Physics of solids

Lecture notes

Ezio Puppin

2025

Summary

Introduction vii
Condensed matter vii
The constituent particles viii
The role of statistical physics ix
Chemical bonds x
Molecular crystals x
The course programme xi
1. Crystal structures in the real space ..... - 1 -
1.1 Crystal structures: lattice + basis ..... - 1 -
1.2 Symmetry and crystal structures ..... - 2 -
1.3 Translational symmetry: the Bravais lattice ..... - 3 -
1.4 The elementary cell ..... - 4 -
1.5 Possible Bravais lattices ..... - 5 -
1.6 Possible rotations of a Bravais lattice ..... - 8 -
1.7 The symmetry operations of a Bravais lattice (the crystal systems) ..... - 9 -
1.8 Possible symmetries of the basis ..... - 10 -
1.9 Point and translation symmetry operations of a structure (lattice + basis) ..... - 11 -
1.10 Physical consequences of symmetry ..... - 12 -
2. Crystal structures in the reciprocal space and diffraction ..... - 13 -
2.1 Fourier series ..... - 13 -
2.2 The reciprocal lattice in 1D ..... - 17 -
2.3 The reciprocal lattice in the general case ..... - 18 -
2.4 Reciprocal lattice properties in 2D and 3D ..... - 22 -
2.5 Miller indices ..... - 23 -
2.6 General considerations on the reciprocal lattice ..... - 24 -
2.7 Diffraction from a Bravais lattice ..... - 25 -
2.8 Diffraction from structures with basis ..... - 29 -
2.9 The probes used for diffraction measurements ..... - 31 -
3. Classical model of lattice oscillations ..... - 32 -
3.1 Approximations ..... - 32 -
3.2 General scheme of the lattice oscillations theory ..... - 33 -
3.3 The equations of motion ..... - 34 -
3.4 The properties of the solutions ..... - 36 -
3.5 Solution of the problem ..... - 38 -

3.6 Chain of equal atoms ..... - 39 -
3.7 Chain with two atoms per cell ..... - 43 -
3.8 Chains with more than two atoms per cell ..... - 45 -
3.9 Experimental dispersion relations ..... - 46 -
4. Specific heats of solids ..... - 48 -
4.1 Phenomenological aspects ..... - 48 -
4.2 The solid as a gas of oscillators ..... - 50 -
4.3 Quantisation of the oscillator energy and quasi-particles ..... - 50 -
4.4 The boundary conditions ..... - 52 -
4.5 Quantization of vectors q ..... - 53 -
4.6 Density of states ..... - 54 -
4.7 Van Hove singularities ..... - 57 -
4.8 A few remind of statistical physics ..... - 58 -
4.9 The general formula for the lattice specific heat ..... - 60 -
4.10 Specific heat for large and small values of T ..... - 61 -
4.11 The Einstein Model ..... - 62 -
4.12 The Debye model ..... - 62 -
4.13 Lattice and electronic specific heat ..... - 64 -
5. Lattice thermal conductivity ..... - 65 -
5.1 Phenomenology and macroscopic description ..... - 65 -
5.2 Modelling ..... - 66 -
5.3 Localized packets of phonons ..... - 67 -
5.4 The transport equation ..... - 69 -
5.5 Scattering mechanisms ..... - 71 -
5.6 The phonon - phonon interaction ..... - 72 -
5.7 High temperature thermal conductivity ..... - 73 -
5.8 Conservation laws in scattering between phonons ..... - 74 -
5.9 Normal and Umklapp processes ..... - 75 -
5.10 Thermal equilibrium and the role of Umklapp processes ..... - 76 -
6. Electron states ..... - 78 -
6.1 Introduction ..... - 78 -
6.2 The formulation of the problem in general terms ..... - 78 -
6.3 Approximations ..... - 79 -
6.4 Electrons in a periodic potential: preliminary considerations ..... - 81 -
6.5 The Bloch theorem in the general form ..... - 83 -
6.6 A few consequences of the Bloch theorem ..... - 84 -
6.7 Demonstration of the Bloch theorem in real space ..... - 86 -
6.8 Possible values of k ..... - 89 -

6.9 Demonstration of the Bloch theorem in the reciprocal space ..... - 90 -
6.10 The solutions of the Schrodinger equation with a periodic potential ..... - 93 -
6.11 The nearly free electron model ..... - 94 -
6.12 The tight binding model ..... - 98 -
6.13 The occupation of the electron states ..... - 99 -
6.14 The Fermi surface ..... - 101 -
6.15 The successive Brillouin zones ..... - 103 -
6.16 The density of states ..... - 104 -
6.17 Comparison of electron states and lattice oscillations ..... - 105 -
7. Transport of electrons ..... - 108 -
7.1 General considerations ..... - 108 -
7.2 The semi-classical model ..... - 108 -
7.3 Free electrons or in the presence of a weakly variable potential ..... - 109 -
7.4 Electrons in a rapidly changing periodic potential ..... - 112 -
7.5 Effective mass ..... - 115 -
7.6 Dynamics of an electron in a band ..... - 118 -
7.7 Conduction from a band ..... - 120 -
7.8 The mechanisms of electron diffusion ..... - 121 -
7.9 The electron escape depth and the independent electrons approximation ..... - 122 -
7.10 The transport equation ..... - 124 -
7.11 Conduction in a metal ..... - 127 -
8. Superconductivity ..... - 129 -
8.1 The cancellation of resistivity ..... - 129 -
8.2 The Meissner-Ochsenfeld effect and perfect diamagnetism ..... - 130 -
8.3 The effects of magnetic fields ..... - 131 -
8.4 The superconducting gap ..... - 132 -
8.5 Applications of superconductivity ..... - 133 -
8.6 Theories of superconductivity ..... - 134 -
8.7 London equations ..... - 134 -
8.8 Phase transitions and their order ..... - 137 -
8.9 The Landau theory of phase transitions ..... - 139 -
8.10 The Ginzburg-Landau theory of superconductivity ..... - 142 -
8.11 The Ginzburg-Landau equations ..... - 143 -
8.12 The characteristic lengths of superconductivity ..... - 144 -
8.13 The lower critical magnetic field ..... - 145 -
8.14 Energy considerations and type II superconductors ..... - 145 -
8.15 The vector potential ..... - 147 -
8.16 Quantization of the magnetic flux ..... - 149 -

