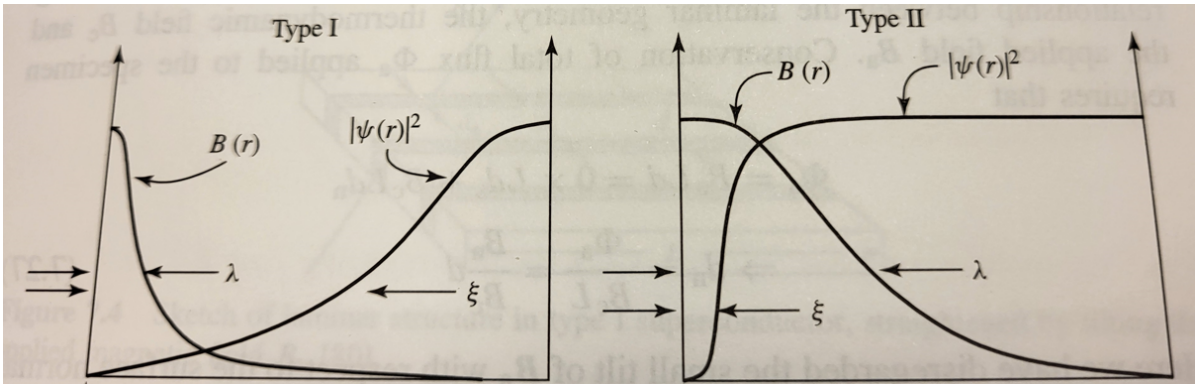

Figure 6.3: **Left:** variation of the magnetic field B and superconducting wavefunction ψ near a boundary where superconductivity vanishes for a type I superconductor. **Right:** Same for a type II superconductor. Figure taken from [10].

The flux penetrating vortices can only take certain quantized values. We now examine such vortices in more detail. Let us start by establishing the quantization of flux. Consider the surface of a superconductor S and a closed path ∂S enclosing the surface. Assume an external magnetic field has been applied so that it penetrates this surface, as shown in the left part of Fig. 6.4.

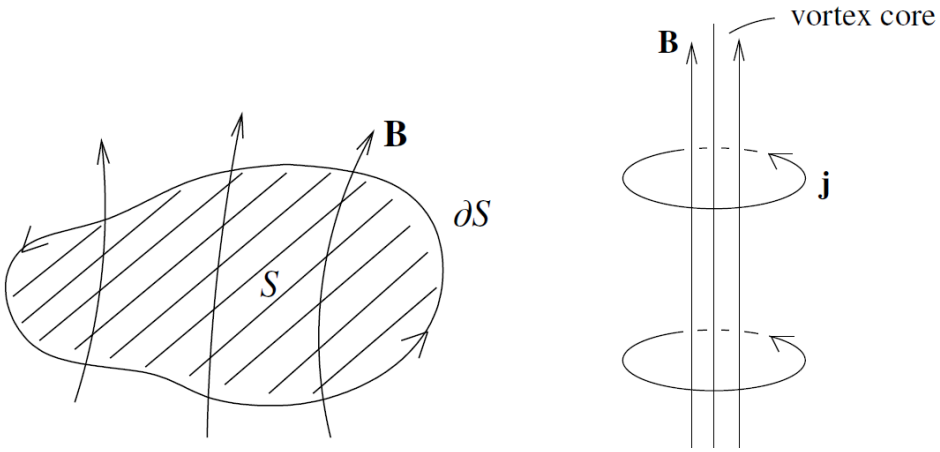

Figure 6.4: **Left:** magnetic field penetrates a superconducting surface. **Right:** magnetic flux can penetrate a type II superconductor through vortices, which are accompanied by a circulating supercurrent at the surface of the superconductor. Figure taken from [2].

The magnetic flux through this loop is

$$\Phi = \int_S d\boldsymbol{a} \cdot \boldsymbol{B} = \int_S d\boldsymbol{a} \cdot (\nabla \times \boldsymbol{A}) = \oint d\boldsymbol{s} \cdot \boldsymbol{A} \tag{6.54}$$

where we used Stokes' theorem. By means of the second Ginzburg-Landau equation

$$\boldsymbol{j} = i\frac{q\hbar}{2m^*}([\nabla\psi^*]\psi - \psi^*\nabla\psi) - \frac{q^2}{m^*}|\psi|^2\boldsymbol{A}, \tag{6.55}$$

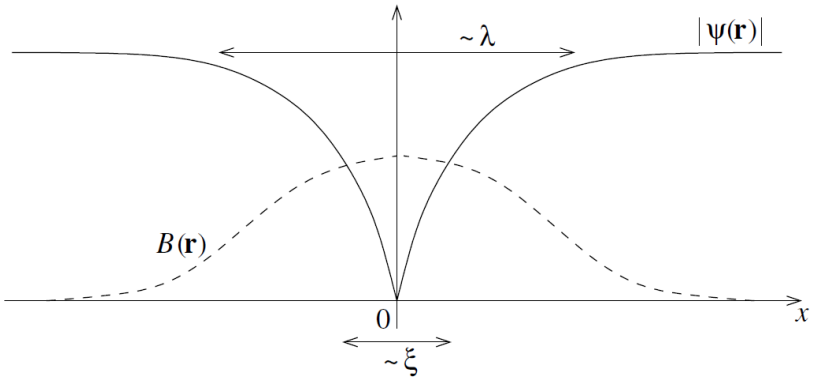

Figure 6.5: Sketch of the solution for $\psi(\boldsymbol{r})$ and $\boldsymbol{B}$ near a vortex in a superconductor. Figure taken from [2].

we can replace $\boldsymbol{A}$ in the flux equation with

$$\begin{aligned}\Phi &= \oint_{\partial S} d\boldsymbol{s}\cdot\Big\{-\frac{m^*}{q^2}|\psi|^2\boldsymbol{j}+i\frac{\hbar}{2q|\psi|^2}([\nabla\psi^*]\psi-\psi^*\nabla\psi)\Big\}\\ &= \oint_{\partial S} d\boldsymbol{s}\cdot\Big\{-\frac{m}{e^2n_s}\boldsymbol{j}+i\frac{\hbar}{2q}(-2i)\nabla\phi\Big\}\\ &= \frac{m}{e}\oint_{\partial S} d\boldsymbol{s}\cdot\boldsymbol{v}_s-\frac{\hbar}{2e}\oint_{\partial S} d\boldsymbol{s}\cdot\nabla\phi. \end{aligned} \tag{6.56}$$

Here, we renamed m_e to m and also defined the superfluid velocity $\boldsymbol{v}_s = -\boldsymbol{j}/en_s$. The order parameter $\psi(\boldsymbol{r})$ must be a single-valued quantity at any point in $\boldsymbol{r}$ to be physically acceptable. This means that the last term containing the integral over $\nabla\phi$ *must* equation an integer multiple of 2π in order for ψ to be single-valued and continuous:

$$\oint_{\partial S} d\boldsymbol{s}\cdot\nabla\phi = -2\pi n,\ n\in\mathbb{Z}. \tag{6.57}$$

The minus sign is conventional, since n can be any positive or negative integer. Thus, we find that

$$\Phi-\frac{m}{e}\oint_{\partial S} d\boldsymbol{s}\cdot\boldsymbol{v}_s=\frac{\hbar}{2e}2\pi n=n\Phi_0 \tag{6.58}$$

where $\Phi_0=\frac{h}{2e}$ is known as the flux quantum. Evidently, the flux and the circulation of the supercurrent around the surface must be quantized in units of Φ_0. However, deep inside a superconducting region the current vanishes (since it decays on a length scale of the penetration depth λ, as we showed previously), and thus the flux Φ alone is quantized. It is experimental measurements of the quantized value of the flux which reveals that the effective charge of the superconducting wavefunction is $q=-2e$, which we alluded to above.

This leads us to the concept of *vortices*. The manner in which flux can pass through the superconductor, accompanied by a circulating superfluid current, is via a vortex (see right part of Fig. 6.4). Consider the smallest amount of flux that is permissible, namely $n=1$ corresponding to a single flux quantum. Following our above arguments, the phase ϕ of ψ changes by -2π as one circles a

vortex in the positive direction, where /by convention) the direction of a vortex is the direction of the magnetic field $\boldsymbol{B}$. In the center of the vortex, called the vortex core, the phase is undefined since the length of the circulating path is zero. This is only consistent with a continuous ψ if $\psi = 0$ in the vortex core. The phase winding of 2π occurs for any closed path around the vortex and thus does not depend on the precise shape of the path, nor on the magnitude of ψ so long that it is finite. In this sense, the phase winding is a topological property of the vortex as it cannot be changed by any continuous deformation of ψ, and can only vanish if ψ itself vanishes. Therefore, vortices are called topological defects of the superconducting state.

To describe the order parameter profile $\psi(\boldsymbol{r})$ and magnetic field $\boldsymbol{B}$ in a superconductor which includes a vortex along the z-axis, one has to solve the GL equations together with Ampere's law with the following boundary conditions

$$\psi(r \to 0) = 0,$$
$$|\psi(r \to \infty)| = \psi_0, \tag{6.59}$$
$$B(r \to \infty) = 0, \tag{6.60}$$

where $r = |\boldsymbol{r}|$. The boundary conditions express that the order parameter should vanish in the vortex core whereas the bulk solution for both ψ and the magnetic field (which is completely screened in the bulk of the superconductor) has to be recovered far away from the vortex. Due to the non-linear term in the first GL euqation, this can only be done numerically in general. We sketch the solution in the figure below.

The order parameter recovers over a distance comparable to the coherence length ξ, whereas the magnetic field is suppressed over a distance comparable with the magnetic penetration depth, perfectly in line with our understanding of these quantities described earlier. In a type II superconductor, the Meissner state thus persists up to a critical field B_{c1} where the field is large enough that one flux quantum can penetrate the superconductor. It does so in the form of a vortex. Afterwards, more and more vortices are generated as B increases, until superconductivity is finally destroyed all together at the second critical field B_{c2}. The complete destruction of superconductivity will take place when the field $B = B_{c2}$ is large enough that there are so many vortices that their vortex cores start to overlap. In this case, there are no more places for the Cooper pairs to exist.

Chapter 7

Diamagnetism, paramagnetism, and ordered magnetic states

7.1 What is magnetism?

It has been known since antiquity that "loadstone" (magnetite, Fe_3O_4) and iron attract each other. Plato and Aristotle mention permanent magnets. They are also mentioned in Chinese texts from the 4th century B.C. The earliest mention of a magnetic compass used for navigation is from a Chinese text dated 1040–1044 A.D., but it may have been invented there much earlier. It was apparently first used for orientation on land, not at sea.

Thus magnetism at first referred to the long-range interaction between ferromagnetic bodies. In this chapter, we will address magnetic order in solids, of which ferromagnetism is the most straightforward case. This begs the question of what it is that is ordering in a ferromagnet.

Oersted (1819) found that a compass needle is deflected by a current-carrying wire in the same way as by a pemanent magnet. This and later experiments led to the notion that the magnetization of a permanent magnet is somehow due to pemanent currents of electrons. Biot, Savart, and Ampere established the relationship of the magnetic induction and the current that generates it. As we know, Maxwell essentially completed the classical theory of electromagnetism.

7.1.1 The Bohr-van Leeuwen theorem

Can we understand ferromagnetism in terms of magnetic dipole moments created by electron currents in the framework of Maxwell's classical electrodynamics? To answer this, consider N classical electrons with positions $\boldsymbol{r}_i$ and momenta $\boldsymbol{p}_i$. The partition function in classical statistical mechanics reads as

$$Z \propto \int \prod_i d^3r_i d^3p_i \exp[-\beta H(\boldsymbol{r}_1, \ldots; \boldsymbol{p}_1, \ldots)], \tag{7.1}$$

where $\beta = 1/k_B T$ and H is the classical Hamilton function

$$H = \frac{1}{2m} \sum_i (\boldsymbol{p}_i + e\boldsymbol{A}(\boldsymbol{r}_i))^2 + V(\boldsymbol{r}_1, \ldots) \tag{7.2}$$

Here, $\boldsymbol{A}$ is the usual vector potential providing the magnetic field $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. This means that the magnetic field is generated due to currents via the Biot-Savart or Ampere-Maxwell laws, in addition

to a possible external magnetic field, because there is no spin degree of freedom included here or magnetic field dependence in the potential V (such as dipole-dipole interactions). The electron charge is $-e$.

But now we may perform the variable substitution $\boldsymbol{p}_i \to \tilde{\boldsymbol{p}}_i = \boldsymbol{p}_i + e\boldsymbol{A}(\boldsymbol{r}_i)$ in the integrals, providing

$$Z \propto \int \prod_i d^3 r_i d^3 \tilde{p}_i \exp\Big[-\beta\Big(\frac{1}{2m}\sum_i \tilde{\boldsymbol{p}}_i^2 + V\Big)\Big]. \tag{7.3}$$

We see that the vector potential $\boldsymbol{A}$ has been eliminated from the partition function. The magnetization is computed from the free energy via

$$\boldsymbol{M} = -\frac{\partial F}{\partial \boldsymbol{B}}, \tag{7.4}$$

and the free energy is determined from the partition function $F = -\frac{1}{\beta}\ln(Z)$: it follows that $\boldsymbol{M} = 0$. This is called the Bohr-van Leeuwen theorem.

This proves that one cannot obtain equilibrium ferromagnetism in a theory that

- is classical *and*
- assumes that the magnetic field to be due to currents alone.

Yet we know that ferromagnetic materials exist, so what is missing? The solution is that we require an intrinsic magnetic moment carried by the electrons, which is attributed to an intrinsic angular momentum, the spin. As we will show in this chapter, ordered magnetic states in solids is a quantum mechanical effect and arises due to two main ingredients: the Coulomb interaction and the Pauli principle, the latter being a purely quantum mechanical concept. In principle, one could argue that the classical dipole-dipole interaction could provide magnetic ordering, but taking into account the strength of this interaction (which depends on the distance between neighboring magnetic moments and the size of the magnetic moment), one finds that the ordering temperature T_c becomes orders of magnitude lower than what is observed experimentally. Thus, there has to exist a different much stronger interaction between magnetic moments which causes them to order even at high temperatures.

7.2 Paramagnetism of the electron gas

Before proceeding to discuss how magnetic ordering arises in materials due to quantum mechanical effects, we consider as a preliminary magnetic properties in the absence of the Coulomb interaction, i.e. the free electron gas.
Paramagnetism is due to the presence of unpaired electrons in the material, meaning that most substances with incompletely filled atomic orbitals act paramagnetically, although exceptions exist (such as copper). A rule of thumb is to say that if an atom or molecule has electrons which all are paired in their orbitals, then it will act diamagnetically - if it has unpaired electrons, it will act paramagnetically.

Electrons possess a spin with quantum number $s = 1/2$ and a spin magnetic moment $m_s = g\mu_B/2 \simeq \mu_B$ oriented oppositely to the spin. The energy of a free electron in a uniform magnetic field is then

$$\varepsilon_{\boldsymbol{k},\sigma} = \frac{\hbar^2 k^2}{2m} + \sigma\frac{g\mu_B B}{2},\ \sigma = \uparrow, \downarrow = \pm 1. \tag{7.5}$$

The total energy is

$$E = \sum_{\boldsymbol{k}\sigma} \varepsilon_{\boldsymbol{k}\sigma} n_F(\varepsilon_{\boldsymbol{k}\sigma} - \mu) \tag{7.6}$$

where μ is the chemical potential. The magnetization is given by the difference of the total spin up and down electrons:

$$M = -\frac{g\mu_B}{2V} \sum_{\boldsymbol{k}\sigma} \sigma n_F(\varepsilon_{\boldsymbol{k}\sigma} - \mu) \tag{7.7}$$

and thus the susceptibility is

$$\begin{aligned} \chi = \left.\frac{\partial M}{\partial H}\right|_{B=0} &\overset{\chi \ll 1}{\simeq} \mu_0 \left.\frac{\partial M}{\partial B}\right|_{B=0} = -\frac{\mu_0 g \mu_B}{2V} \sum_{\boldsymbol{k}\sigma} \sigma n_F'(\hbar^2 k^2/2m - \mu) \sigma \frac{g\mu_B}{2} \\ &= -\frac{\mu_0 g^2 \mu_B^2}{2V} \sum_{\boldsymbol{k}} n_F'(\hbar^2 k^2/2m - \mu) \end{aligned} \tag{7.8}$$

where we defined $n_F'(x) = dn_F(x)/dx$. The density of states for a single spin direction is $D(\varepsilon) = \frac{1}{V}\sum_{\boldsymbol{k}} \delta(\varepsilon - \hbar^2 k^2/2m)$, giving us

$$\chi = -\frac{\mu_0 g^2 \mu_B^2}{2} \int d\varepsilon D(\varepsilon) n_F'(\varepsilon - \mu). \tag{7.9}$$

So long that $k_B T \ll \mu$, which is typically the case for metals since a thermal energy of 1 eV requires $T = 10^4$ K, we can approximate n_F by a step function so that n_F' is close to a Dirac delta-function: $n_F'(\varepsilon) \simeq -\delta(\varepsilon)$. We arrive at

$$\chi = \frac{\mu_0 g^2 \mu_B^2 D(\mu)}{2} = \chi_{\text{Pauli}}. \tag{7.10}$$

This is called the Pauli susceptibility, describing Pauli paramagnetism of an electron gas. The result is valid for any didspersion, not just for free electrons, if the appropriate density of states is inserted. Note that the above expression is equal to our expression for the Pauli suscpetibility derived in an earlier chapter, by noting that the total DOS is $N(\mu) = 2D(\mu)$ and $g \simeq 2$, giving $\chi = \mu_0 \mu_B^2 N(\mu)$.