8.17 Symmetry breaking with a complex order parameter ..... - 150 -
8.18 Excitations with a complex order parameter ..... - 151 -
8.19 Gauge invariance and superconductivity ..... - 153 -
8.20 BCS Theory: Cooper Pairs ..... - 155 -
8.21 The BCS state ..... - 158 -
Suggested texts ..... - 161 -

Traducción / Anotaciones

Definición

Contradicciones/Contraejemplos

Definiciones propias/Conceptos Puntuales/Investigación

Definiciones complementarias / Claridades Ejemplos

> “I understand what an equation means if I have a way of figuring out the characteristics of its solutions without actually solving it.” P. A. M. Dirac

Introduction

Condensed matter

The term condensed matter is used when dealing with macroscopic physical systems consisting of a large number of particles that forces of different origin tend to keep together. These constituent particles can be atoms or molecules and the attractive forces between them are those of electrostatic type that are exerted between the charged particles of which the atoms are constituted.

Even a gas is a macroscopic physical system made up of many particles, but in this case we do not speak of condensed matter for the obvious reason that there is no "condensation" of these particles which, on the contrary, retain their individuality, and what results is a system whose density can assume arbitrary values. When, on the other hand, atoms and molecules aggregate and maintain close contact with each other, the density of the medium is almost constant.

The physics of condensed matter is a vast field of investigation, for example it is sufficient to scroll through the list of research subjects collected in the so called PACS (Physics And Astronomy Classification Scheme). In Italy alone, several thousand scientific articles are published every year in the field of condensed matter physics and in the world they go well beyond the figure of 100.000 publications.

"brevemente" "en pocas palabras"

Wanting to briefly classify the many forms of condensed matter, a first distinction to be made is that between solids and liquids. In the first, the relative positions of the constituent atoms are fixed, while in liquids the constituent particles are free to move freely with respect to each other while remaining in close contact.

In solid materials, atoms can be arranged in an orderly fashion, in which case we speak of crystals or crystalline solids. If, on the other hand, the positions of the atoms do not follow a precise rule but they are arranged in a disorderly manner, then we speak of amorphous solids, a word that means "without form".

The radial distribution function g(r) is defined as the probability of finding an atom at a distance r (moving in all directions) from an assigned reference point. This is a curve that can be obtained by diffraction experiments. In a perfect crystal, the g(r) consists of a series of well-defined peaks as shown in the figure.

Amorphous materials can be imagined as distorted crystals in which the distance between atoms is not always exactly the same but varies randomly. In this case the g(r) assumes a course as shown in the figure, with well defined peaks at low values of r, that indicate the presence of a short range order, while for great values of r the order disappears and we find ourselves in the same situation of a gas. This trend in g(r) is also observed in liquids. In gases, on the other hand, there is no order, even at short range, and the g(r) has no structure.

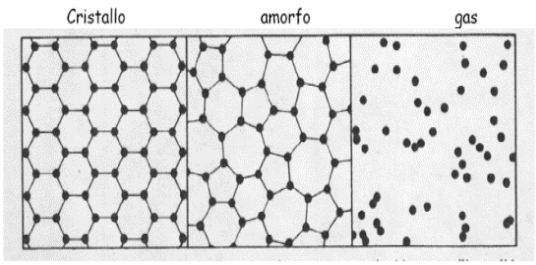


Fig.1

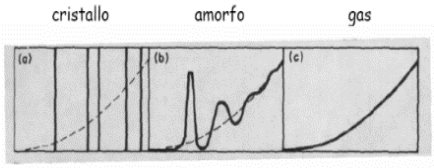


Fig.2 - Funzione di distribuzione radiale per un cristallo, per un solido amorfo e per un gas.

The g(r) of liquids is like that of amorphous solids. If the temperature T increases, the peaks widen

There are also the so called quasicrystals, discovered in the 80s of the '900, which do not respect the rule of perfect order that is observed in the crystals, but which are not even completely disordered as the amorphous materials. In scientific language all possible forms of solid state are catalogued as hard matter.

Even the distinction between solids and liquids is not as clear as we might think because there is a wide category of materials whose characteristics can be considered as intermediate. These include colloids, liquid crystals and living matter. In these cases the term soft matter is used, incidentally a field of research that is currently in a phase of enormous development.

The constituent particles

In the theories used to describe the properties of condensed matter, it is almost always assumed that atoms maintain their individuality and that only the outermost electrons modify their properties to give rise to chemical bonds. In this approximation we speak of ions to indicate the part of the atom that is not modified after the formation of the condensed state and of valence electrons to indicate the electrons directly involved in the formation of bonds.

The physics of the solids we are going to deal with is based precisely on this assumption, but before proceeding, it is worth remembering that there are also other cases of condensed matter that do not fall within this scheme.

A first example is represented by the physical systems in which the pressure and/or the temperature increase to levels such as to destroy the atoms or even the nuclei as happens in the neutron stars (that too is condensed matter, higly condensed).

Another very important situation occurs when quantum effects related to the dual nature of material particles prevails. To understand how this can happen, remember that the de Broglie wavelength λ_{DB} of an atom with mass m and momentum p is:

$$\lambda_{DB} = \frac{h}{p} = \frac{h}{\sqrt{2mE}}$$

In obtaining this expression it was assumed that $E = p^2/2m$, as required by Newtonian mechanics while for relativistic particles it is necessary to use a different expression. In the case of the atoms of ordinary matter, the Newtonian formulation is still perfectly adequate because the speeds involved are still much lower compared to the speed of light.

On the basis of the principle of energy equipartition we can think that each of the atoms of the solid has an energy equal to 3/2kT and this fixes the link between the De Broglie wavelength and the temperature:

$$\lambda_{DB} = \frac{h}{\sqrt{3mkT}}$$

At room temperature kT = 25 meV = 4×10^{-22} J and, for a sodium atom, m = 37×10^{-27} kg. This results in $\lambda_{DB} = 10^{-14}$ m. In a solid the lattice parameter is of the order of 10^{-10} m and therefore, in the case we are considering, the associated wavelength is much smaller than the space in which the atom is located. In these conditions it is possible to represent the atom as a classical object built by superimposing waves of much shorter length than the dimensions within which the object is confined.

However, when the temperature drops to levels such that λ_{DB} becomes comparable with the distance between the atoms, then the wave functions that represent the particles overlap, resulting in objects of macroscopic size but with quantum behavior. These are so called Bose-Einstein condensates. These phenomena occur at temperatures of the order of hundreds of nanoKelvin and have been experimentally observed. Similar phenomena, that can be traced back to quantum effects on a macroscopic scale, are those of superfluidity or superconductivity.

The role of statistical physics

The particles that make up the condensed matter are in enormous numbers and the macroscopic properties of the materials are the expression of their collective behavior. There is a branch of physics that, based on a few fundamental principles, is able to provide forecasts on the macroscopic behavior of physical systems. We are talking about thermodynamics which, as known, is based on some general principles and on the data obtained through experiments, such as for example the values actually assumed by the physical quantities and the empirical laws that link them to each other, the so called state equations.

In the case of systems made up of many particles, the physical laws that regulate their reciprocal interactions are generally known and it would be nice to be able to start from them to obtain the thermodynamic properties. This is not possible, however, relying only on thermodynamics because, although very powerful, it has an intrinsic limit, that of not allowing to deduce from first principles the values of the physical quantities and the corresponding state equations.

In order to do this, it is necessary to use the tools of statistical physics, which for this reason plays a fundamental role in the study of condensed matter. This is the reason why, as we go deeper into the study of the properties of solids, we will devote some effort to the study of statistical physics.

Chemical bonds

In materials, what keeps the atoms in the positions defined by the peaks of the radial distribution are the chemical bonds that are formed by the most external electrons. The bonds can be of various types and are usually ranked as shown in the table below.

TABLE 1-3 SUMMARY OF THE CLASSIFICATION OF BONDING TYPES IN SOLIDS

Bond Type	Typical Examples				Some Properties of the Type
	Material	Crystal Structure	Binding Energy (eV/molecule)	Nearest-Neighbor Distance (Å)	
Van der Waals (Molecular crystals)	Argon	F.C.C.	0.1	3.76	Low melting and boiling points.
	Chlorine	Tetragonal	0.3	4.34	Easily compressible.
	Hydrogen	H.C.P.	0.01	3.75	Electrically insulating. Transparent for photon energies to far U.V.
Covalent	Silicon	Cubic (Diamond)	3.7	2.35	Rigid and hard. Often high melting points.
	InSb	Cubic (Zinc blende)	3.4	2.80	Electrical insulator or semiconductor. Strongly absorbing for photon energies above intrinsic edge, transparent (when pure) for longer wavelengths.
	Mg_2Sn	Cubic (Fluorite)	1.0	2.92	
Ionic	KCl	Cubic (Rocksalt)	7.3	3.14	Rather plastic. Often dissociates on heating.
	AgBr	Cubic (Rocksalt)	5.4	2.88	Electrical insulator at low temperatures.
	BaF_2	Cubic (Fluorite)	17.3	2.69	Lattice disorder promotes ionic conduction at higher temperatures. Has reststrahlen absorption in I.R. as well as intrinsic absorption.
Hydrogen Bond	Ice	Hexagonal	0.5	1.75	Many allotropic forms. Electrically non-conducting. Dielectric activity. Optically transparent.
Metallic	Sodium	B.C.C.	1.1	3.70	Large spacing and coordination number. Electrical conductor. Opaque and highly reflecting in infrared and visible light; transparent in U.V.
	Silver	F.C.C.	3.0	2.88	
	Nickel	F.C.C.	4.4	2.48	

The type of bond that is established determines all the mechanical, electrical, optical, magnetic, ... properties of the materials. One of the fundamental objectives of solid state physics is to explain these properties in relation to the type of bond between the constituent atoms.

Molecular crystals

A little-known type of bond is that which is established in so called "molecular crystals". These are solids in which the constituent atoms (or molecules) maintain their individuality and the electrons of the external shells remain in the same configuration as they have in the atom or in the isolated molecule without mixing with those of other atoms to form chemical bonds. What keeps these atoms together, then, if the external electrons do not play the role of "glue" that they have in covalent and metallic crystals or if there is no transfer of electrons from one atomic species to another as it happens in ionic crystals?

Usually textbooks offer a rather hasty explanation: it is said that the so called "van der Waals forces" act, attractive forces of an electrostatic nature that act between atoms and molecules. Like all we have to do is name something in order to explain what it is.

In some cases the origin of the attractive forces between the neutral molecules is quite simple to understand, as in the case of polar molecules, i.e. those that possess an intrinsic electric dipole moment such as water. In this case the dipoles are oriented in such a way to minimize the energy of the system giving rise to an attractive force. This does not entirely clarify what happens, because the way in which these dipoles are arranged is anything but easy to describe, but at least the essential ingredient is known: there are electric dipoles that exert forces on each other.

Much more subtle is the case of solids consisting of individual atoms or non polar molecules. Let us consider the case, for example, of all noble gases that, below a certain temperature, first liquefy and then solidify (except helium, which liquefies at 4.2 K but in order to become solid requires that the pressure is greater than the atmospheric one). Noble gas atoms are charge distributions without an electric dipole moment. If they join together to form a solid, then the existence of <u>induced</u> dipole moments must be hypothesized.