Note that this is quite different from the Curie law for the paramagnetic behavior of local magnetic moments, where $\chi \propto 1/T$. Unlike Pauli paramagnetism derived above, which is suitable for metals where conduction electrons are weakly interacting and delocalized in space forming a free gas, Curie paramagnetism occurs in materials with a very different electronic structure cmopared to simple metals such as rare earths (Sc, Y, La) and transition metals. Take the iron group transition elements: the 4s state is filled before the 3d ones, so that the inner electrons form an incomplete shell and produces a permanent magnetic moment. This causes the transition elements to produce a strong susceptibility, as indicated by the $1/T$ dependence in the Curie law. Note that they are still paramagnetic, and not ferromagnet, so long that the permanent magnetic moments of the atoms do not align spontaneously with each other.

7.3 Diamagnetism of the electron gas

There is also a diamagnetic contribution to the susceptibility of free electrons, which is due to the magnetic moments generated by charge currents. It would still be present if the electrons had

no spin (or rather no spin magnetic moment). We know from the Bohr-van Leeuwen theorem that this diamagnetic response has to be a quantum-mechanical phenomenon. Interestingly, the dimensionality of the electron gas is crucial with respect to the magnitude of the diamagnetic contribution.

7.3.1 The 2D electron gas

Consider a two-dimensional electron gas in a uniform magnetic field

$$H = \frac{1}{2m}[\boldsymbol{p} + e\boldsymbol{A}(\boldsymbol{r})]^2 \tag{7.11}$$

where we ignored the Zeeman term, since we already know that this leads to Pauli paramagnetism. Consider a uniform field $\boldsymbol{B} = B\hat{\boldsymbol{z}}$ and choose a Landau gauge $\boldsymbol{A}(\boldsymbol{r}) = (-By, 0, 0)$. In this gauge, we find that $[H, p_x] = 0$ since H does not contain x. Therefore, p_x is a constant of motion and can be replaced by its eigenvalue $\hbar k_x$ since the wavefunction must have a planewave form in the x-direction. Defining

$$y_0 = \frac{\hbar}{eB}k_x,\ \omega_c = \frac{eB}{m}, \tag{7.12}$$

we can write

$$H = \frac{1}{2m}p_y^2 + \frac{1}{2}m\omega_c^2(y - y_0)^2. \tag{7.13}$$

Here, ω_c is the cyclotron frequency. The resulting Hamiltonian has a well-known form: it describes a harmonic oscillator with potential minimum shifted to y_0. The eigenvalues are immediately known:

$$E_{n,k_x} = \hbar\omega_c(n + 1/2),\ n = 0, 1, 2, \ldots \tag{7.14}$$

The energies do not depend on k_x and are thus hugely degenerate. Note that the apparent asymmetry between the x- and y-direction in $\boldsymbol{B}$ is gauge-dependent and therefore without physical consequence. We could have chosen the vector potential $\boldsymbol{A}$ to point in any direction within the xy-plane. The isotropy of two-dimensional space is not broken by the choice of a special gauge.
The presence of the magnetic field transforms the continuous spectrum of a 2D electron gas, $E_{\boldsymbol{k}} = \hbar^2(k_x^2 + k_y^2)/2m$, into a discrete spectrum of Landau levels enumerated by the quantum number n. For $B = 0$, the density of states is

$$\begin{aligned} D(\varepsilon) &= \int \frac{d^2k}{(2\pi)^2}\delta(\varepsilon - \hbar^2k^2/2m) \\ &= \frac{1}{4\pi}\int_0^\infty dE\frac{2m}{\hbar}\delta(\varepsilon - E) \\ &= \frac{m}{2\pi\hbar^2}. \end{aligned} \tag{7.15}$$

The DOS is thus constant. For $B > 0$, it is replaced by equidistant δ-function peaks as shown in Fig. 7.1, since the integrand becomes momentum-independent.
The degeneracy of the Landau levels can be obtained by considering a system of size $L \times L$. Using periodic boundary conditions, *e.g.* $\psi(x + L_x) = \psi(x) \propto e^{ik_x x}$ for the wavefunction, the allowed values of the momentum is $k_x = 2\pi n_x/L_x$ where n_i are integers. Note that using periodic boundary conditions allows us to use free-particle wavefunctions to count the number of states in contrast to hard-wall boundary conditions where $\psi = 0$ at the edges, while still obtaining the same density of

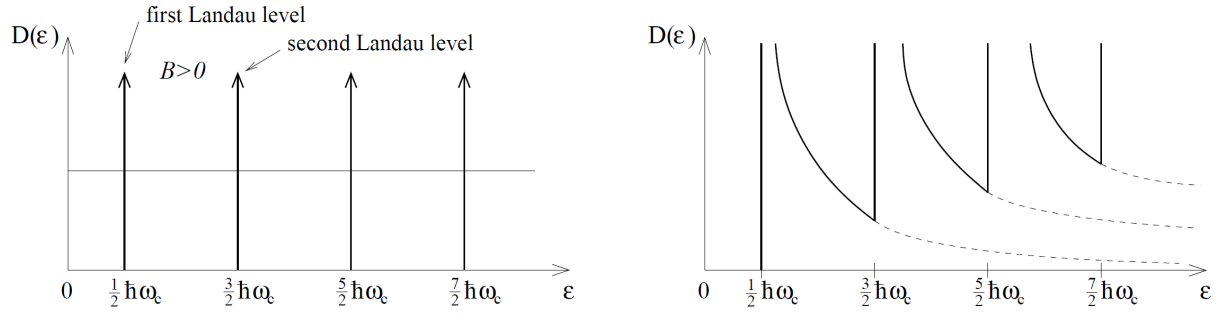

Figure 7.1: **Left:** DOS for a 2D electron gas with magnetic field. **Right:** DOS for a 3D electron gas with magnetic field. Figure taken from [1].

states. In turn, this means that the allowed values for $y_0 = \hbar k_x/eB$ are separated by $\Delta y_0 = h/eBL_x$. The number of available positions for y_0 then becomes:

$$\frac{L}{\Delta y_0} = L^2\frac{eB}{h} = \frac{\Phi_{\text{tot}}}{h/e} \equiv N_L. \tag{7.16}$$

Here, $\Phi_{\text{tot}} = L^2B$ is the total magnetic flux through the area $L \times L$. We may conclude that each Landau level contains the same number of states: $\Phi_{\text{tot}}/(h/e)$. Each state is then associated with the flux quantum $\Phi_0 = h/e$.

The degeneracy level is thus the same for all levels. At low temperatures, the low-energy states are filled successively until all N electrons are accommodated. If $N = 2\pi N_L$ with $n = 1, 2, \ldots$, the lowest n Landau levels are completely filled while the others are empty. The factor 2 comes from two possible spin directions, as we have neglected the Zeeman-energy. The total energy of the N electrons is

$$E = \sum_{n=0}^{\tilde{n}-1} 2N_L\hbar\omega_c(n + 1/2) + (N - \tilde{n}2N_L)\hbar\omega_c(\tilde{n} - \frac{1}{2}) \tag{7.17}$$

where we defined $\tilde{n} = [\frac{N}{2N_L}]$ and the function $[x]$ is the largest integer smaller or equal to x. The first term in E describes the filled Landau levels, and the second the partially filled one. Introducing the filling factor $\nu \equiv N/N_L$, we obtain the energy per electron:

$$E/N = \sum_{n=0}^{\tilde{n}-1} \frac{2}{\nu}\hbar\omega_c(n + 1/2) + (1 - 2\tilde{n}/\nu)\hbar\omega_c(\tilde{n} + 1/2). \tag{7.18}$$

While this function is continuous as a function of B, it is not everywhere differentiable, as seen in the figure. In the figure, B_t is the field for which $\nu = 2$, i.e. for which the lowest Landau level is completely filled:

$$\frac{N}{N_L} = \frac{N}{eB_tL^2/h} = 2 \rightarrow B_t = \frac{h}{2e}\frac{N}{L^2}. \tag{7.19}$$

The magnetization is generally $M = -\frac{1}{V}\partial F/\partial B$ where F is the free energy of the system. At low temperatures, this is simply the internal energy E of the system (without the contribution $-TS$

from entropy), so the magnetization at low T can be estimated as

$$M = -\frac{1}{L^2}\frac{\partial E}{\partial B} \tag{7.20}$$

where we use L^2 for V since we have a 2D system. Numerical evaulation of this expression shows that it oscillates with a periodicity $\propto 1/B$, a phenomenon known as de Haas-van Alpen oscillations.

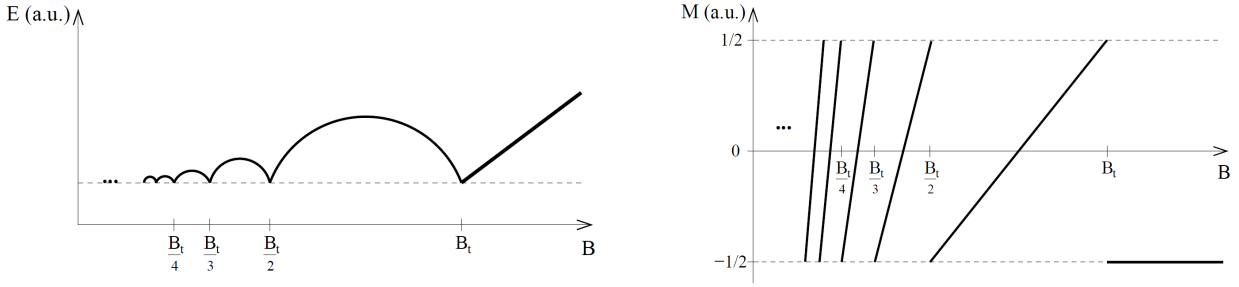

Figure 7.2: **Left:** energy of the 2D Landau level system as a function of magnetic field. **Right:** magnetization of the 2D Landau level system as a function of magnetic field. "a.u." means arbitrary units. Figure taken from [1].

We would now like to compute the susceptibility corresponding to this magnetization. The problem is that $\lim_{B\to 0} M$ does not exist and neither does the limit $\lim_{B\to 0}\chi$ for the susceptibility $\chi = \mu_0 \partial M/\partial B$. The reason for this unphysical behavior is that we set T to strictly zero. At any $T > 0$, the thermal energy $k_B T$ will be large compared to the energy spacing $\hbar\omega_c$ between the Landau levels as long as B is sufficiently small. In this regime, we can neglect the discreteness of the levels and write

$$\begin{aligned} E &\overset{N\gg 2N_L}{\simeq} \int_0^{N/2N_L} dn 2N_L \hbar\omega_c n = \frac{2eB}{h} L^2 \frac{\hbar eB}{m}\left[\frac{n}{2}\right]_0^{Nh/2eBL^2} \\ &= \frac{N^2 h^2}{8\pi m}\frac{1}{L^2}. \end{aligned} \tag{7.21}$$

The magnetization thus vanishes

$$M = -\frac{1}{L^2}\frac{\partial E}{\partial B} = 0. \tag{7.22}$$

This means that the susceptibility $\chi = \mu_0 \partial M/\partial B = 0$ of a 2D electron gas vanishes: it cannot be magnetized.

7.3.2 The 3D electron gas

The Hamiltonian of free electrons in 3D in the presence of a uniform magnetic field $\boldsymbol{B} = B\hat{\boldsymbol{z}}$ reads as

$$H = \frac{1}{2m}[\boldsymbol{p} + e\boldsymbol{A}(\boldsymbol{r})]^2 = \frac{1}{2m}p_y^2 + \frac{1}{2}m\omega_c^2(y - y_0)^2 + \frac{p_z^2}{2m} \tag{7.23}$$

upon choosing again the Landau gauge. Now, we have free motion along the z-direction (generally, the direction of the applied field) in addition to shifted harmonic oscillators in the xy-plane. The

eigenenergies can immediately be written down for such a physical system

$$E_{n,k_x,k_y,k_z} = \hbar\omega_c(n+1/2) + \frac{\hbar^2 k_z^2}{2m} \tag{7.24}$$

The DOS is therefore the sum of the density of states of a 1D electron gas (in the z-direction) shifted to the minimum energies $\hbar\omega_c(n+1/2)$, $n = 0, 1, 2, \ldots$. For one spin direction, this gives

$$D(\varepsilon) = \frac{N_L}{L^2}\sum_{n=0}^{\infty}\frac{1}{\pi\hbar}\sqrt{\frac{m}{2[\varepsilon-\hbar\omega_c(n+1/2)]}}\Theta\Big(\varepsilon-\hbar\omega_c(n+1/2)\Big) \quad = \frac{1}{\pi\hbar}\sqrt{\frac{m}{2}}\frac{N_L}{L^2}\sum_{n=0}^{\infty}\frac{\Theta[\varepsilon-\hbar\omega_c(n+1/2)]}{\sqrt{\varepsilon-\hbar\omega_c(n+1/2)}}.$$

For $k_BT \ll \hbar\omega_c$, the low-energy states are again filled up until all electrons are accommodated. We therefore also expect to find special features whenever the chemical potential μ reaches $\hbar\omega_c(n+1/2)$, as seen for instance in the DOS.

We now want to compute the magnetic susceptibility for small fields. Thus, we have $\hbar\omega_c \ll k_BT$ and we still assume $k_BT \ll \mu$ as for a typical metal. The total energy at low temperature (where the Fermi-Dirac distribution is unity below the Fermi level) is

$$E = 2V\int_0^{\mu} d\varepsilon\varepsilon D(\varepsilon) \tag{7.25}$$

where 2 is for spin. Inserting the DOS provides (see exercise for details):

$$E = E(B=0) + \frac{1}{12\pi^2}V\frac{e^2\sqrt{\mu}}{\sqrt{2m}\hbar}B^2 \tag{7.26}$$

so that

$$M = -\frac{1}{V}\frac{\partial E}{\partial B} = -\frac{1}{6\pi^2}\frac{e^2\sqrt{\mu}}{\sqrt{2m}\hbar}B \tag{7.27}$$

and finally we arrive at our desired susceptibility

$$\chi = \frac{\mu_0\partial M}{\partial B} = -\frac{1}{6\pi^2}\frac{e^2\sqrt{\mu}}{\sqrt{2m}\hbar}. \tag{7.28}$$

This can be expressed in terms of the zero-field density of states in the system $D(\mu) = m\sqrt{2m\mu}/(2\pi^2\hbar^3)$:

$$\chi = -\frac{e^2\hbar^2}{6m^2}D(\mu) \overset{g\simeq 2}{\simeq} -\frac{g^2\mu_B^2}{6}D(\mu). \tag{7.29}$$

This shows that the diamagnetic susceptibility, called the Landau susceptibility, satisfies:

$$\chi = \chi_{\text{Landau}} = -\frac{1}{3}\chi_{\text{Pauli}}. \tag{7.30}$$

This describes Landau diamagnetism: the orbital motion of the electrons generates a magnetic moment which opposes the external field. The result is essentially temperature-independent as long as $k_BT \ll \mu$. The total susceptibility is obtained by adding the paramagnetic and diamagnetic contribution, so that

$$\chi = \chi_{\text{Pauli}} + \chi_{\text{Landau}} = \frac{2}{3}\chi_{\text{Pauli}}. \tag{7.31}$$

Examples of Pauli-paramagnetic metals are W, Al, Li, Mg, but in some metals the diamagnetic contribution can actually be stronger. This occurs when the diamagnetic contribution from closed shell inner electrons (which have been neglected in this free electron gas analysis) is larger than the paramagnetic term of the almost free electrons of the outer shells, such as in Au.

Thus, we have seen that Pauli paramagnetism and Landau diamagnetism are consequences of the spin and the free electron model: the paramagnetism occurs due to intrinsic spin of electrons, whereas diamagnetism arises because of their orbital motion. It should be noted that diamagnetism is always present because such a response does not require any finite magnetic moment to exist. In contrast, a paramagnetic effect can only take place is a net magnetic moment of the atom exists, causing this magnetic moment to align with the external field.

7.4 Exchange interaction

Since the magnitude of the dipole-dipole interactions apparently do not cut it in order to explain magnetic order at high temperatures, we need to look elsewhere to find a stronger interaction between electrons that could account for the observations. It was Heisenberg who in 1928 realized that the Coulomb repulsion between electrons can explain magnetic order at high temperatures. This might seem strange at first since this interaction is spin-independent. The point is that when combined with the intrinsic spin selectivity of quantum mechanics coming from the Pauli principle, stating that two electrons with parallell spins cannot be at the same place at the same time, a spin-dependent strong interaction is produced. Below, we derive how precisely a spin-spin interaction of the type $\boldsymbol{s}_i \cdot \boldsymbol{s}_j$ arises in this way.

The Coulomb interaction reads

$$H_C = \frac{1}{2}\frac{1}{4\pi\epsilon_0}\int d^3r_1 d^3r_2 \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}. \tag{7.32}$$

It has the familiar form of the product of the total electron density (hence sum over spin) at two points $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ divided on the distance between the points. In the second quantized formalism, $\rho(\boldsymbol{r})$ becomes the operator of charge density

$$\rho(\boldsymbol{r}) = -e\sum_\sigma \psi^\dagger_\sigma(\boldsymbol{r})\psi_\sigma(\boldsymbol{r}) \tag{7.33}$$

According to the general result in Chapter 2 for how second quantization of a two-particle operator term in the Hamiltonian should be constructed, the operators must be normal-ordered in the second quantized version of the Hamiltonian:

$$H_C = \frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\int d^3r_1 d^3r_2 \sum_{\sigma\sigma'} \frac{\psi^\dagger_\sigma(\boldsymbol{r}_1)\psi^\dagger_{\sigma'}(\boldsymbol{r}_2)\psi_{\sigma'}(\boldsymbol{r}_2)\psi_\sigma(\boldsymbol{r}_1)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}. \tag{7.34}$$

The second quantized field operators were discussed in the initial chapter of this course and satisfy the usual fermionic anticommutation relations such as $\{\psi_{\sigma_1}(\boldsymbol{r}_1), \psi^\dagger_{\sigma_2})(\boldsymbol{r}_2)\} = \delta_{\sigma_1,\sigma_2}\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$.

The field operators ψ can be expanded into any orthonormal set of single-electron wave functions. Considering a crystal of ions as our model for a solid-state system, it is reasonable to use a set of orbital wavefunctions $\phi_{\boldsymbol{R}}(\boldsymbol{r})$ which are localized around ionic positions $\boldsymbol{R}$ (Wannier functions):

$$\psi_\sigma(\boldsymbol{r}) = \sum_{\boldsymbol{R}} \phi_{\boldsymbol{R}}(\boldsymbol{r}) c_{\boldsymbol{R}\sigma}, \tag{7.35}$$

where $c_{\boldsymbol{R}\sigma}$ are fermion annihilation operators, satisfying the same anticommutation rules as $\psi_\sigma(\boldsymbol{r})$. Inserting this into H_C, we obtain:

$$H_C = \frac{1}{2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2,\boldsymbol{R}_3,\boldsymbol{R}_4} \langle \boldsymbol{R}_1,\boldsymbol{R}_2|\frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}_1-\boldsymbol{r}_2|}|\boldsymbol{R}_3\boldsymbol{R}_4\rangle \times \sum_{\sigma_1\sigma_2} c^\dagger_{\boldsymbol{R}_1\sigma_1}c^\dagger_{\boldsymbol{R}_2\sigma_2}c_{\boldsymbol{R}_3,\sigma_2}c_{\boldsymbol{R}_4,\sigma_1} \tag{7.36}$$

by using the orthonormality of the spinors and we have defined the quantity:

$$\langle \boldsymbol{R}_1,\boldsymbol{R}_2|\frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}_1-\boldsymbol{r}_2|}|\boldsymbol{R}_3\boldsymbol{R}_4\rangle \equiv \int d^3r_1 d^3r_2 \phi^*_{\boldsymbol{R}_1}(\boldsymbol{r}_1)\phi^*_{\boldsymbol{R}_2}(\boldsymbol{r}_2)\frac{e^2}{4\pi|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_{\boldsymbol{R}_3}(\boldsymbol{r}_2)\phi_{\boldsymbol{R}_4}(\boldsymbol{r}_1). \tag{7.37}$$

To make the sum over ionic position more managable, we consider the contribution $\boldsymbol{R}_1 = \boldsymbol{R}_2 = \boldsymbol{R}_3 = \boldsymbol{R}_4$ first, i.e. the on-site contribution of the Coulomb interaction. We then get:

$$H_{C,\text{on-site}} = \frac{1}{2}\sum_{\boldsymbol{R}}\sum_{\sigma_1,\sigma_2} K c^\dagger_{\boldsymbol{R},\sigma_1}c^\dagger_{\boldsymbol{R},\sigma_2}c_{\boldsymbol{R},\sigma_2}c_{\boldsymbol{R},\sigma_1}, \tag{7.38}$$

with

$$K = \int d^3r_1 d^3r_2 |\phi(\boldsymbol{r}_1)|^2 \frac{e^2}{4\pi|\boldsymbol{r}_1-\boldsymbol{r}_2|}|\phi(\boldsymbol{r}_2)|^2. \tag{7.39}$$

Using the fermionic anticommutation relations for the $c_{\boldsymbol{R},\sigma}$-operators and rearranging the operators in $H_{C,\text{on-site}}$, we obtain:

$$H_{C,\text{on-site}} = \frac{1}{2}\sum_{\boldsymbol{R}} K n_{\boldsymbol{R}} n_{\boldsymbol{R}} + \frac{1}{2}\sum_{\boldsymbol{R}} K n_{\boldsymbol{R}}, \tag{7.40}$$

with the number operator $n_{\boldsymbol{R}} = \sum_\sigma c^\dagger_{\boldsymbol{R},\sigma}c_{\boldsymbol{R},\sigma}$. The last term simply corresponds to a change in the reference energy level for each electron (renormalization of the chemical potential) and can be disregarded. The first term is the density-density interaction of the Coulomb repulsion (the so-called Hubbard U-term).

To see how spin-dependent interactions come into play in our model, we now consider the contribution to H_C from different ionic sites $\boldsymbol{R}_i$. As a first-order approximation, we can assume that orbitals $\phi_{\boldsymbol{R}}(\boldsymbol{r})$ at different sites $\boldsymbol{R}$ have negligible overlap. In this case, we only have a contribution to the integral when $\boldsymbol{R}_1 = \boldsymbol{R}_4$ and $\boldsymbol{R}_2 = \boldsymbol{R}_3$ or if $\boldsymbol{R}_1 = \boldsymbol{R}_3$ and $\boldsymbol{R}_2 = \boldsymbol{R}_4$. Inserting this into the Hamiltonian, and using that

$$\sum_{\sigma_1,\sigma_2} c^\dagger_{\boldsymbol{R},\sigma_1}c_{\boldsymbol{R},\sigma_2}c^\dagger_{\boldsymbol{R},\sigma_2}c_{\boldsymbol{R},\sigma_1} = \frac{1}{2}n_{\boldsymbol{R}}n_{\boldsymbol{R}} + 2\boldsymbol{s}_{\boldsymbol{R}}\cdot\boldsymbol{s}_{\boldsymbol{R}} \tag{7.41}$$

where $\boldsymbol{s}^\alpha_{\boldsymbol{R}} = \sum_{\sigma\sigma'} c^\dagger_{\boldsymbol{R},\sigma}(\sigma^\alpha_{\sigma\sigma'})c_{\boldsymbol{R},\sigma'}$, we get:

$$H_{C,\text{exchange}} = \frac{1}{2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2}\Big[(K_{12}-\frac{1}{2}J_{12})n_1n_2 - 2J_{12}\boldsymbol{s}_1\cdot\boldsymbol{s}_2\Big] \tag{7.42}$$

upon defining

$$\begin{aligned} K_{12} \equiv K_{\boldsymbol{R}_1,\boldsymbol{R}_2} &= \int d^3r_1 d^3r_2 |\phi_{\boldsymbol{R}_1}(\boldsymbol{r}_1)|^2 \frac{e^2}{4\pi|\boldsymbol{r}_1-\boldsymbol{r}_2|}|\phi_{\boldsymbol{R}_2}(\boldsymbol{r}_2)|^2, \\ J_{12} \equiv J_{\boldsymbol{R}_1,\boldsymbol{R}_2} &= \int d^3r_1 d^3r_2 \phi^*_{\boldsymbol{R}_1}(\boldsymbol{r}_1)\phi^*_{\boldsymbol{R}_2}(\boldsymbol{r}_2)\frac{e^2}{4\pi|\boldsymbol{r}_1-\boldsymbol{r}_2|}\phi_{\boldsymbol{R}_1}(\boldsymbol{r}_2)\phi_{\boldsymbol{R}_2}(\boldsymbol{r}_1). \end{aligned} \tag{7.43}$$

We see that a spin-dependent interaction has arisen and in fact one may show that $J_{12} \geq 0$. This means that the Coulomb repulsion between electrons in orthogonal orbitals always leads to a ferromagnetic exchange interaction. Physically, one may think of this as follows: by making their spins parallell, the electrons cannot occupy the same position in space and thus avoid strong Coulomb repulsion. This makes their energy lower compared to if their spins were antiparallell. To see that $J_{12} \geq 0$, we use that

$$\frac{1}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} = \int d^3k (2\pi)^3 \mathrm{e}^{i\boldsymbol{k}(\boldsymbol{r}_1 - \boldsymbol{r}_2)} \frac{4\pi}{k^2} \tag{7.44}$$

we get

$$J_{12} = \int d^3k (2\pi)^3 \frac{e^2}{\epsilon_0 k^2} |I(\boldsymbol{k})|^2 \geq 0 \tag{7.45}$$

having defined $I(\boldsymbol{k}) = \int d^3 r_1 \phi^*_{\boldsymbol{R}_1}(\boldsymbol{r}_1) \phi_{\boldsymbol{R}_2}(\boldsymbol{r}_1) \mathrm{e}^{i\boldsymbol{k}\boldsymbol{r}_1}$.

Had we allowed for the possibility of multiple orbitals m per ionic site, i.e.

$$\psi_\sigma(\boldsymbol{r}) = \sum_{\boldsymbol{R},m} \phi_{\boldsymbol{R},m}(\boldsymbol{r}) c_{\boldsymbol{R},m,\sigma}, \tag{7.46}$$

we would also have obtained an exchange integral involving wavefunctions at different orbitals even when just considering one particular ionic site $\boldsymbol{R}_i = \boldsymbol{R}$. The conclusion, similarly to before, is that electrons with parallel spins cannot occupy the same orbital and therefore avoid the strong intra-orbital Coulomb repulsion.

7.5 The Heisenberg model

The above form of the spin-spin interactions can be generalized to the Hamiltonian on a lattice:

$$H = \frac{1}{2} \sum_{i \neq j} J_{ij} \boldsymbol{S}_i \cdot \boldsymbol{S}_j. \tag{7.47}$$

Here, we allow the coupling constant J_{ij} to vary in strength depending on which lattice sites one considers (a homogeneous coupling is achieved by setting $J_{ij} = J$). Moreover, $\boldsymbol{S}_i$ are spin operators defined on lattice sites. They will obey the standard angular momentum commutation relations

$$[S^\alpha_j, S^\beta_j] = i \sum_\gamma \epsilon_{\alpha\beta\gamma} S^\gamma_j, \ (\alpha, \beta, \gamma = x, y, z) \tag{7.48}$$

and spins on different sites commute with each other. The spin operators all have spin S where S in principle can be integer or half-integer. This is a more general model than the spin-spin interaction we obtained for fermions in the previous section, and it is known as the Heisenberg model. The origin of the spin-spin interaction in the fermionic case is the exchange interaction coming from the Coulomb interaction and the Pauli principle.

While any lattice geometry can be considered, we restrict our attention to a hybercubic lattice in d dimension, $d = 1, 2, 3$. This is simply a chain of equidistant atoms in 1D, a square lattice in 2D, and a cubic lattice in 3D. Moreover, we will make the simplifying assumption (which is often realistic) that the spin interactions are negligible between spins that are not nearest-neighbors, i.e.

J_{ij} is nonzero only if i and j are nearest-neighbor lattice sites, in which case we further assume $J_{ij} = J$ where J is a constant, and $J_{ij} = 0$ for any other sites. There are then two different cases to consider, $J < 0$ and $J > 0$. For $J < 0$ the interaction energy of two spins favors them to be parallel; this is the ferromagnetic case. For $J > 0$ antiparallel orientation is instead favored; this is the antiferromagnetic case.

In the previous section, we saw how the Coulomb interaction could give rise to a ferromagnetic spin-spin interaction. We now show how antiferromagnetic spin-spin interaction also can come about from the Coulomb interaction. It turns out that the magnetic properties of many insulating crystals can be quite well described by Heisenberg-type models of interacting spins. Let us consider an example based on assuming that the electrons can be described in terms of the so-called Hubbard model, with Hamiltonian:

$$H = -t \sum_{\langle i,j \rangle \sigma} (c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.}) + U \sum_i n_{i\uparrow} n_{i\downarrow}. \tag{7.49}$$

This is probably the most important (and famous) lattice model of interacting electrons. $c^\dagger_{i\sigma}$ creates an electron on site i with spin σ, and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ counts the number of electrons with spin σ on site i. In this model there is therefore one electronic orbital per site. The first term in Eq. (7.49) is the kinetic energy describing electrons hopping between nearest-neighbor sites i and j, and the second term is the interaction energy describing the energy cost $U > 0$ associated with having two electrons on the same site (these electrons must have opposite spin, as having two electrons on the same site with the same spin would violate the Pauli principle). Note that the interaction energy between electrons which are not on the same site is completely neglected in this model. The Hubbard model is the simplest model de- scribing the fundamental competition between the kinetic energy and the interaction energy of electrons on a lattice. Despite much research, there is still a great deal of controversy about many of its properties in two and three dimensions (the model has been solved exactly in one dimension, but even in that case the solution is extremely complicated).

Rather than giving a strict mathematical derivation (which is definitely possible), we will instead provide a physical picture for how the Hubbard model under specific circumstances gives rise to an antiferromagnetic Heisenberg model, which provides a much stronger qualitative understanding of the underlying physics than a purely technical derivation.

Consider a system with N lattice sites, and assume that there are also N electrons in the system, so the average number of electrons per site is 1. This is called the half-filled case because maximally the system could contain 2 electrons per site (one for each spin projection $\uparrow, \downarrow$) and thus a total electron number of 2N. Assume further that $U \gg t$. This suggests that to find the low-energy states we should first minimize the interaction energy and then treat the kinetic energy as a perturbation. The interaction energy is minimized by putting exactly one electron on each site; then no site is doubly occupied so the total interaction energy is zero. Furthermore, whether the electron on a given site has $\uparrow$ or $\downarrow$-spin is clearly unimportant; thus any spin distribution gives the same interaction energy, leading to a large ($2N$-fold) degeneracy. Moving an electron to a different site containing an electron with opposite spin creates a doubly occupied site which is penalized by a large energy cost U. Thus as long as we're only interested in understanding the physics of the system for energies (and/or temperatures) much less than U, we can neglect configurations with double occupancies. Then, the interaction energy completely determines the charge distribution of the electrons while putting no constraints on their spin distribution.

However, if we now consider the kinetic energy term ($\propto t$) as a perturbation, it is clear that (see Fig. 7.3) if neighboring electrons have opposite spins, an electron can hop (virtually, in the sense of 2nd order perturbation theory) to a neighboring site and back; this virtual delocalization reduces the kinetic energy. The fact that the kinetic energy is reduced by this process can be understood from the formula for the energy correction in 2nd order perturbation theory, which is always negative for the ground state. In contrast (see Fig. 7.3), if neighboring spins are parallel such hopping is forbidden, as the intermediate state with two electrons on the same site with the same spin would violate the Pauli principle. Therefore in this situation an effective interaction is generated which favours neighbouring electrons to have opposite spin, i.e. antiparallel orientation. The resulting effective model, valid at temperatures and energies $\ll U$, is thus the Heisenberg antiferromagnet for the $S = 1/2$ electron spins, with $J = 4t^2/U$. Tthis expression for J can be understood from 2nd order perturbation theory: there is one factor of t for each of the two hops (first to the neighboring site, then back) and a factor U coming from the energy denominator due to the larger energy of the intermediate state. Therefore, the antiferromagnetic exhange interaction J comes about due to an interplay between electron hopping, the electron-electron interactions, and the Pauli principle; the effect of the antiferromagnetic exchange is to reduce the kinetic energy of the electrons.