These dipole moments cannot be permanent, as in the case of water molecules. The only possibility is that these are dipole moments due to temporary fluctuations in the distribution of atomic charges. However, since we are talking about atoms in their fundamental state (and since they are noble gases with the first level excited at very high energies) we need to identify the source of energy capable of producing these fluctuations. The answer to this question was first given by Fritz London in 1930. His reasoning was as follows. Atoms can be classically seen as harmonic oscillators, with two charge distributions of opposite sign that attract each other. Quantum mechanics tells us that the minimum energy of an oscillator cannot be zero but there is the so called "zero point energy" equal to $\frac{1}{2}\hbar\omega_o$, being ω_o the pulsation of the oscillator. This is precisely the source of energy that produces the fluctuations of the charge distributions, thereby inducing the temporary dipole moments.

To distinguish this type of interaction between atoms and non polar molecules, the term <u>London force</u> or <u>dispersion force</u> is used. Ultimately, the chemical binding in non polar molecular crystals is a purely quantum effect. A comprehensive but very complex theory of molecular bond formation in solids was formulated by Lifshitz in the 1950s. The study of fluctuations due to zero point energy is a very advanced topic of the so called quantum field theory. Fluctuations of this kind are present even in vacuum and give rise to an expected and then observed physical effect, the Casimir effect.

The course programme

- To start we will take care of the "crystal structures" in which the atoms are arranged in an orderly manner. We will see how this type of order can be described and classified in more precise terms and how it can be determined experimentally.

- We will then move on to deal with how the atoms of the solid move around the normally occupied positions of equilibrium, giving rise to the so called "lattice oscillations".

- There are many atoms and therefore the collective behaviour of the solid can only be understood by means of statistical physics. This will help us to understand the thermodynamic and thermal transport properties.

- Valence electrons are those that give rise to chemical bonds and therefore to most of the physical properties of condensed matter. For this reason the study of the states of these electrons is so important and we will dedicate a part of the course to them.

- In some cases, electrons may move within solids to produce electric currents. This is clearly a subject of enormous importance for the infinite applications that result from it, and this will also be discussed in the course. We will also deal diffusely with a very special and strange case of electronic transport, superconductivity.

Based on a knowledge of these arguments which constitute the fundamentals of solid state physics, in successive courses it will be possible to address more specific topics such as magnetic properties, optical properties, semiconductor devices, the role of defects, the role of surfaces, ...

The aim of this course, however, is not to go into the details of specific aspects, but rather to provide the conceptual basis for understanding the details of the individual topics at a later stage. We will then see that some of the models we develop fall within the scope of what is called "Quantum Field Theory" (QFT). The Lifshitz theory mentioned above was one of the first examples of application of QFT to condensed matter. We will talk about lattice oscillations and superconductivity.

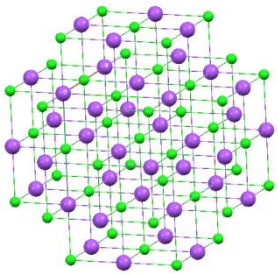
The quantum field theory itself is the one on which all modern particle physics is based, in particular the so called "standard model". That is, there is a substantial methodological unity in the most advanced theories of condensed matter and those of elementary particles. In formulating the physics of solids as we will do in this course we will try to give at least a taste of all this.

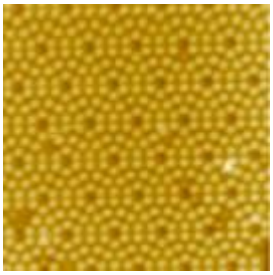
1. Crystal structures in the real space

1.1 Crystal structures: lattice + basis

Until now it has been possible to formulate relatively simple physical theories, but nevertheless adequate to understand many properties of condensed matter, only if the constituent atoms are arranged in an ordered manner, in which case we speak of crystalline solids or crystal structures. In these materials, the microscopic order is also reflected at the macroscopic level, as is the case for many minerals that appear as regular geometric figures. In fact, many of the concepts that serve to classify the structures of solids, that is, the way in which atoms are disposed on a regular basis, were conceived by mineralogists when the existence of atoms was still controversial.

In a crystal structure a given figure, called basis, repeats itself indefinitely and periodically in space. We can think that there is an underlying lattice, which defines the periodicity, at each point of which a copy of the basis is fixed. To understand what this means, see the two examples in the figure. The left image shows the 3D crystal structure of sodium chloride. The basis that repeats itself indefinitely is constituted by a couple of different atoms (Na and Cl). The other image shows the 2D surface of a Si crystal, in the so called "7x7 reconstruction". This image was acquired with a scanning tunnel effect microscope (STM) and the basis is a more complex figure than in the case of sodium chloride.





Red o Red Cristalina en este contexto
A lattice is an infinite, periodic arrangement of points in space generated by a set of discrete translation vectors.

In real solids, order cannot extend to infinity because they have a finite dimension and there is always a surface. The physics we are going to deal with is that of solids large enough to be able to neglect the effects due to the surface, thus taking into account only the effects that manifest in a supposed infinite solid. These effects are called volume effects or bulk effects. p

This does not mean that surfaces are not important: chemical reactions, catalysis, optical properties of metals, ... are all surface properties. Another situation in which surface effects cannot be overlooked is that of nanometric objects.

The reason for the relative simplicity of periodic structures is that for them the properties of the whole system can be traced back to that of the basis only. But unlike molecular physics, which studies a single object of finite size, in crystal structures there are endless bases next to each other. It is therefore necessary to take into account how the individual properties of a basis "connect" to those of the other bases. As we will see, this is dealt with by a fundamental theorem, the Bloch theorem. When this periodicity is lacking, as for example in amorphous solids, then everything becomes much more complicated.

Why do atoms in many materials arrange themselves in an ordered manner, resulting in a certain structure? Strange as it may seem, it is not possible to answer this question because, in order to do

this, we should be able to calculate a priori the structure that minimizes the energy of the system, and this is not yet possible, despite the enormous progresses of the methods of calculation. To do this we should be able to calculate the energy of formation of different possible structures so as to choose the one that represents the fundamental state. However, this would reveal that the difference in energy between the different structures is generally very small, less than the errors due to the approximations of the models adopted. The fact that there are different structures with comparable energy explains why many materials are presented in different allotropic forms where the atoms are the same but the structure can change under the effect of external factors such as pressure or temperature.