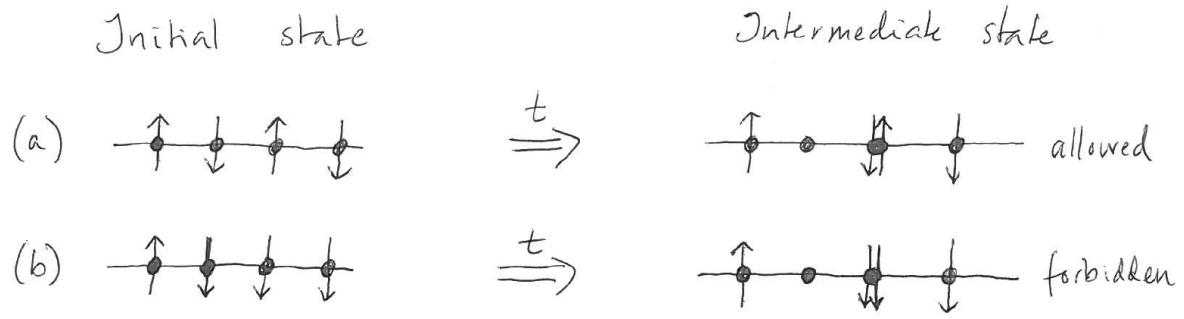

Figure 7.3: Illustration of why an effective antiferromagnetic interaction is generated in the half-filled Hubbard model with $U \gg t$. In both **(a)** and **(b)** the transition shown is between an initial state with an electron on each site and an intermediate state where the electron initially on site 2 has hopped to site 3. See text for further explanation. Figure taken from [6].

Finally, let us discuss the behavior of the spins, as a function of temperature, in materials described by the Heisenberg model. At high temperatures there are strong thermal fluctuations so that the spins are disordered, meaning that the expectation value of each spin vanishes: $\langle \hat{\boldsymbol{S}}_i \rangle = 0$. (Here the brackets represent both a thermal and quantum-mechanical expectation value.) However, below some critical temperature T_c it may be that the spins order magnetically, meaning that the spins on average point in some definite direction in spin space, $\langle \hat{\boldsymbol{S}}_i \rangle \neq 0$. Whether or not such magnetic order occurs depends on the dimensionality and type of lattice, and the range of the interactions (we will limit ourselves to hypercubic lattices and nearest- neighbor interactions in our explicit investigation of the Heisenberg model). If magnetic order occurs with $T_c > 0$, then, as the temperature is lowered from T_c down to zero, $\langle \hat{\boldsymbol{S}}_i \rangle$ will increase and reach some maximum value at zero temperature. The critical temperature T_c is called the Curie temperature (T_C) in ferromagnets and the Neel temperature (T_N) in antiferromagnets. The spin ordering pattern in two magnetically ordered phases are illustrated for a square lattice in Fig. 7.4: In the ferromagnetic case ($J < 0$) all spins point in the same direction while in the antiferromagnetic case ($J > 0$) neighboring spins point in

opposite directions.

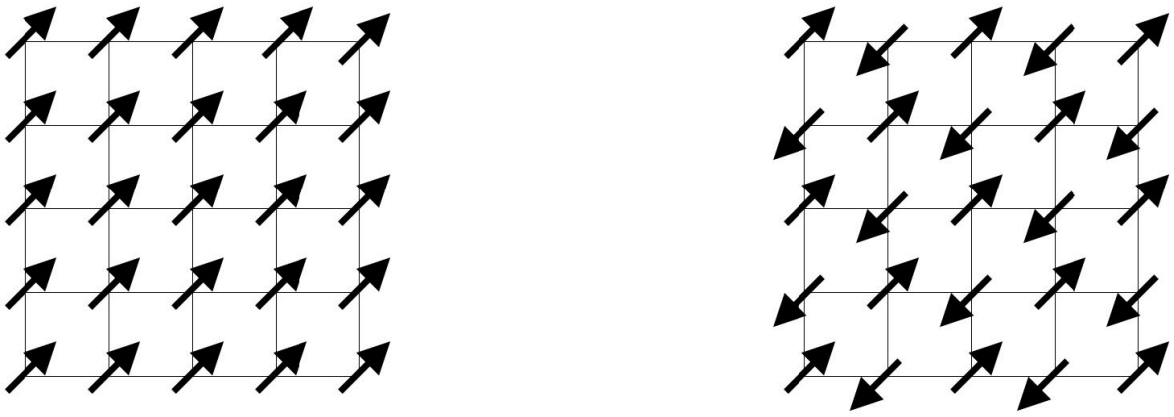

Figure 7.4: Average spin directions in phases with (left) ferromagnetic and (right) antiferromagnetic order on the square lattice. The ordering direction is arbitrary. Figure taken from [6].

7.6 The Holstein-Primakoff representation

Note that the commutator of two spin operators is itself an operator (see Eq. (7.48)), rather than just a complex number (c-number). This makes it much more complicated to work directly with spin operators than with canonical bosonic (fermionic) creation and annihilation op- erators whose commutators (anticommutators) are just c-numbers. It would therefore be advantageous if one could represent the spin operators in terms of such canonical bosonic or fermionic operators and work with these instead. Fortunately quite a few such representations are known. In these lectures we will make use of the so-called Holstein-Primakoff (HP) representation which expresses the spin operators on a site j in terms of canonical boson creation and annihilation operators $a_j^\dagger$ and a_j as follows:

$$\begin{aligned} S_j^+ &= \sqrt{2S - \hat{n}_j}\, a_j, \\ S_j^- &= a_j^\dagger \sqrt{2S - \hat{n}_j}, \\ S_j^z &= S - \hat{n}_j. \end{aligned} \tag{7.50}$$

Here, we have introduced the raising and lowering operators $S_j^\pm = S_j^x \pm i S_j^y$. The operator $\hat{n}_j \equiv a_j^\dagger a_j$ is the number operator for site j, counting the number of bosons on this site. AS the allowed eigenvalues of S_j^z are $-S, -S+1, \ldots$, we see that the boson number must satisfy the constraint

$$\langle \hat{n}_j \rangle \leq 2S. \tag{7.51}$$

Also note that the expressions for S_j^+ and S_j^- are Hermitian conjugates of each other, as they should be. It can be shown that the spin commutation relations Eq. (7.48) and the relation $\boldsymbol{S}^2 = S(S+1)$ follow from the HP representation and the bosonic commutation relations $[a_j, a_j^\dagger] = 1$ etc. As will hopefully become clear in the next couple of sections, the HP representation is very useful for studying magnetically ordered states and their excitations. The reason for this is related to the fact that

the vacuum of the a_j bosons (i.e. the state with no such bosons) is the state corresponding to the maximum eigenvalue S of S_j^z.

Performing the HP transformation on the Heisenberg Hamiltonian, one generally obtains linear terms in $a, a^\dagger$. However, these terms must vanish so long that the we are expanding around the classical ground-state spin configuration. The reason is as follows. The classical ground-state yields the lowest energy. Thus, introducing any additional parameters in the model which change the spin configuration (which is what the HP transformation does) can only increase the energy. Considering then $a, a^\dagger$ as such additional parameters, it is clear that the ground-state of $H(a, a^\dagger)$ is obtained at $\partial H/\partial a = \partial H/\partial a^\dagger = 0$ since this minimizes H with respect to the additional parameters. Therefore, the coefficients associated with the linear terms must vanish, and thus the linear terms vanish all together. In summary, the zeroth-order contribution gives the classical ground-state energy, the terms linear in the boson operators vanish when the classical ground-state is used to define the local spin frames, and the quadratic terms describe the spin-excitations.

7.7 Ferromagnetic Heisenberg model: magnons and Mermin-Wagner theorem

We will now begin our study of the Heisenberg model Eq. (7.47) using spin-wave theory. We will first consider the ferromagnetic case, with $J < 0$ (which can be written $J = -|J|$). It will be convenient to rewrite the Hamiltonian in terms of the raising and lowering operators $S_j^\pm$. Also since we're only considering nearest neighbor interactions we can write $j = i + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a vector connecting nearest-neighbor sites. Actually to avoid counting each interaction twice we will take $\boldsymbol{\delta}$ to run over only half the nearest neighbor vectors. We will consider hypercubic lattices only. Thus in 1D, we take $\boldsymbol{\delta} = +\hat{x}$, in 2D $\boldsymbol{\delta} = +\hat{x}, +\hat{y}$ and so on. This gives

$$H = J \sum_{i,\boldsymbol{\delta}} \Big[\frac{1}{2} (S_i^+ S_{i+\boldsymbol{\delta}}^- + S_i^- S_{i+\boldsymbol{\delta}}^+ + S_i^z S_{i+\boldsymbol{\delta}}^z \Big]. \tag{7.52}$$

The state with all spins pointing along z, i.e. $S_j^z = S$ is a ground state of the Heisenberg ferromagnet. This should be intuitively quite reasonable, and it can be proven easily. First let us show that it is an eigenstate. When we apply Eq. (7.52) to this state, the part involving the ladder operators give exactly zero because the operators S_j^+ kill the ground state since it already has maximum S_j^z which therefore cannot be increased further. Acting with the $S_i^z S_j^z$ gives back an energy $-|J|S^2 Nz/2$ times the same state, where z is the number of nearest neighbors. This shows that this state is an eigenstate. To show that it is also the ground state, we note that the minimal energy possible is given by

$$E_0 = -|J| \sum_{i,\boldsymbol{\delta}} \max \langle \boldsymbol{S}_i \cdot \boldsymbol{S}_{i+\boldsymbol{\delta}} \rangle. \tag{7.53}$$

It can be shown that $\max\langle \boldsymbol{S}_i \cdot \boldsymbol{S}_{i+\boldsymbol{\delta}} \rangle = S^2$, which gives $E_0 = -|J|S^2 Nz/2$, the same as above. Hence, the state in question is indeed a ground state.

We will now use the HP representation to study the Heisenberg ferromagnet, especially its excitations. As noted earlier we will consider a hypercubic lattice in d spatial dimensions, i.e. a standard lattice in one dimension, a square lattice in 2 dimensions, and a cubic lattice in 3 dimensions. The

theory that will be developed is known as spin-wave theory. A natural guess for the low-energy excitations would be that they just correspond to small collective oscillations of the spins around the ordering direction (which we choose to be the z direction). Thus these oscillations, which are called spin waves, make $\langle S_j^z \rangle$ less than the maximum value S. In terms of the HP representation this means that the boson number $\langle \hat{n}_j \rangle$ is nonzero, see Eq. (7.50). If this boson number $\langle \hat{n}_j \rangle$ is much smaller than S the reduction in S_j^z is small (i.e. very weak oscillations) and one might expect that an expansion in a small parameter proportional to $\langle \hat{n}_j \rangle / S$ would make sense. One might expect this to work better the larger S is, since one might guess that increasing S would make this parameter smaller (we'll verify this explicitly later). On the other hand, if it should turn out that $\langle \hat{n}_j \rangle / S$ is not small (this is something we'll have to check at the end of our calculation), then our basic assumption, that spin-waves are just weak oscillations around an ordered state, is wrong or at least questionable, and we may have to conclude that the system is not magnetically ordered after all. For example, this will be the conclusion if $\langle \hat{n}_j \rangle / S$ turns out to be divergent, which we'll see some examples of later.

The above indicates that spin-wave theory is essentially a $1/S$ expansion. It is semi- classical in nature, which follows since the limit $S \to \infty$ corresponds to classical spins, which can be seen e.g. from the fact that the eigenvalues of $\hat{\boldsymbol{S}}_i^2$ are $S(S+1) = S^2(1+1/S)$. If the spins were just classical vectors of length S, the square of their length should be just S^2. Instead we see that there is a correction factor $(1+1/S)$ due to the quantum nature of the spins. As the correction factor goes to 1 in the limit $S \to \infty$, this limit corresponds to classical spins.

7.7.1 Magnons

Let us now discuss the spin-wave theory for the Heisenberg ferromagnet in detail. As noted earlier, in both the quantum and classical ground state all the spins point along the same direction, which we will take to be the z direction. We then rewrite the spin operators in the Heisenberg Hamiltonian in terms of boson operators using the HP representation. We write $\sqrt{2S - \hat{n}_j} = \sqrt{2S}\sqrt{1 - \hat{n}_j/2S}$ and expand the last square root here in a series in the operator $\hat{n}_j/2S$. This gives

$$H = -|J|NS^2 z/2 - |J|S \sum_{i,\boldsymbol{\delta}} [a_{i+\boldsymbol{\delta}}^\dagger a_i + a_j^\dagger a_{i+\boldsymbol{\delta}} - a_i^\dagger a_i - a_{i+\boldsymbol{\delta}}^\dagger a_{i+\boldsymbol{\delta}}] \tag{7.54}$$

Here, N is the total number of lattice sites and z is the number of nearest neighbors and given by $z = 2d$ for a hybercubic lattice. Note that after performing the summations, the two last terms inside the square brackets are in fact identical. We have only included terms of $\mathcal{O}(S)$ and $\mathcal{O}(S^2)$ in the Hamiltonian. To this order, the HP representation reduces to

$$\begin{aligned} S_j^+ &\simeq \sqrt{2S} a_j, \\ S_j^- &\simeq \sqrt{2S} a_j^\dagger, \\ S_j^z &= S - a_j^\dagger a_j. \end{aligned} \tag{7.55}$$

The term linear in S is quadratic (also said to be bilinear) in the boson operators and can be diagonalized. As it is quadratic, it corresponds to noninteracting bosons. Terms in the Hamiltonian which are higher order in the $1/S$ expansion (not shown) contain four or more boson operators and thus represent interactions between bosons. However, these are suppressed at least by a factor $1/S$ compared to the $\mathcal{O}(S)$ noninteracting term and one can thus hope that their effects are small (at least at large S and when the boson number is small) so that they can either be neglected to a first

approximation or be treated as weak perturbations on the noninteracting theory.

We now proceed to diagonalize the Hamilton-operator. In the ferromagnetic case, this can be accomplished by first introducing Fourier-transformed boson operators as follows:

$$a_{\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_i \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}_i} a_j. \tag{7.56}$$

This is just a variable transformation so that there as many operators $a_{\boldsymbol{k}}$ as there are operators a_i. The inverse transformation is

$$a_i = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{r}_i} a_{\boldsymbol{k}}. \tag{7.57}$$

By using periodic boundary conditions, we impose specific values that $\boldsymbol{k}$ can take. Such boundary conditions imply that for instance $a_i = a_{i+N_x\hat{x}}$ where N_x is the number of sites in the x-direction. This is satisfied if $\mathrm{e}^{ik_xN_x} = 1$, so that $k_x = 2\pi n_x/N_x$ where n_x is an integer. It is customary to choose n_x to take N_x successive values

$$-\frac{N_x}{2}, -\frac{N_x}{2}+1, \ldots \frac{N_x}{2}-1. \tag{7.58}$$

For simplicity, we assumed here that N_x is even so that $N_x/2$ is in fact an integer, but one can choose such a series of values even if N_x is odd. Then, k_x takes values in the interval $[-\pi, \pi\rangle$. Doing the same for all directions, the resulting $\boldsymbol{k}$ values lie within the first BZ. An important aspect of the above transformation of the a-operators is that it preserves the commutation relations, so that $[a_{\boldsymbol{k}}, a^\dagger_{\boldsymbol{k}'}] = \delta_{\boldsymbol{k},\boldsymbol{k}'}$ and so on. Using that $\sum_i \mathrm{e}^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}_i} = N\delta_{\boldsymbol{k},\boldsymbol{k}'}$ (see exercise for proof), one obtains

$$H = E_0 + \sum_{\boldsymbol{k}} \omega_k a^\dagger_{\boldsymbol{k}} a_{\boldsymbol{k}}, \tag{7.59}$$

where

$$E_0 = -|J|NS^2z/2,\ \omega_{\boldsymbol{k}} = 2|J|S\sum_{\boldsymbol{\delta}}(1-\cos\boldsymbol{k}\cdot\boldsymbol{\delta}). \tag{7.60}$$

This equation describes a Hamiltonian which is simply a bunch of independent harmonic oscillators, each labelled by a wavevector $\boldsymbol{k}$. The quanta of the harmonic oscillators are called magnons: they are the quantized spin wave excitations (just like phonons are the quantized lattice vibrations in a crystal) with energy $\omega_{\boldsymbol{k}}$. It is also referred to as a free magnon gas. In the limit $\boldsymbol{k} \to 0$, we have

$$\omega_{\boldsymbol{k}} \simeq |J|S|\boldsymbol{k}|^2. \tag{7.61}$$

As a magnon with wavevector $\boldsymbol{k}$ costs an energy $\omega_{\boldsymbol{k}} > 0$, the ground state has no magnons, i.e. $\langle\hat{n}_{\boldsymbol{k}}\rangle = 0$ at zero temperature. The ground state energy is therefore simply E_0 which is the interaction energy of all spins pointing in the same direction with maximal projection S along the z axis. As the temperature is increased, magnons will be thermally excited. Since they are just noninteracting bosons, the mean number of magnons with momentum $\boldsymbol{k}$ is given by the Bose-Einstein distribution function

$$\langle\hat{n}_{\boldsymbol{k}}\rangle = \frac{1}{\mathrm{e}^{\beta\omega_{\boldsymbol{k}}}-1}. \tag{7.62}$$

The magnetization $\boldsymbol{M} = \frac{1}{N}\sum_j \langle \boldsymbol{S}_i \rangle$ is a natural measure of the strength of the amgnetic order in the system. If $M \equiv |\boldsymbol{M}|$ is positive (zero) we say that the system is (is not) ferromagnetically ordered. The larger M, the stronger ferromagnetic order. We say that the magnetization is an order parameter for the ferromagnetic phase. By definition, as we have seen earlier, an order parameter for a given type of order is a quantity that is nonzero in the phase(s) where that order is present and is zero in other phases. With the ordering direction being the z direction, we get

$$M = \frac{1}{N}\sum_i \langle S_i^z \rangle = S - \frac{1}{N}\sum_{\boldsymbol{k}} \langle \hat{n}_{\boldsymbol{k}} \rangle \equiv S - \Delta M. \tag{7.63}$$

Let us compute ΔM at low temperatures. First, introduce an artificial wavevector cutoff k_0 which is the smallest wavevector in the $\boldsymbol{k}$-sum: the real system is then obtained by taking $k_0 \to 0$. We also introduce another wavevector $\bar{k} > k_0$ which is defined by $\omega_{\bar{\boldsymbol{k}}} \ll k_B T \ll |J|S$. This means in particular that for $|k| < \bar{k}$, the quadratic form of the dispersion is valid. This gives

$$\Delta M = \frac{1}{N}\Big(\sum_{k_0 < |\boldsymbol{k}| < \bar{k}} \frac{1}{\mathrm{e}^{|J|Sk^2/k_B T} - 1} + \sum_{|\boldsymbol{k}| > \bar{k}} \frac{1}{\mathrm{e}^{\omega_{\boldsymbol{k}}/k_B T} - 1} \Big). \tag{7.64}$$

The second term is independent on k_0 and finite. However, we may neglect it because the first term turns out to be infinite! To see this, convert the sum to an integral and expand the exponential (using $|J|S^2k^2 \ll k_B T$) to get

$$\Delta M \propto \int_{k_0}^{\bar{k}} dk\; k^{d-1} \frac{k_B T}{|J|Sk^2} \propto \frac{k_B T}{|J|S} \cdot \begin{cases} \frac{1}{k_0} + \ldots, & d = 1 \\ -\log k_0 + \ldots, & d = 2 \end{cases} \tag{7.65}$$

Therefore we see that, at nonzero temperatures in one and two dimensions, ΔM diverges as the cutoff k_0 is sent to zero. Therefore our initial assumption that this correction is small, is found to be wrong for these cases (note that the quantity $\Delta M/2S$ is the expectation value of the average over all sites of our original expansion parameter $\hat{n}_j/2S$, which we assumed to be small when we expanded the square roots in the HP expression). Thus we conclude that $M = 0$ (i.e. there is no ferromagnetic order) at finite temperatures for the Heisenberg model in one and two dimensions. For the case of three dimensions, it can be shown that $\Delta M \propto T^{3/2}$ as $T \to 0$. Thus, spin-wave theory predicts that ferromagnetic order is stable, in effect $M > 0$, at sufficiently low temperatures in three dimensions (see one of the exercises for derivation).

7.7.2 Mermin-Wagner theorem

In quantum field theory and statistical mechanics, there exists a famous theorem which provides limitations on when long-range order, such as a ferromagnetic phase, can exist. This is known as the Mermin-Wagner theorem [Mermin and Wagner, Phys. Rev. Lett. **17**, 1133 (1966)]:

At any non-zero temperature, there can be no long-range order coming from the spontaneous breaking of a continuous symmetry in a one- or two-dimensional system with isotropic and sufficiently short-ranged interactions.

This theorem indicates that long-range fluctuations can be generated with little energy, thus lowering the entropy of the free energy $F = U - TS$ of the system, and destroying any order. This is exactly what we saw happen due to the spin-waves above for $d = 1$ and $d = 2$ dimensions. In

the ferromagnetic case, we are spontaneously breaking the spin-rotational symmetry of the Hamiltonian, causing a long-range ordering of the electron spins.

The situation changes if we add a magnetic easy-axis to the system. This means that there exists a direction in space along which it is most favorable for the spins to point. The physical origin of such a magnetic easy-axis is spin-orbit interactions. This causes an anisotropy in the Hamiltonian, described by an additional term $-\sum_i K S_{z,i}^2$. The origin of a favored magnetization direction can be understood by noting the the electron spins in the atomic orbitals will interact with the crystal structure via spin-orbit interactions in the following way. The electron spin moments are coupled to the spatial distribution of the electron cloud (the orbitals) via spin-orbit interactions $\propto \boldsymbol{S} \cdot \boldsymbol{L}$ where $\boldsymbol{L}$ is the orbital angular momentum of the electron. If we try to rotate the electron spins with a magnetic field, the orbitals will thus try to follow them by also rotating in space, as shown in Fig. 7.5. The orientation of the orbitals relative neighboring orbitals will now determine the energy cost of this. If the orbitals overlap along a certain direction, an additional energy cost can be incurred due to Coulomb repulsion. Therefore, some orientations of the magnetization caused by the electrons are favorable. This produces so-called easy-axes in a magnet.

Put differently, the electron spin orientation affects the orientation of the electronic wave-functions due to spin-orbit interactions. Since different electronic wavefunctions have different Coulomb interaction, the spin orientation has an associated energy.

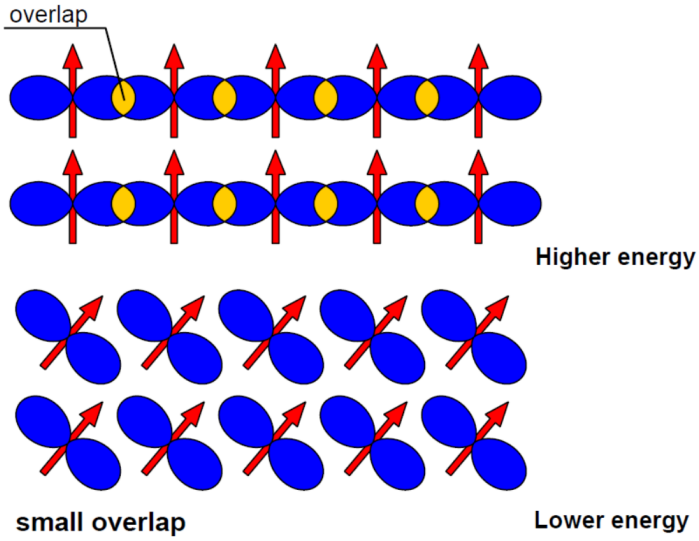

Figure 7.5: Illustration of how orbital overlap determines a preferred magnetization orientation in a ferromagnetic crystal. Figure taken from `https://magnetism.eu/esm/2023/slides/esm2023-iglesias-slides-1.pdf`.

The presence of an anisotropy breaks the assumption of isotropic interactions in the Mermin-Wagner theorem. The system is no longer isotropic: the z-axis is special in the example above. Inserting the HP-transformation into such a term can straightforwardly be seen to introduce a gap $\propto K$ in the dispersion relation $\omega_{\boldsymbol{k}}$ for the magnons, so that $\omega_{\boldsymbol{k}=0} \propto K$. With such a dispersion, magnetic order is stabilized even in $d = 1$ or $d = 2$.

Another way to break the Mermin-Wagner theorem is by making the interactions sufficiently long-ranged. The precise mathematical criterion for 'sufficiently' can be found elsewhere. Here, we merely give an example. For instance, had we assumed a longer-range interaction between the spins in our model of a ferromagnet (coming *e.g.* from dipole-dipole interactions), the dispersion relation $\omega_{\boldsymbol{k}}$ would have changed. In turn, this could remove the divergence coming from the lower integration limit (k_0) in the magnetization integral, again providing a finite magnetization even at zero temperature.

7.8 Antiferromagnetic Heisenberg model and magnons

We next turn to the antiferromagnetic case $J > 0$ of the Heisenberg model. As the ground state for classical spins has the spins on neighbouring sites pointing in opposite directions, one might naively guess that the ground state in the quantum case is analogous, thus having maximal and opposite spin projections $\pm S$ on neighboring sites. Such a state can be written

$$\prod_{j\in A} |S\rangle_j \prod_{l\in B} |-S\rangle_l \equiv |N\rangle. \tag{7.66}$$

Here A and B denote the two sublattices such that the spins on A (B) sites have spin projection S ($-S$), i.e. the states $|\pm S\rangle$ are eigenstates of S_z for the given lattice site with eigenvalue $\pm S$. It is however easy to see that $|N\rangle$ can not be the ground state, and is in fact not even an eigenstate, for the Heisenberg model for finite values of S. Acting with the Heisenberg H on this state, the quantum fluctuation terms involving the spin raising and lowering operators change the state so that $H|N\rangle$ is not proportional to $|N\rangle$. Note that this did not happen in the ferromagnetic case because then the S^+ operators always killed the ferromagnetic ground state, leaving only the contribution from the S^zS^z part of the Hamiltonian. Consequently, quantum uctuations play a much more important role in the antiferromagnetic case, as they change the ground state (and its energy) away from the "classical" result. Although the ground state is not given by $|N\rangle$, it may still be that the ground state has antiferromagnetic order, i.e. that the spins on sublattice A point predominantly in one direction (taken to be the z direction here) and the spins on sublattice B point predominantly in the opposite direction. (If so the state $|N\rangle$ captures the structure of the true ground state at least in a qualitative sense.) To investigate this possibility we again develop a spin-wave theory based on expanding the square roots in the HP expansion. On the A sublattice where the spin projection is $+S$ we use the standard expressions:

$$\begin{aligned} S^+_{Aj} &= \sqrt{2S - a_j^\dagger a_j}\, a_j, \\ S^-_{Aj} &= a_j^\dagger \sqrt{2S - a_j^\dagger a_j}, \\ S^z_{Aj} &= S - a_j^\dagger a_j. \end{aligned} \tag{7.67}$$

However, on the B sublattice where the spin projection is $-S$, we must modify the HP expressions according to reflect this:

$$\begin{aligned} S^+_{Bj} &= b_j^\dagger \sqrt{2S - b_j^\dagger b_j}, \\ S^-_{Bj} &= \sqrt{2S - b_j^\dagger b_j}\, b_j, \\ S^z_{Bj} &= -S + b_j^\dagger b_j. \end{aligned} \tag{7.68}$$

Compared to the expressions on the A sublattice, these modified expressions correspond to the changes $S^+ \leftrightarrow S^-$, $S^z \to -S^z$, which preserve the commutation relations, which shows that the HP expressions for sublattice B are indeed correct. Note that different boson operators a_j and b_j have been introduced for sublattices A and B respectively. The indices j run over the sites in A and B respectively. Inserting the HP expressions in the Hamiltonian, expanding the square roots and keeping terms to order S in the Hamiltonian, we get

$$\begin{aligned} H &= J \sum_{j\in A,\boldsymbol{\delta}} [S(a_j b_{j+\boldsymbol{\delta}} + \text{h.c.}) + S(a_j^\dagger a_j + b_{j+\boldsymbol{\delta}}^\dagger b_{j+\boldsymbol{\delta}}) - S^2] \\ &+ J \sum_{j\in B,\boldsymbol{\delta}} [S(b_j a_{j+\boldsymbol{\delta}} + \text{h.c.}) + S(b_j^\dagger b_j + a_{j+\boldsymbol{\delta}}^\dagger a_{j+\boldsymbol{\delta}}) - S^2]. \end{aligned} \tag{7.69}$$

Introducing Fourier-transformed operators we get

$$\begin{aligned} a_{\boldsymbol{k}} &= \frac{1}{\sqrt{N_A}} \sum_{j\in A} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}_j} a_j, \\ b_{\boldsymbol{k}} &= \frac{1}{\sqrt{N_B}} \sum_{j\in B} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{r}_j} a_b \end{aligned} \tag{7.70}$$

where $N_A = N_B = N/2$ is the number of lattice sites in each sublattice.

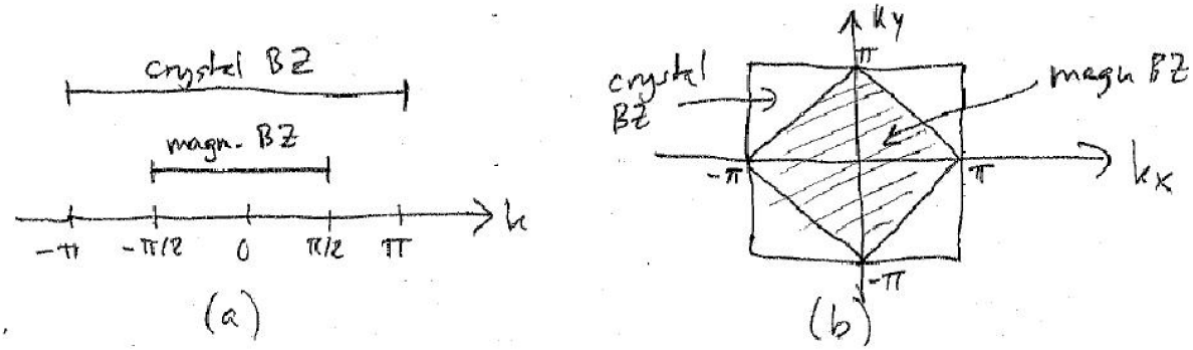

Figure 7.6: Crystal Brillouin zone and magnetic Brillouin zone in **(a)** one dimension and **(b)** two dimensions. Figure taken from [6].

The commutation relations are standard bosonic, i.e. the only nonzero commutators are $[a_{\boldsymbol{k}}, a_{\boldsymbol{k}'}^\dagger] = [b_{\boldsymbol{k}}, b_{\boldsymbol{k}'}^\dagger] = \delta_{\boldsymbol{k},\boldsymbol{k}'}$. We choose the $\boldsymbol{k}$-vectors to lie in the Brillouin zone associated with each sublattice. (Since the two sublattices are identical, their Brillouin zone is also identical). This Brillouin zone is called the magnetic Brillouin zone to distinguish it from the Brillouin zone associated with the full lattice which is called the crystal Brillouin zone. In one dimension periodic boundary conditions gives $a_{j+N_a\cdot 2\hat{x}} = a_j$ (the factor of 2 comes from the fact that the spacing between neigboring sites in the sublattice is 2, not 1) which gives $k_x = 2\pi n_x/(2N_A)$. Choosing the N_A values of n_x which lie closest to 0 (i.e. $n_x = -N_A/2, \ldots N_A/2 - 1$) then gives $k_x \int [\pi/2, \pi/2\rangle$. Thus the length of the magnetic Brillouin zone is half the length of the crystal Brillouin zone $|\pi, \pi\rangle$. In two dimensions the two sublattices are oriented at a 45 degree angle with respect to the full lattice and have a lattice

spacing which is $\sqrt{2}$ larger. It follows from this that the magnetic Brillouin zone is a square oriented at a 45 degree angle with respect to the crystal Brillouin zone and has half its area. These facts are illustrated in Fig. 7.6. To distinguish $\boldsymbol{k}$-sums over the full and magnetic BZ, we will use $\boldsymbol{k} \in \square$ and $\boldsymbol{k} \in \diamond$, respectively.