An interesting observation is that atoms often come together in order to create the maximum possible packaging, i.e., to place as many atoms as possible in the available volume. In mathematics, the Kepler conjecture has to do with the packaging of spheres in the usual three dimensional Euclidean space. It states that there is no way to place spheres in space with an average density higher than that of cubic packing with centered faces or that of hexagonal packing. The density of these two ways of arranging the spheres is the same and is slightly higher than 74%. Indeed, the majority of the elements crystallize just in one of these two ways.

Incidentally, the Kepler conjecture has only been satisfactorily demonstrated in very recent times using computers. The first formulation of the demonstration was presented in 1998 but its verification by the scientific community took about 15 years and was fully accepted only in 2014.

1.2 Symmetry and crystal structures

We are talking about periodic structures so let us try to better define what is meant by periodicity from the physical point of view. To do this, we use a simple example of a periodic structure, an infinite floor in which the basis is made up of the single tile. In any tile where we decide to stay, the image of the floor we observe is always the same and it is not possible to determine where we are since all the positions are equivalent to each other.

We can see it differently. Instead of moving from one tile to another, we can think of staying in the same place in space and move the entire floor. It is easy to understand that in reality the two operations are equivalent, what counts is the relative movement between the observer and the infinite periodic structure.

Symmetry operations are generally defined as manipulations that, after being carried out, lead to a situation that is completely indistinguishable from the initial situation. In the case of molecules, there are only the so called point operations in which at least one point remains fixed and all the others move. These include, for example, rotations around an axis (in which all the points on the axis remain stationary), or mirroring with respect to a plane (in which case all the points on the plane remain fixed). In the case of solids, in addition to point operations, there are also translations, which are only possible for systems of infinite dimensions.

There is a formal theory of symmetry of periodic structures based on group theory. We will not deal with this, but we will limit ourselves to a quick overview aimed, above all, at introducing a series of terms that are widely used. The starting point is that, given a structure, there is a certain set of symmetry operations that characterize it. The symmetry operations that can be performed on a crystal are:

1) Translations of the entire structure, in which each point of the structure undergoes the same displacement.

2) Point operations, which leave fixed at least one of the points of the structure.

- Inversion with respect to a point that does not necessarily have to be one of the atoms of the structure but can be an arbitrary point of space. In practice, this means bringing all the points with coordinate **r** with respect to the inversion point to points with –**r** coordinates. If the structure obtained in this way coincides with the initial one, the system is said to possess an inversion center.

- Rotations of the whole structure, considered as a rigid entity, around an axis.

- Reflections with respect to a plan.

3) Mixed operations that can be expressed as the sum of reflections, rotations and translations (which, if carried out individually, would not be symmetry operations):

- Reflection + translation (glide).

- Rotation + translation (screw).

Given a particular structure, the set of symmetry operations that can be performed is called the spatial group of the structure. It is easy to demonstrate that this set of operations constitutes, from an algebraic point of view, a group. The reason is that symmetry operations quite obviously satisfy all the conditions that serve to define a group:

- Applying two symmetry operations subsequently results in another symmetry operation, and this defines the product of two elements of the group.

- The product has associative properties (but not necessarily commutative properties).

- There is the identity operation E (doing nothing to the structure).

- For each operation R of the group there is its inverse R^{-1} defined as follows: $RR^{-1} = E$

The point group of a structure is the subset of the spatial group that includes only point operations, i.e. those that leave at least one point of the structure fixed.

1.3 Translational symmetry: the Bravais lattice

The lattice is the set of points at which the basis is fixed to give rise to the periodic structure. We can define a lattice in several equivalent ways:

Def. 1 (physical) A lattice is a set of points (also called nodes) such that each one of them has the same configuration as far as the first neighbors, i.e. the nodese closer to it, are concerned.

Def. 2 (mathematical) A lattice is a set of points obtained by means of the formula:

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

The n_i are integers with sign, the vectors $\mathbf{a}_i$ (not complanar, i.e. linearly independent) are the primitive vectors of the lattice and it is said that they generate the lattice itself. Given a lattice,

$a_i \cdot a_j = 0 \quad (i \neq j)$

the choice of the primitive vectors is not univocal, we can use arbitrary triads of vectors, provided they are not complanar.

Def. 3 (based on symmetry operations) Instead of considering the vectors **R** that identify the points of the lattice, we consider a lattice and we think that the vectors **R** represent rigid translations of the lattice itself that lead the lattice to overlap itself. Obviously, these are symmetry operations. If we consider the combined effect of two translations, we obtain another translation, therefore the set of translations is closed with respect to the sum operation. In addition, for each translation **R** there is the opposite translation, that we indicate with −**R**. If we make the two in succession we get the equivalent of a zero translation. In this way we can also define the difference operation of two translations. From the definition of the vectors **R** that identify the translations emerges a further way of defining a lattice:

> *a discrete set of nonplanar vectors closed with respect to the operations of sum and difference.*

This definition is quite abstract and to make it more understandable let us try to think of a blank sheet on which an origin is fixed. On the same sheet we draw a series of points with the idea that they represent vectors referred to the chosen origin. If the points are drawn randomly, taken two of these vectors hardly their sum will coincide with one of the other points already present. Only if the points are those of a Bravais lattice this can happen. This definition will be useful when we will discuss diffraction.

The lattices thus defined are called Bravais lattices, from the name of the French scientist (physicist, mineralogist and meteorologist) Auguste Bravais who, in the '800, introduced them to classify the forms of minerals.

In the following we will address the problem of enumerating the possible Bravais lattices in 1, 2 and 3 dimension. Before doing so, however, we will introduce the concept of elementary cell.

1.4 The elementary cell

Given the points of the lattice it is possible to fill the whole space by choosing a suitable geometric figure and connecting it in the same way to all the points of the lattice. In this case the selected region is called a cell.