Using $\sum_{j\in A} \mathrm{e}^{i(\boldsymbol{k}-\boldsymbol{k}')'\cdot \boldsymbol{r}_j} = N_A \delta_{\boldsymbol{k},\boldsymbol{k}'}$ and inserting our Fourier-transformed operators, we obtain

$$H = -NJS^2 z/2 + JSz \sum \boldsymbol{k} \in \diamond [\gamma_{\boldsymbol{k}} (a_{\boldsymbol{k}} b_{-\boldsymbol{k}} + a^\dagger_{\boldsymbol{k}} b^\dagger_{-\boldsymbol{k}}) + a^\dagger_{\boldsymbol{k}} a_{\boldsymbol{k}} + b^\dagger_{\boldsymbol{k}} b_{\boldsymbol{k}}] \tag{7.71}$$

where we defined

$$\gamma_{\boldsymbol{k}} = \frac{2}{z} \sum_{\boldsymbol{\delta}} \cos(\boldsymbol{k}\cdot\boldsymbol{\delta}). \tag{7.72}$$

This Hamiltonian can now be diagonalized (see one of the exercises) and one arrives at the following Hamiltonian describing a free magnon gas with two species of magnons α and β:

$$H = H_0 + \sum_{\boldsymbol{k}\in\diamond} \omega_{\boldsymbol{k}} (\alpha^\dagger_{\boldsymbol{k}} \alpha_{\boldsymbol{k}} + \beta^\dagger_{\boldsymbol{k}} \beta_{\boldsymbol{k}}). \tag{7.73}$$

Here, we have defined the magnonic dispersion relation

$$\omega_k = JSz\sqrt{1-\gamma^2_{\boldsymbol{k}}} \tag{7.74}$$

and the ground-state energy

$$E_0 = -NJS^2 z/2 - NJSz/2 + \sum_{\boldsymbol{k}\in\diamond} \omega_{\boldsymbol{k}} \tag{7.75}$$

The operators $\alpha^\dagger_{\boldsymbol{k}}$ and $\beta^\dagger_{\boldsymbol{k}}$ create magnon excitations with wavevector $\boldsymbol{k}$ and energy $\omega_{\boldsymbol{k}}$. These magnons are bosons, and are noninteracting in the approximation used here (i.e. when only including terms of $\mathcal{O}(S)$ and $\mathcal{O}(S^2)$ in the Hamiltonian). Note that in this antiferromagnetic case, for each $\boldsymbol{k}$ there are two types of magnons which are degenerate in energy. On the other hand, the $\boldsymbol{k}$ sum goes over the magnetic Brillouin zone which only has $N/2$ $\boldsymbol{k}$-vectors, so the total number of magnon modes is $2\cdot N/2 = N$, the same as for the ferromagnetic case. Note that as $\boldsymbol{k} \to 0, \omega_{\boldsymbol{k}} \to 0$ as in the ferromagnetic case, but unlike the ferromagnetic case, for which a quadratic dispersion $\omega_{\boldsymbol{k}} \propto |\boldsymbol{k}|^2$ was found in this limit, in the antiferromagnetic case we have instead a linear dispersion,

$$\omega_{\boldsymbol{k}} \propto |\boldsymbol{k}| \text{ as } \boldsymbol{k} \to 0. \tag{7.76}$$

In the ground state of H, let's call it $|G\rangle$, there are neither α or β magnons since they cost an energy $\omega_{\boldsymbol{k}} > 0$. Thus, we can define the ground state by the relations

$$\alpha_{\boldsymbol{k}}|G\rangle = 0,\ \beta_{\boldsymbol{k}}|G\rangle = 0 \text{ for all } \boldsymbol{k}. \tag{7.77}$$

This gives $H|G\rangle = E_0|G\rangle$. The first term in E_0, which is $\propto S^2$, is just the ground state energy of a classical nearest-neighbor antiferromagnet of spins with length S. The other terms, which are $\propto S$, represent quantum corrections to the classical ground-state energy. The resulting correction ΔE is actually negative:

$$\Delta E = \sum_{\boldsymbol{k}} \omega_{\boldsymbol{k}} - NJSz/2 = JSz \sum_{\boldsymbol{k}} [\sqrt{1-\gamma^2_{\boldsymbol{k}}} - 1] < 0 \tag{7.78}$$

Thus, quantum fluctuations lower the energy of the system. Note that $\Delta E = 0$ in the ferromagnetic case, so there are no fluctuations in the FM ground state.

Let us next investigate the amount of magnetic order in the system. Thus we need to identify an order parameter for antiferromagnetic order. Note that the magnetization $\boldsymbol{M} = (1/N)\sum_i \langle \boldsymbol{S}_i \rangle$ can not be used since it is zero in the presence of antiferromagnetic order, because the two sublattices give equal-magnitude but opposite-sign contributions to $\boldsymbol{M}$. Instead the natural order parameter is the so-called sublattice magnetization, defined by averaging $\langle \boldsymbol{S}_i \rangle$ only over the sites of one of the two sublattices. Without loss of generality, let us pick sublattice A, where the putative ordering is in the z direction. The magnitude of the sublattice magnetization is thus

$$M_A = \frac{1}{N_A}\sum_{j\in A}\langle S_j^z \rangle = S - \frac{1}{N_A}\sum_{\boldsymbol{k}\in\diamond}\langle a_{\boldsymbol{k}}^\dagger a_{\boldsymbol{k}} \rangle. \tag{7.79}$$

The correction ΔM_A to the classical result S is therefore

$$\begin{aligned}\Delta M_A &= \frac{1}{N_A}\sum_{\boldsymbol{k}\in\diamond}\langle a_{\boldsymbol{k}}^\dagger a_{\boldsymbol{k}} \rangle \\ &= \frac{2}{N_A}\sum_{\boldsymbol{k}\in\diamond}[u_{\boldsymbol{k}}^2\langle \alpha_{\boldsymbol{k}}^\dagger \alpha_{\boldsymbol{k}} \rangle + v_{\boldsymbol{k}}^\dagger \langle \beta_{-\boldsymbol{k}}\beta_{-\boldsymbol{k}}^\dagger + u_{\boldsymbol{k}}v_{\boldsymbol{k}}\langle \alpha_{\boldsymbol{k}}\beta_{-\boldsymbol{k}} + \text{h.c.}\rangle].\end{aligned} \tag{7.80}$$

These expectation values are both thermal and quantum, and are evaluated (see one of the exercises) to yield

$$\Delta M_A = -\frac{1}{2} + \frac{2}{N_A}\sum_{\boldsymbol{k}\in\diamond}\Big(n_{\boldsymbol{k}} + \frac{1}{2}\Big)\frac{1}{\sqrt{1-\gamma_{\boldsymbol{k}}^2}}. \tag{7.81}$$

This expression has a temperature-dependent part coming from nk (note that $n_{\boldsymbol{k}} = 0$ at $T = 0$) and a temperature-independent part coming from the two terms containing the factor $1/2$. Let us first consider the case of zero temperature. In one dimension the $\boldsymbol{k}$-sum (note that $\gamma_k = \cos k$ in one dimension) becomes:

$$\frac{1}{N_A}\sum_k \frac{1}{\sqrt{1-\gamma_k^2}} \propto \lim_{k_0\to 0}\int_{k_0}^{\pi/2}\frac{dk}{\sin k}. \tag{7.82}$$

The most important contribution to this integral comes from the small-k region where we can approximate $\sin k \simeq k$. Thus the leading term becomes $\int_{k_0} dk/k = -\log(k_0) \to \infty$ for $k_0 \to 0$. Thus ΔM_A diverges even at zero temperature. We must therefore conclude that for this one-dimensional case our assumption that the system was magnetically ordered is invalid and so is our truncated spin-wave expansion. Note that this conclusion holds for any S. In two dimensions at zero temperature ΔM_A is still nonzero so quantum uctuations do reduce the magnetization, but the correction turns out to be small enough, $\Delta M_A \simeq 0.2$ so that even for the lowest spin, $S = 1/2$, spin-wave theory indicates that the system is ordered at zero temperature for a square lattice. This is also in agreement with other methods. In three dimensions ΔM_A is even smaller so the order is more robust then. Next we consider finite nonzero temperatures. We just summarize the results that are ob- tained by analyzing the above expression for ΔM_A (which, we stress, is valid for a hypercubic lattice). In one dimension there is of course no antiferromagnetic order since none existed even at zero temperature. In two dimensions it turns out that the order does not survive at finite temperatures, so the spin-wave approach again is invalid. In three dimensions the system is ordered at

sufficiently low temperatures.

In Fig. 7.7 below we summarize the main conclusions we have obtained from applying spin-wave theory to the Heisenberg ferromagnet and antiferromagnet on a hypercubic lattice in d spatial dimensions ($d = 1, 2, 3$). "Ordered" means spin-wave theory predicts magnetic ordering of the ferromagnetic/antiferromagnetic type, "disordered" means spin-wave theory predicts the absence of such order.

	Ferromagnet	Antiferromagnet
$d = 1$, $T = 0$	Ordered	Disordered
$d = 1$, $T > 0$	Disordered	Disordered
$d = 2$, $T = 0$	Ordered	Ordered
$d = 2$, $T > 0$	Disordered	Disordered
$d = 3$, $T = 0$	Ordered	Ordered
$d = 3$, $T > 0$	Ordered (at low T)	Ordered (at low T)

Figure 7.7: Summary of magnetic order at zero and finite temperature for different dimensionality for both FM and AFM order. Figure taken from [6].

7.8.1 Experimental measurement of magnon dispersion

One of the most prominent ways to characterize magnetic materials is neutron scattering. This is made possible due to the fact that neutrons have spin and a magnetic moment [1] This allows neutrons to investigate the magnetic structure of materials, including magnetic excitations. The energy lost in the neutron beam is left in the sample, as shown in Fig. 7.8, and can excite spin waves.

By measuring the intensity of neutrons at wavenumbers shifted from the initial neutrons, the magnon dispersion can be reconstructed. More specifically, the scattering cross section of the neutrons for scattering on phonons and magnons with a given dispersion relation can be derived analytically. One then compares the experimentally observed differential scattering cross section and its angular dependence with this analytical expression. This provides information about how the phonon or magnon dispersion must look like in order to match the experimental data. One can distinguish between the contribution from phonons and magnons by looking at for instance how the scattering cross section depends on the momentum transfer $\boldsymbol{Q} = \boldsymbol{k}_f - \boldsymbol{k}_i$ of the neutrons. This dependence is different for phonons and magnons. The magnitude of the inelastic scattering cross section of the neutrons can be of similar magnitude for phonon and magnon scattering.

7.9 Magnetic resonance

In this chapter we discuss dynamical magnetic effects associated with the spin angular momentum of nuclei and of electrons, rather than static phenomena and ordering considered in the previous sections. The principal phenomena are often identified in the literature by their initial letters, such as

[1] The neutrons have zero net charge but still feature a magnetic moment due to its internal charge-structure from the quarks.

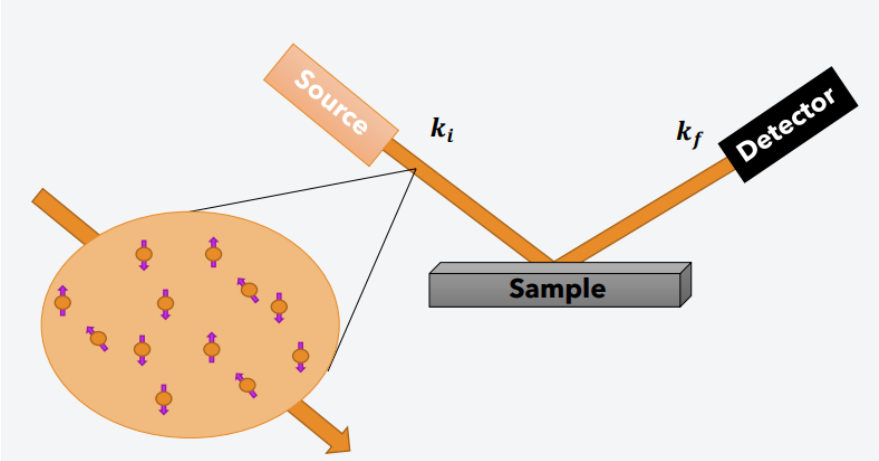

Figure 7.8: Bombarding a sample with neutrons, which carry magnetic moment, allows them to interact with magnetic excitations in the material due to their dipole-dipole interaction. By comparing the outgoing neutron momentum with the incoming momentum, one infers information about the magnetic excitations in the sample that the neutrons have scattered on. Figure taken from [13].

- NMR: nuclear magnetic resonance
- EPR: electron paramagnetic resonance
- FMR: ferromagnetic resonance
- AFMR: antiferromagnetic resonance

Resonance studies provide information about the internal effective magnetic field in a sample (which can have a lot of different physical contributions), mechanisms providing relaxation of magnetization dynamics, and collective spin excitations in a material.

It is best to discuss NMR as a basis for a brief account of the other resonance experiments. A great impact of NMR has been in organic chemistry and biochemistry, where NMR provides a powerful tool for the identification and the stnicture determination of complex molecules. A major medical application of NMK is magnetic resonance imaging (MRI), which allows the resolution in 3D of abnormal growths, configurations, and reactions in the whole body.

7.9.1 Nuclear magnetic resonance

Consider a nucleus that has a magnetic moment $\boldsymbol{\mu}$ and an angular momentum $\hbar \boldsymbol{I}$. They are related by

$$\boldsymbol{\mu} = \gamma \hbar \boldsymbol{I} \tag{7.83}$$

where the gyromagnetic ratio $\gamma < 0$ is constant. The energy of interaction with an applied magnetic field $\boldsymbol{B}_a$ is

$$U = -\boldsymbol{\mu} \cdot \boldsymbol{B}_a. \tag{7.84}$$

For $\boldsymbol{B}_a = B_0\hat{\boldsymbol{z}}$, we get

$$U = -\gamma\hbar B_0 I_z \tag{7.85}$$

and the allowed values for I_z are $m_I = I, I-1, \ldots -I$. In a magnetic field of a nucleus with $I = 1/2$ there are two energy levels corresponding to $m_I = \pm 1/2$. Let $\hbar\omega_0$ denote the energy difference between the two levels, from which it follows that $\omega_0 = \gamma B_0$. This will turn out to be the fundamental condition for magnetic resonance absorption: when the frequency of an applied ac radiation field matches ω_0, a resonance occurs.

For the proton, $\gamma = 2.675 \times 10^8$ s^{-1} T^{-1}, giving

$$\nu(\text{MHz}) = 42.58 B_0(\text{Tesla}) \tag{7.86}$$

where ν is the frequency of the applied radiation field. For the electron spin, one instead obtains

$$\nu(\text{GHz}) = 28.0 B_0(\text{Tesla}) \tag{7.87}$$

Equations of motion

The rate of change of angular momentum of a system is equal to the torque that acts on the system. The torque on a magnetic moment $\boldsymbol{\mu}$ in a magnetic field $\boldsymbol{B}$ is $\boldsymbol{\mu} \times \boldsymbol{B}$, so that we get the equation

$$\hbar d\boldsymbol{I}/dt = \boldsymbol{\mu} \times \boldsymbol{B}_a \tag{7.88}$$

which can be rewritten to

$$d\boldsymbol{\mu}/dt = \gamma\boldsymbol{\mu} \times \boldsymbol{B}_a. \tag{7.89}$$

The nuclear magnetization $\boldsymbol{M}$ is the sum $\sum_i \boldsymbol{\mu}_i$ over all the nuclei in a unit volume. Thus,

$$d\boldsymbol{M}/dt = \gamma\boldsymbol{M} \times \boldsymbol{B}_a \tag{7.90}$$

Note that the same equation would apply to the magnetization from other sources, like from polarized itinerant electrons or localized polarized electrons in orbitals. Let $\boldsymbol{B} = B_0\hat{\boldsymbol{z}}$ be a static field. In tehrmal equilibrium, the magnetization will point along $\hat{\boldsymbol{z}}$: $M_z = M_0$ where M_0 is a material-specific constant.

The magnetization of a system of spins with $I = 1/2$ originates from the population difference $N_1 - N_2$ of the lower and upper energy levels, so that $M_z = (N_1 - N_2)\mu$ where the N's refer to the population in a unit volume. The population ratio in thermal equilibrium is provided by the Boltzmann factor from statistical mechanics [2] for the energy difference $2\mu B_0$:

$$\left(\frac{N_2}{N_1}\right)_0 = \exp(-2\mu B_0/k_B T) \tag{7.91}$$

so that the equilibrium magnetization is $M_0 \propto \mu \cdot \tanh(\mu B/k_B T)$. When the magnetization M_z is not in thermal equilibrium, it is reasonable to suppose that it over time will approach its equilibrium value at a rate proportional to the departure from M_0:

$$\frac{dM_z}{dt} = \frac{M_0 - M_z}{T_1}. \tag{7.92}$$

[2] Boltzmann postulated that in a system with a macroscopic number of particles, the probability that a certain energy level should be occupied is a function of the temperature and energy of the level according to $\mathrm{e}^{-\beta E_i}/Z$ where Z is the partition function. Thus, the ratio of probabilities is governed by the energy difference between two states.