Usually the unit cell, or elementary cell, is considered. A unit cell shifted over all the points of the lattice fills the whole space without overlapping. There are infinite ways to choose the unit cell, but obviously the volume of all the possible cells is the same as they must fill the whole space without overlapping. A very simple elementary cell consists of the parallelepiped defined by three primitive vectors, the primitive cell. A case of great importance is the Wigner-Seitz cell which, as we shall see, preserves the symmetries of the basis. The W-S cell centered on a particular point of the lattice is defined as *the set of points of space closest to the point itself*. We will see some examples to understand this definition.

In some cases, instead of the unit cell, a larger one called the conventional cell is used: it is a cell chosen for reasons of convenience for the purpose of describing the structure and it is larger than the unit cell and therefore, by translating it to the other points of the lattice, there are overlaps.

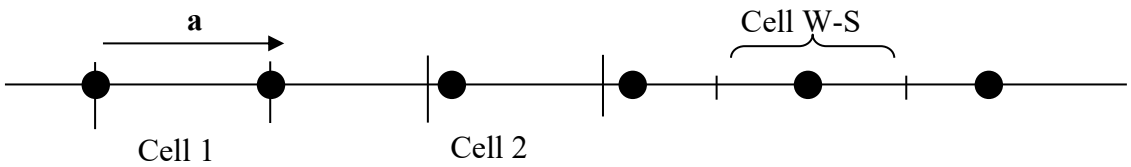
The elementary cell should not be confused with the basis. The first is a portion of space that, repeating itself with the periodicity of the underlying lattice, must fill the whole space without overlapping, while the basis is a physical object, a set of atoms, that repeats itself in each of the points of the lattice.

1.5 Possible Bravais lattices

Once the elementary cell has been introduced, it becomes possible to establish a criterion for identifying how many and which Bravais lattices are possible. For simplicity let us start with the trivial case in 1D and then move on to those in 2D and 3D.

1D lattices

The easiest way to cover a whole line without overlapping with a simple geometric figure is to use a segment of length a as an elementary cell. After having "tiled" in this way the whole line we can trace the corresponding Bravais lattice by choosing, for each of the cells, a point at will, obviously the same for all cells. In this simple case it is immediately understood that the underlying lattice is a set of points separated by the same distance a and generated by the formula $\mathbf{R} = n\mathbf{a}$ starting from an arbitrary origin. The primitive vector or generator of the lattice **a** of this type of lattice is uniquely defined (not so in 2D and 3D).



If the point of the segment used to connect the basis to the points of the lattice is the one that divides them into two equal parts, then the corresponding elementary cell is that of W-S.

2D lattices

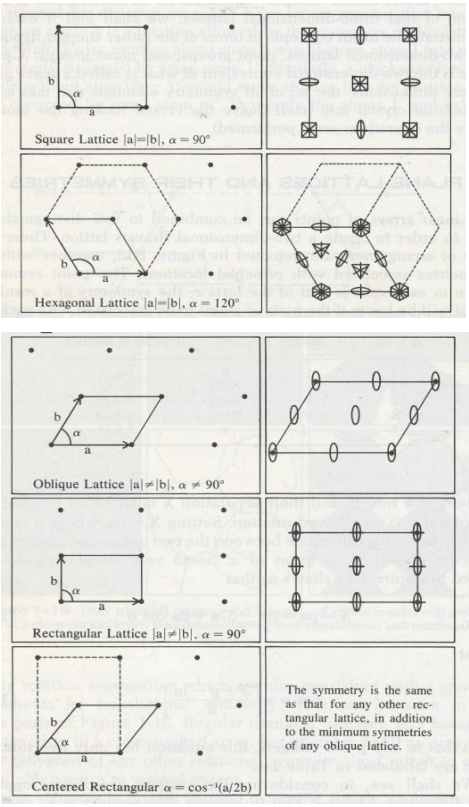
Here, too, we start from the question: how is it possible to cover the entire surface without superpositions? As we can easily guess, if we consider regular figures, the thing can be done with squares and equilateral triangles. Choosing, as already done in 1D, a point at will of the cell we build the corresponding lattices that, for the two cells in question, are the square one and the triangular one, usually called hexagonal.

In addition to simple squares, it is possible to obtain the same result, i.e. to fill the whole space without overlapping, using a cell obtained by deforming the square.

- If we change the length of one of the two sides we get a rectangle, which corresponds to the rectangular lattice.

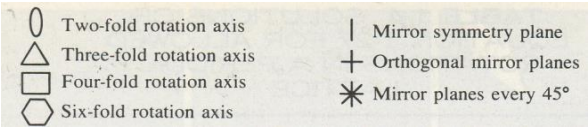
- If we tilt one of the sides we get a rhombus, which generates the oblique lattice.

- A particular case of oblique lattice is the one that is obtained starting from a rectangular lattice and putting at the center of each of the rectangles a further point, in which case we have the rectangular lattice with <u>centered faces</u>.



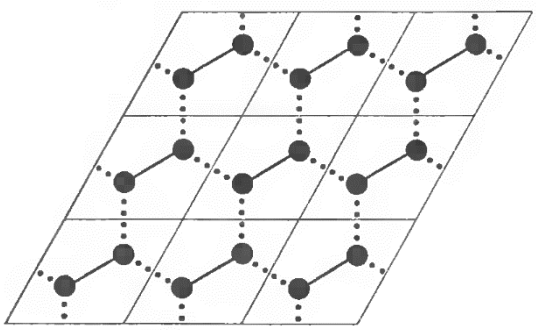
The 2D lattices can be constructed starting from a square and a triangle, by means of which it is possible to cover the whole plane

Other three lattices are built by deforming the square first making it a rectangle and then by tilting it.