It is standard to refer to T_1 as the longitudinal relaxation time or the spin-lattice relaxation time. If an unmagnetized speciment is placed in a magnetic field $B_0\hat{\boldsymbol{z}}$ at $t = 0$, then using this as a boundary condition for the above equation of motion provides the solution

$$M_z(t) = M_0[1 - \mathrm{e}^{-t/T_1}]. \tag{7.93}$$

The physical origin of the relaxation is the interaction of the magnetization with other degrees of freedom in the system, such as phonons, causing it to relax back to its ground-state. Typical processes whereby the magnetization approaches equilibrium are shown in Fig. 7.9. The dominant spin-lattice interactions of paramagnetic ions in crystals is by the phonon modulation of the crystalline electric field. Relaxation proceeds by three principal processes: direct emission or absorption of a phonon; Raman scattering of a phonon; and Orbach intervention of a third state.

We can now include the relaxation coming from spin-lattice relaxation in the equation of motion in Eq. (7.90):

$$\frac{dM_z}{dt} = \gamma(\boldsymbol{M} \times \boldsymbol{B}_a)_z + \frac{M_0 - M_z}{T_1}. \tag{7.94}$$

If in a static field $B_0\hat{\boldsymbol{z}}$ the transverse magnetization component M_x and M_y is not zero, then they will decay to zero which is their equilibrium value. To describe this, we introduce the transverse relaxation time T_2 entering the equation of motion as

$$\frac{dM_{x,y}}{dt} = \gamma(\boldsymbol{M} \times \boldsymbol{B})_{x,y} - \frac{M_{x,y}}{T_2}. \tag{7.95}$$

Since the magnetic energy $-\boldsymbol{M} \cdot \boldsymbol{B}_a$ does not change as $M_{x,y}$ varies, no energy needs to flow out of the magnetic part of the system during relaxation of $M_{x,y}$. Which one is greater of T_1 and T_2 is material-specific. The above set of equations for the magnetization components $M_{x,y,z}$ are collectively known as Bloch equations. The time T_2 is a measure of the time during which the individual moments that contribute to $M_{x,y}$ remain in phase with each other. Different local magnetic fields at the different spins will cause them to precess at different frequencies. If initially the spins have a common phase, the phases will become random in the course of time and the values of $M_{x,y}$ will become zero. We can think of T_2 as a dephasing time.

In experiments an rf magnetic field (this is what we previously referred to as an AC radiation field) is usually applied along the $\hat{\boldsymbol{x}}$ or $\hat{\boldsymbol{y}}$ axis. Our main interest is in the behavior of the magnetization in the combined rf and static fields, as in Fig. 7.10. We first determine the frequency of free precession of the spin system in a static field $\boldsymbol{B}_a = B_0\hat{\boldsymbol{z}}$ with $M_z = M_0$. The Bloch equations turn into

$$\frac{dM_x}{dt} = \gamma B_0 M_y - \frac{M_x}{T_2},\ \frac{dM_y}{dt} = \gamma B_0 M_x - \frac{M_y}{T_2},\ \frac{dM_z}{dt} = 0. \tag{7.96}$$

We look for damped oscillatory solutions of the form

$$M_x = m\mathrm{e}^{-t/T'}\cos(\omega_0 t),\ M_y = -m\mathrm{e}^{-t/T'}\sin(\omega_0 t). \tag{7.97}$$

Substituting into the Bloch equations gives

$$\omega_0 = \gamma B_0,\ T' = T_2. \tag{7.98}$$

This is exactly the announced resonance frequency condition for free precession. We could have started with a more general ansatz for $M_{x,y}$ in terms of amplitudes and harmonic functions, but the

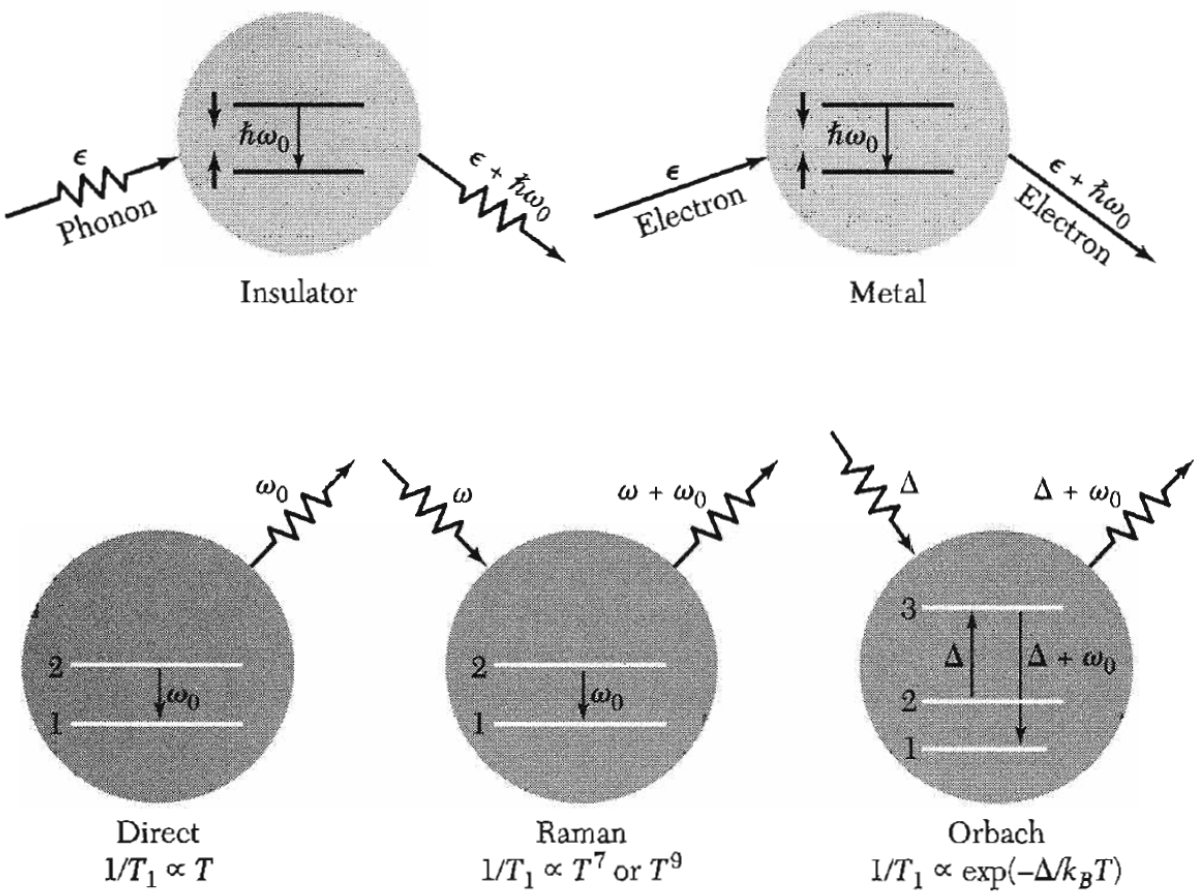

Figure 7.9: **Upper row:** Some important processes that contribute to longitudinal magnetization relaxation in an insulator and in a metal. For the insulator we show a phonon scattered inelastically by the spin system. The spin system moves to a lower energy state, and the emitted phonon has higher energy by $\hbar\omega_0$, than the absorbed phonon. For the metal we show a similar inelastic scattering process in which a conduction electron is scattered. **Lower row:** Spin relaxation from $2 + 1$ by phonon emission, phonon scattering, and a two-stage phonon process. The temperature dependence of the longitudinal relaxation time T_1 is shown for the several processes. Figure taken from [7].

solution would nevertheless have ended up being the same. The behavior of $M_{x,y}(t)$ is similar to that of a damped harmonic oscillator in two dimensions. The analogy suggests correctly that the spin system will show resonance in the absorption of energy from a driving field near the frequency $\omega = \omega_0$, so that constantly supplying radiation at this frequency will enable the magnetic system to uphold its precession. The frequency width $\Delta\omega$ of the peak shown in the absorption of the system as a function of the frequency of the driving field will scale like $1/T_2$.

The absorption of the energy can be shown to be determined by the imaginary part of the dynamic magnetic susceptibility. In other words, magnetization relaxation will cause a broadening of the resonance peak. This is a general feature of states with a finite lifetime, also seen in completely different fields of physics. In particle physics, the production of particles with a finite lifetime

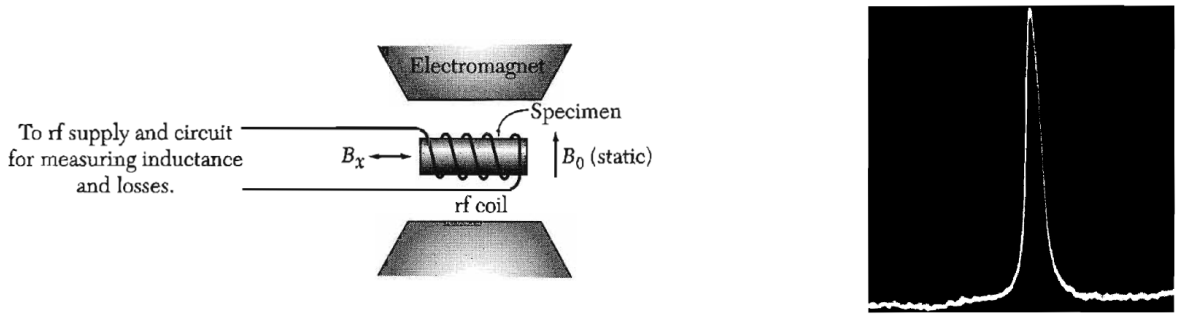

Figure 7.10: **Left:** schematic arrangement for magnetic resonance experiments. **Right:** experimental measurement of proton resonance absorption in water as a function of the frequency of the rf field. Figure taken from [7].

will turn up as resonance peaks when measuring the scattering cross section for a process. The shorter the lifetime of the particle is, the broader the peak is. In our particular example above, one proves formally that the width of the resonance peak scales like $1/T_2$ by including the rf magnetic field in the analysis and identifying the steady-state solution of the Bloch equations (i.e. where $dM_i/dt = 0$). Fig. 7.10 shows the absorption peak due to magnetic resonance of protons in water.

Note that the presence of internal fields, such as dipolar interactions, will affect both the value of the resonance frequency and the line width of the peak. For instance, at a fixed frequency the resonance of a nuclear spin is observed at a slightly different magnetic field in a metal than in a diamagnetic solid. The effect is known as the *Knight shift* or metallic shift and is vahiable as a tool for the study of conduction electrons. The interaction energy of a nucleus of spin $\boldsymbol{I}$ and gyromagnetic ratio γ_I is:

$$U = -(-\gamma_I \hbar B_0 + a\langle S_z \rangle) I_z \tag{7.99}$$

where the first term is the interaction with the applied magnetic field B_0, and the second is the average hyperfine interaction of thc nucleus with the conduction electrons. The average conduction electron spin $\langle S_z \rangle$ is related to the Pauli spin susceptibility χ_{Pauli}, of the conduction electrons: $M_z = gN\mu_B \langle S_z \rangle = \chi_{\text{Pauli}} B_0$. Therefore, the interaction may be written as

$$U = \Big(-\gamma_I \hbar + \frac{a\chi_{\text{Pauli}}}{gN\mu_B} \Big) B_0 I_z = -\gamma_I \hbar B_0 \Big(1 + \frac{\Delta B}{B_0} \Big) I_z. \tag{7.100}$$

The Knight shift K is defined as

$$K = -\frac{\Delta B}{B_0} = \frac{a\chi_{\text{Pauli}}}{gN\mu_B \gamma_I \hbar} \tag{7.101}$$

and thus provides an effective fractional change in the gyromagnetic ratio γ_I due to the response of the conduction electrons. In metallic elements like Al, K, V, Cr, the Knight shift is typically of order a few percent.

7.10 Electron paramagnetic resonance

The phenomenon of electron paramagnetic resonance is conceptually similar to nuclear magnetic resonance, except that it is the spins of unpaired electrons rather than nuclei that respond to the

magnetic field and radiation.

In the presence of an external field B_0, the separation between the lower and upper energy level for an electron due to Zeeman-splitting is $\Delta E = g_e \mu_B B_0$. An unpaired electron can change its electron spin by either absorbing or emitting a photon of energy $\hbar\omega$ so long that the resonance condition $\Delta E = \hbar\omega$ is fulfilled. By experimentally measuring at what frequencies electromagnetic radiation is absorbed/emitted from a material subject to an external field B_0, one thus gains information about the spin splitting of energy levels in the materials. This can in turn be used to determine the effective g-factor for electrons in a material, which due to interactions between the electron and its environment can deviate strongly from the free electron value $g_e \simeq 2$.

A simplified way to think of this is the following. An unpaired electron will feel not only an applied external field B_0, but also any local magnetic field provided by the atom or molecules. The effective field B_{eff} felt by the electron can therefore be written

$$B_{\text{eff}} = B_0(1 - \sigma) \tag{7.102}$$

where σ can be both positive or negative and accounts for the effect of local fields. The resonance condition for EPR is then

$$h\nu = g_e \mu_B B_{\text{eff}} \tag{7.103}$$

which can be rewritten as

$$h\nu = g_{\text{eff}} \mu_B B_0 \text{ where } g_{\text{eff}} = g_e(1 - \sigma). \tag{7.104}$$

This equation can then be used to determine the effective g in an experiment by measuring the field and the frequency at which the resonance occurs. The magnitude of change of g, which can arise for instance due to spin-orbit coupling, then gives information about the atomic orbital containing the unpaired electron.

7.10.1 Ferromagnetic resonance

Spin resonance at microwave frequencies in ferromagnets is similar in principle to nuclear spin resonance. The total electron magnetic moment of the specimen precesses about the direction of the static magnetic field, and energy is absorbed strongly from the rf transverse field when its frequency is equal to the precessional frequency. We may think of the macroscopic vector S representing the total spin of the ferromagnet as quantized in the static magnetic field, with energy levels separated by the usual Zeeman splitting.

It turns out that the shape of the specimen plays an important role due to *demagnetization factors* coming from the demagnetization field. The demagnetization field is a magnetic field generated by the magnetization of a material. It is a field that therefore appears in ferromagnets, but also in superconductors due to the magnetization created by screening currents. The demagnetization field can be understood by considering the fundamental equations $\nabla \cdot \boldsymbol{B} = 0$ and $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$ which are combined into

$$\nabla \cdot \boldsymbol{H} = -\nabla \cdot \boldsymbol{M}. \tag{7.105}$$

In the simplest case where $\boldsymbol{M}$ is constant inside a magnet (homogeneous ferromagnet) and zero outside, the gradient is finite right at the edge of the material. Due to the minus sign in the above

equation, the demagnetization field $\boldsymbol{H}$ has to provide a flux in the opposite direction as $\boldsymbol{M}$ and thus has to act as to partially cancel the total magnetic field inside the ferromagnet. The manner in which this is realized is illustrated in Fig. 7.11.

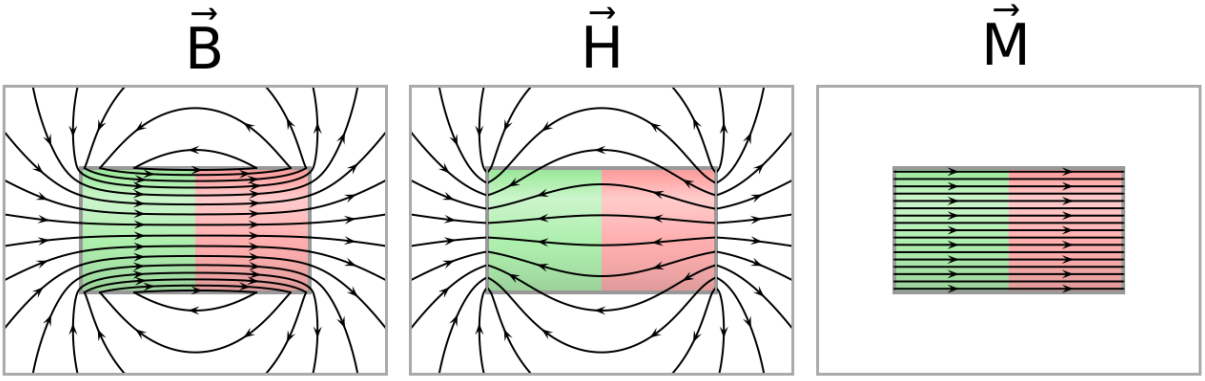

Figure 7.11: Illustration of the magnetic field $\boldsymbol{B}$ and the corresponding $\boldsymbol{H}$ and $\boldsymbol{M}$ fields for a cylindrical bar magnet. The field $\boldsymbol{H}$ is referred to as the demagnetization field. The picture is taken from Wikipedia.

Let us show analytically why the demagnetization field (known as the *stray field* outside of the ferromagnet) points in the opposite direction of $\boldsymbol{M}$. The main point is that we can show from Maxwell's equations for the magnetic field that in the situation above $\boldsymbol{H}$ acts mathematically analogously to an electric field $\boldsymbol{E}$ created by a charge distribution. Therefore, we can think of $\boldsymbol{H}$ as created by fictious magnetic monopoles just as an $\boldsymbol{E}$-field can be created by electric monopoles. These magnetic monopoles are positioned at the N and S poles of the magnet, and give rise to an $\boldsymbol{H}$-field which points either inward or outward as shown in Fig. 7.11.

Consider Maxwell's equation for the $\boldsymbol{H}$-field in the absence of any macroscopic currents in the material and time-varying electric fields. This is the situation of a ferromagnet in thermodynamic equilibrium. Maxwells equation then reads

$$\nabla \times \boldsymbol{H} = 0 \tag{7.106}$$

which means that we can write $\boldsymbol{H} = -\nabla U(\boldsymbol{r})$ where $U(\boldsymbol{r})$ is referred to as a magnetic potential. Furthermore, since $\nabla \cdot \boldsymbol{B} = \mu_0(\boldsymbol{B} + \boldsymbol{M})$, it follows that $\nabla^2 U(\boldsymbol{r}) = \nabla \cdot \boldsymbol{M}$. Outside of the material where $\boldsymbol{M} = 0$, we have $\nabla^2 U(\boldsymbol{r}) = 0$.

To solve the above equations, we need accompanying boundary conditions. The continuity conditions for $H_{\parallel}$ and $B_{\perp}$ at the surface of the material dictate that $U(\boldsymbol{r})$ has to be continuous at the surface and that

$$\left(\frac{\partial U}{\partial n}\right)_{\text{int}} - \left(\frac{\partial U}{\partial n}\right)_{\text{ext}} = \boldsymbol{M} \cdot \boldsymbol{n} \tag{7.107}$$

where $\boldsymbol{n}$ is the surface normal.

The above system of equations and boundary conditions is analogous to the problem of an electro-

static potential produced by charge distribution. The solution for $U(\boldsymbol{r})$ can be shown to be:

$$U(\boldsymbol{r}) = \frac{1}{4\pi}\Big[- \int_V \frac{\nabla \cdot \boldsymbol{M}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r}' + \int_S \frac{\boldsymbol{n} \cdot \boldsymbol{M}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} dS' \Big]. \tag{7.108}$$

The first term is zero in our case of a homogeneous ferromagnet since $\nabla \cdot \boldsymbol{M} = 0$ inside the material. The second term of the potential is the potential associated with the surface density of "magnetic poles" $\sigma \equiv \boldsymbol{M} \cdot \boldsymbol{n}$. These poles arise due to the discontinuity of the magnetization at the surface. The magnetic poles are positive at N, where the magnetization points out of the surface, whereas they are negative at S. Therefore, the belonging $\boldsymbol{H}$-field points in opposite directions at the two poles, just as the $\boldsymbol{E}$-field of two opposite charges.

The demagnetization field is thus created by the magnetization itself and acts to reduce the total $\boldsymbol{B}$-field inside the magnet. If an external field is also present, we would write the total $\boldsymbol{H}$-field as

$$\boldsymbol{H} = \boldsymbol{H}_a + \boldsymbol{H}_d \tag{7.109}$$

where $\boldsymbol{H}_a$ is the applied magnetic field and $\boldsymbol{H}_d = \boldsymbol{H}_d(\boldsymbol{M})$ in general has a complicated dependence on the magnetization $\boldsymbol{M}$, as shown by the analytical expressions above. For certain shapes of the magnet, the expression for the demagnetization field becomes more tractable. In particular, for an ellipsoidally shaped magnet the demagnetization field inside the material is uniform if the magnetization is uniform and takes the form

$$\boldsymbol{H}_d = -N\boldsymbol{M} \tag{7.110}$$

where N is a tensor.

We can then generally write the components B_i of total magnetic field inside the material as

$$B_i = B_{i,\text{ext}} + \mu_0(M_i - N_i M_i) \tag{7.111}$$

The term $\mu_0 M_i$ will not contribute to the spin equation of motion since we should take the cross product with the magnetization itself. Thus, we obtain for an applied static field $B_0\hat{\boldsymbol{z}}$:

$$\begin{aligned} \frac{dM_x}{dt} &= \gamma(M_y B_z - M_z B_y) = \gamma[B_0 + (N_y - N_z)\mu_0 M]\mu_0 M_y, \\ \frac{dM_y}{dt} &= \gamma[\mu_0 M(-N_x M_x - Mx(B_0 - N_z M)] = -\gamma[B_0 + (N_x - N_z)M]\mu_0 M_x \end{aligned} \tag{7.112}$$

where we approximated $M_z = M$ and $dM_z/dt = 0$ by assuming a weak precession with small transverse components $M_{x,y}$. Stationary solutions of the above equation system with a time dependence $\mathrm{e}^{-i\omega t}$ (the actual magnetization components are then obtained by taking the real part of $M_{x,y}$, which is allowed since the equations are linear in $M_{x,y}$) exist if

$$\begin{vmatrix} i\omega & \gamma[B_0 + (N_y - N_z)\mu_0 M] \\ -\gamma[B_0 + (N_x - N_z)\mu_0 M] & i\omega \end{vmatrix}. \tag{7.113}$$

One obtains this equation by demanding that the linear system of coupled equations has a non-trivial solution.

$$\omega_0^2 = \gamma^2[B_0 + (N_y - N_z)\mu_0 M][B_0 + (N_x - N_z)\mu_0 M]. \tag{7.114}$$

The frequency ω_0 is called the frequency of the uniform mode, in distinction to the frequencies of magnon and other spatially nonuniform modes. In the uniform mode all the moments precess together in phase with the same amplitude.
For a sphere $N_x = N_y = N_z$, we obtain $\omega_0 = \gamma B_0$. For a flat plate with B_0 perpendicular to the plate, $N_x = N_y = 0$ and $N_z = 1$, so that

$$\omega_0 = \gamma(B_0 - \mu_0 M). \tag{7.115}$$

If B_0 is instead parallell to the plane of the plate, then $N_x = N_z = 0$ and $N_y = 1$, giving

$$\omega_0 = \gamma[B_0(B_0 + \mu_0 M)]^{1/2}. \tag{7.116}$$

Spin-wave resonance

Uniform rf magnetic fields can excite long-wavelength spin waves in thin ferromagnetic films if the electron spins on the surfaces of the film see different anisotropy fields than the spins within the films. In effect, the surface spins may be pinned by surface anisotropy interactions. If the rf field is uniform, it can excite waves with an odd number of half-wavelengths within the thickness of the film, see Fig. 7.12. Waves with an even number of half-wavelengths have no net interaction energy with the field.

The condition for spin wave resonance (SWR) with the applied magnetic field normal to the film is obtained by adding to the right-hand side of Eq. (7.115) the exchange contribution to the frequency. The exchange contribution may be written as Dk^2 (as we derived for magnons in ferromagnets earlier), where D is called the spin wave exchange constant. The wavevector k for a mode of n half-wavelengths in a film of thickness L is $k = n\pi/L$ and n has to be odd to obtain SWR as mentioned above.

7.10.2 Antiferromagnetic resonance

The magnetizations on each sublattice in an antiferromagnet, displaying staggered magnetic order, can also exhibit resonance behavior. Consider a uniaxial antiferromagnet with spins on two sublattices, 1 and 2. Assume that the magnetization $\boldsymbol{M}_1$ on sublattice 1 is directed along the $+\hat{\boldsymbol{z}}$ direction by an anisotropy field $B_A\hat{\boldsymbol{z}}$. The anisotropy field is an internal field in the ferromagnet originating from spin-orbit interactions which makes the sublattice spins have an energetically preferred direction in space. Conversely, $\boldsymbol{M}_2$ on sublattice 2 is directed along $-\hat{\boldsymbol{z}}$.

In an antiferromagnet, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are coupled in an antiparallell way due to the exchange interaction. Therefore, we can introduce the exchange fields

$$\boldsymbol{B}_{1,\text{exc}} = -\lambda \boldsymbol{M}_2, \ \boldsymbol{B}_{2,\text{exc}} = -\lambda \boldsymbol{M}_1. \tag{7.117}$$

When $\lambda > 0$, these internal fields acting on the sublattice magnetizations will favor an antiparallell alignment of $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$. In the absence of an external magnetic field, the total fields acting on $\boldsymbol{M}_{1,2}$ are shown in Fig. 7.13.
We again assume a small deviation from equilibrium, so that $M_1^z = M$ and $M_2^z = -M$. The

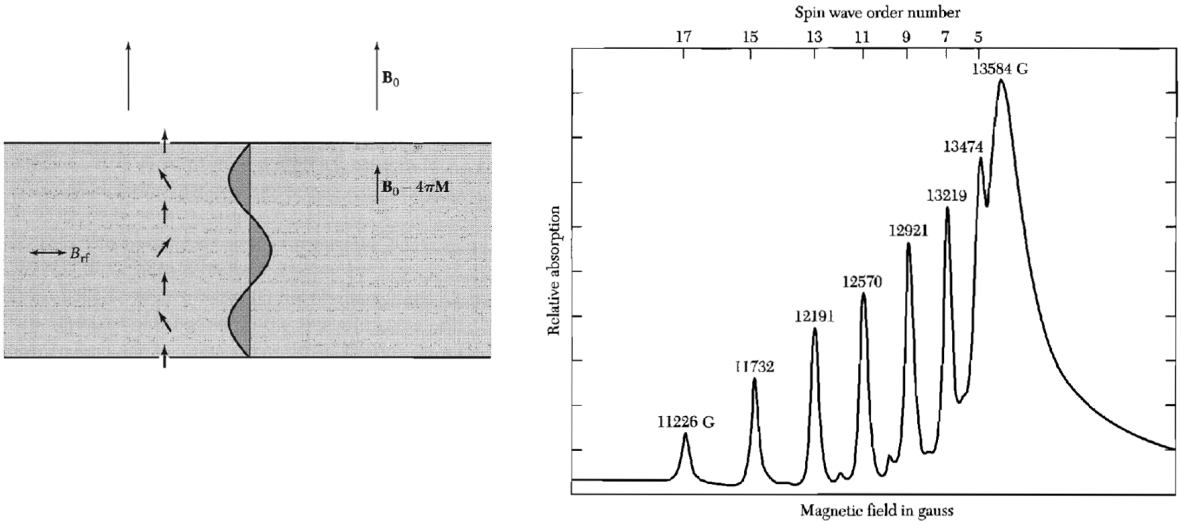

Figure 7.12: **Left**: spin wave resonance in a thin film. The plane of the film is normal to the applied magnetic field B_0. A cross section of the film is shown here. The internal magnetic field is $B_0 - \mu_0 M$ (shown in CGS units in the figure). The spins on the surfaces of the film are assumed to be held fixed in direction by surface anisotropy forces. A uniform rf field will excite spin wave modes having an odd number of half-wavelengths. The wave shown is for $n = 3$ half-wavelengths. **Right:** Spin wave resonance spectrum in a Permalloy (80Ni20Fe) film at 9 GHz. The order number is the number of half-wavelengths in the thickness of the film. Figure taken from Ref. [7].

linearized equations of motion are then

$$\begin{aligned}
\frac{dM_1^x}{dt} &= \gamma[M_1^y[\lambda M + B_A) - M(-\lambda M_2^y)], \\
\frac{dM_1^y}{dt} &= \gamma[M(-\lambda M_2^x) - M_1^x(\lambda M + B_A)], \\
\frac{dM_2^x}{dt} &= \gamma[M_2^y(-\lambda M - B_A) - (-M)(-\lambda M_1^y)], \\
\frac{dM_2^y}{dt} &= \gamma[(-M)(-\lambda M_1^x) - M_2^x(\lambda M - B_A)].
\end{aligned} \tag{7.118}$$

We define $M_1^+ = M_1^x + iM_1^y$, $M_2^+ = M_2^x + iM_2^y$. Assuming again a time-dependence $\mathrm{e}^{-i\omega_0 t}$ for $M_{1,2}^+$, we obtain

$$\begin{aligned}
-i\omega M_1^+ &= -i\gamma[M_1^+(B_A + \lambda M) + M_2^+ \lambda M], \\
-i\omega M_2^+ &= i\gamma[M_2^+(B_A + \lambda M) + M_1^+ \lambda M].
\end{aligned} \tag{7.119}$$

These equations have a solution if

$$\begin{vmatrix} \gamma(B_A + B_E) - \omega_0 & \gamma B_E \\ \gamma B_E & \gamma(B_A + B_E) + \omega \end{vmatrix} \tag{7.120}$$

where we defined the exchange field $B_E \equiv \lambda M$. Thus, the antiferromagnetic resonance frequency is given by

$$\omega_0^2 = \gamma^2 B_A(B_A + 2B_E). \tag{7.121}$$

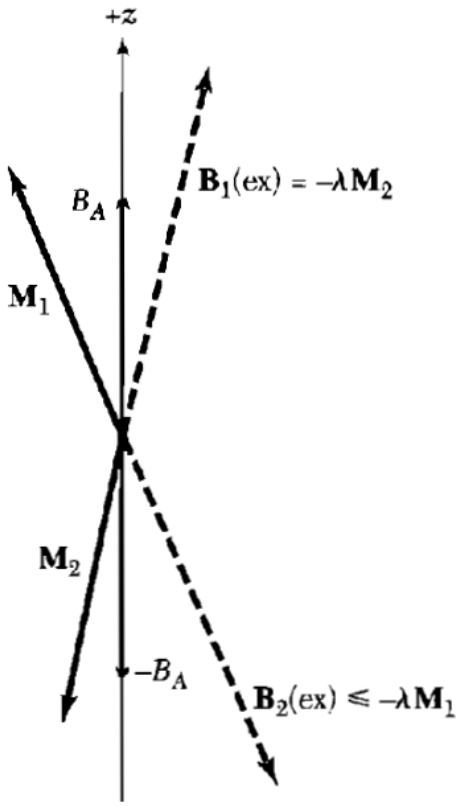

Figure 7.13: Effective fields in AFM resonance. The magnetization $\boldsymbol{M}_1$ of sublattice 1 sees a field $-\lambda\boldsymbol{M}_2 + \boldsymbol{B}_A\hat{\boldsymbol{z}}$. The magnetization $\boldsymbol{M}_2$ sees $-\lambda\boldsymbol{M}_1 - B_A\hat{\boldsymbol{z}}$. Both ends of the crystal axis are "easy axes" of magnetization. Figure taken from Ref. [7].

A notable quantitative difference between the magnitude of the resonance frequency for ferromagnets and antiferromagnets is that while the former typically lies in the range of GHz, the latter has resonance frequencies in the range of THz due to the very strong exchange field B_E. For instance, antiferromagnets MnO and NiO have resonance frequencies 0.82 THz and 1.09 THz, respectively.

Bibliography

[1] C. Timm, lecture notes (https://obelix.physik.uni-bielefeld.de/~schnack/molmag/material/timm-lecturenotes.pdf).

[2] C. Timm, lecture notes (https://tu-dresden.de/mn/physik/itp/cmt/ressourcen/dateien/skripte/Skript_Supra.pdf?lang=en).

[3] M. Sigrist, lecture notes (https://edu.itp.phys.ethz.ch/fs13/sst/Lecture-Notes.pdf).

[4] Advanced Solid State Physics, lecture notes by P. Hadley, TU Graz (https://lampz.tugraz.at/~hadley/ss2/introduction/index.php).

[5] M. van Veenendaal, lecture notes (https://www.niu.edu/veenendaal/_internal/_pdf/667.pdf).

[6] J. O. Fjærestad, lecture notes (https://folk.ntnu.no/johnof/magnetism-2014.pdf and https://johnof.folk.ntnu.no/second-quantization-2014.pdf).

[7] C. Kittel, *Introduction to Solid State Physics*, New York: Wiley (8th edition, 1953).

[8] D. Tong, *Lectures on applications of quantum mechanics*, https://www.damtp.cam.ac.uk/user/tong/aqm.html and *Lectures on kinetic theory*, https://www.damtp.cam.ac.uk/user/tong/kintheory/four.pdf.

[9] N. Spaldin, Journal of Solid State Chemistry **195**, 2 (2012).

[10] K. Fossheim and A. Sudbø, *Superconductivity: Physics and Applications* (Wiley, 2005)

[11] https://nanoscale.blogspot.com/2018/08/what-is-dielectric-polarization.html

[12] Oxford University, lecture notes (https://www2.physics.ox.ac.uk/sites/default/files/2011-10-04/crystalstructure_handout10_pdf_19544.pdf).

[13] https://neutrons2.ornl.gov/conf/nxs2013/lecture/pdf/NXS_Lee_2013.pdf