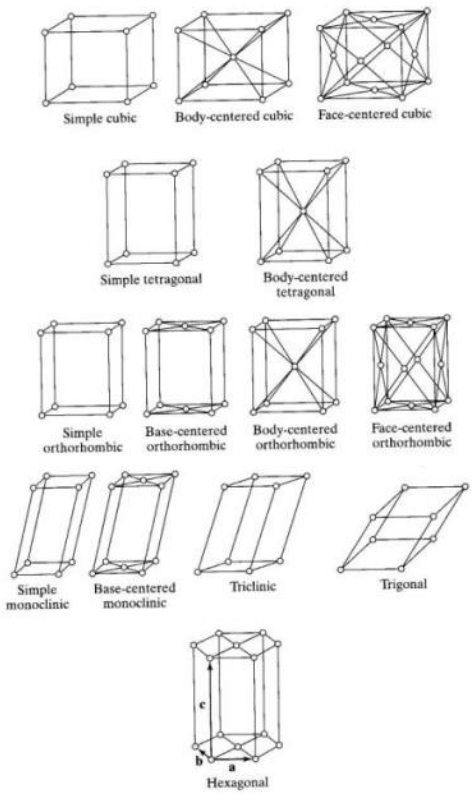
Starting from a square then we can generate 4 distinct Bravais lattices, one using the same square and the other three by deforming it. The triangle, on the other hand, cannot be deformed (doing so would result in one of the lattices generated from the square) and therefore only one lattice is constructed from it, the <u>hexagonal</u> one. In total, therefore, the number of possible Bravais lattices in 2D is 5.

<u>The honeycomb</u> structure is a set of hexagons placed side by side so as to fill the entire surface. If we consider the vertices of these hexagons we get a set of points that, although periodic, is NOT a Bravais lattice. In fact, the points are not equivalent because the configurations of the first neighbors are not the same for all. It is possible to build the honeycomb structure starting from an hexagonal lattice and decorating it with a basis consisting of two points.



3D lattices

In 3D there are 14 Bravais lattices. To understand their origin we follow the same reasoning done in 2D and we try to find a way to fill the whole space with regular figures without overlapping. In 3D the possible figures are the cube and the prism with an equilateral triangular basis. Also in this case we think to progressively deform the cube (which we can't do with the triangular prism, just like for the 2D triangle). In this way we obtain the 14 lattices shown in the figure.



1) Simple cube, which gives rise to the cubic system, from which descend three distinct Bravais lattices (simple, body-centered and faces-centered).

2) If we vary the length of one of the three sides we get a rectangular prism, which generates the tetragonal system. Two associated lattices (simple and body-centered).

3) If we change the lengths of all three sides, obtaining a prism, we have the orthorhombic system, with four associated lattices.

4) By making one of the corners different from 90° we have the monoclinic system, with two associated lattices.

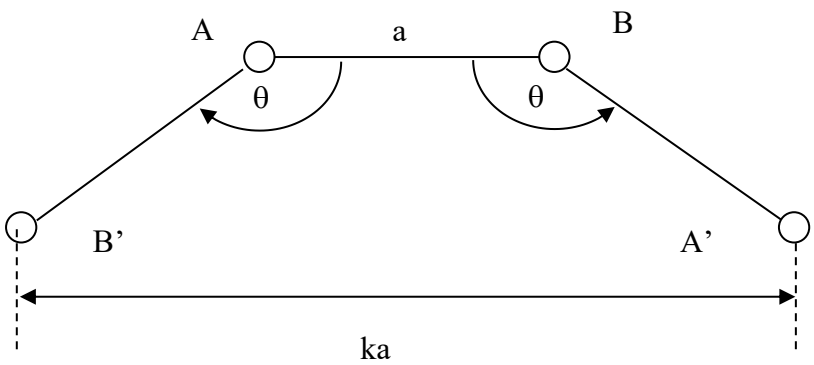
5) If all corners are made different from 90° we have the triclinic system. In this case there is only one associated lattice.

6) If the cube is deformed along the diagonal we have the trigonal or rhombohedral system. Only one associated lattice.

7) If we start from a triangular prism we have the hexagonal system. Only one associated lattice.

1.6 Possible rotations of a Bravais lattice

In a crystal structures, as we have said, translational and point symmetry operations can be defined. Before starting to think in more detail about the symmetries of crystal structures, we demonstrate a simple theorem of great importance, according to which only rotation axes of order 0, 1, 2, 3, 4 and 6 are possible for Bravais lattices. Let us consider the figure in which two points, A and B, of a Bravais lattice are shown. First we think to operate a rigid rotation of the lattice around an axis perpendicular to the sheet passing through A, being θ the angle of rotation. After rotation, point B moves to point B'. Now let us do the same rotation but in the opposite direction and around an axis passing through B (the symmetry operations admit that there is their inverse, so for a rotation there must also be the one in the opposite direction). In this case, point A, after the rotation, is at point A'.



The condition of invariance of the structure following rotations is that points A' and B' are in turn points of the lattice. Their distance along a direction parallel to the conjunction of A and B must therefore be an integer multiple of the distance between A and B because it must be possible to pass from A' to B' by steps of length a. Translated into a formula:

$$a + 2a\sin\left(\theta - \frac{\pi}{2}\right) = a - 2a\cos(\theta) = ka$$

Where k is an integer number (with sign). The condition is that:

$$\cos(\theta) = \frac{1-k}{2}$$

However, we know that $|\cos(\theta)| = \left|\frac{1-k}{2}\right| \leq 1$ and this means that k can only assume the values 0, 1, 2 and 3 (k = −1, which corresponds to θ = 0, i.e. identity, is also fine). These values of k correspond, respectively, to rotation angles of 60, 90, 120 and 180° and therefore rotations of order n = 6, 4, 3 and 2, QED. Then we must add as a trivial case also the rotations of order 1, that is 360°, which is equivalent to the operation identity.

1.7 The symmetry operations of a Bravais lattice (the crystal systems)

Let us first consider the symmetry operations possible for a simple Bravais lattice (i.e. a lattice without a basis). Obviously, all Bravais lattices are equipped with translational symmetry and inversion. This is why we are going to consider only rotations and mirroring.

1D

Since in one dimension there is only one possible Bravais lattice, there is also only one group of symmetry operations. In this case, to use the language of crystallography, it is said that there is only one crystal system. A crystal system is a *distinct group of symmetry operations that can be applied to the existing Bravais lattices*. In the 1D case the thing is trivial and not very significant because there is only one possible lattice, so let us wait to see what happens in two or three dimensions where, as we have seen, the possible lattices are, respectively, 5 and 14.

To list the symmetry operations possible on the only existing Bravais lattice in one dimension, we observe that, in this case, it makes no sense to talk about rotations and therefore the only possible operations are inversion and specular reflections. The points around which we can perform these operations are the points of the lattice itself and the intermediate points between a point of the lattice and the other. In reality, in one dimension the two operations of inversion and mirroring are actually coincident.

It is clear that the W-S cell possesses the same symmetry operations of the lattice, in this case only inversions. All the other possible cells are not symmetrical by inversion (around the point of the lattice to which they are attached). This property of the W-S cell, having the same symmetries as the lattice to which it refers, also occurs in 2D and 3D, and this is precisely the reason for its importance.

2D

Let us refer to the figure shown above, which illustrates the 2D lattices.

1) The oblique lattice, with fundamental vectors inclined at an arbitrary angle and $a_1 \neq a_2$. It allows only order 2 rotations.

2) The rectangular lattice with perpendicular vectors and $a_1 \neq a_2$. Admits order 2 rotations and perpendicular mirroring planes. The centred rectangular lattice also has the same symmetries. As we can see, there is a group of symmetry operations that is shared by two different Bravais lattices. This group defines the rectangular crystal system.

3) The square lattice with perpendicular vectors and $a_1 = a_2$. It allows rotations of order 2 and 4 and mirroring planes every 45°. This group defines the square crystal system.

4) The hexagonal lattice with vectors at 60° and $a_1 = a_2$. It allows rotations of order 2, 3 and 6 and various planes of mirroring. This group defines the hexagonal crystal system.

The 5 possible lattices therefore correspond in reality only to 4 distinct groups of symmetry, since two of these lattices share the same symmetry operations. In 2D, therefore, there are 4 crystal systems.

3D

In 3D there are 7 crystal systems (cubic, tetragonal, rhombic, monoclinic, triclinic, trigonal, hexagonal, see the previous discussion on Bravais lattices in 3D). The demonstration is obviously much more complicated than the 2D case.

1.8 Possible symmetries of the basis

So far, we considered a lattice without basis. In this way we could only deal with the symmetries of the lattice. The conclusions reached are obviously valid even if we consider the presence of a basis with spherical symmetry because this kind of basis possesses all the symmetries considered for the lattices.

If instead the basis is there and it is not trivially spherical, then it is right to ask oneself what happens to the symmetry of the structure constituted by the lattice and by the basis. To understand the role of the basis, we begin by considering only those operations that keep a point on the lattice fixed. The set of these operations is the crystallographyc point group. Since this is a basis connected to a lattice, the symmetry operations to be considered are only those that are compatible with translational symmetry, that is to say:

- Rotations of order 1 (only 360° rotations, no rotational symmetry), 2, 3, 4 and 6

- Reflections

- Inversions

A trivial case is one in which the basis has no symmetry. In this case, the set of symmetry operations of the system is limited to translation operations and the point group is empty, or rather contains only the identity.

1D lattices

In 1D rotations do not make sense, so they are not taken into account. In addition, the mirror reflection and inversion operations coincide. In this way the number of possible symmetry groups is reduced to only two: one is the one that contains only the identity operation and applies to the bases without symmetry. The other also includes reflections (or reversals) and is valid if the basis is symmetrical for mirroring.

2D lattices

In 2D there are ten crystallographic point groups. To understand why we think that there are five possible types of rotation (of order 1, 2, 3, 4, 6). See figure. There are therefore 5 possible basis types and in turn each of them can possess mirroring symmetry or not. So, we have 5x2 = 10 crystallographic groups.

3D lattices

In 3D there are 32 groups of crystallographic point groups.

Bases without mirroring symmetry	Bases with mirroring symmetry

Possible rotations of order 1,2,3,4,6

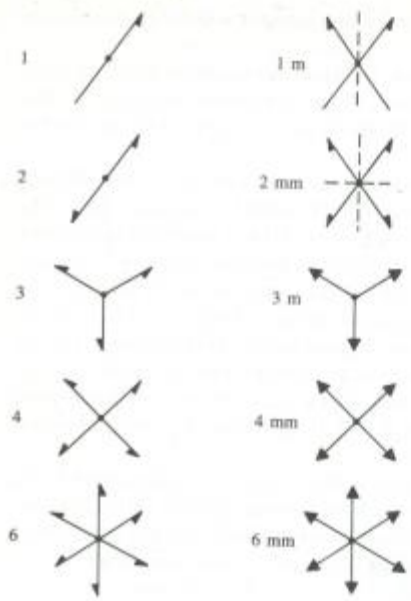


Figure 1-19 The ten types of point group which can be consistent with the symmetries of a plane lattice.

1.9 Point and translation symmetry operations of a structure (lattice + basis)

If the different Bravais lattices are combined with the possible point groups compatible with the translational symmetries (point symmetries), all the possible spatial groups are obtained. The space group is the set of all the possible symmetry operations of a given structure.

In one dimension there is only one Bravais lattice and two spatial point groups, both compatible with this lattice. So, there are two space groups.

The table shows, for each of the crystallographic systems, the point groups compatible with it in 2D (for the meaning of the acronyms identifying the point operations, see the previous figure). The total number of space groups resulting from this is 17.

TABLE 1-5 SUMMARY OF TWO-DIMENSIONAL CRYSTAL SYSTEMS, LATTICES, AND PLANE GROUPS

Crystal System	Point Groups Compatible With System	Bravais Lattices Included in System	Number of Plane Groups Compatible With Lattice
Oblique	1, 2	P (primitive)	2
Rectangular	1m, 2mm	P (primitive) C (centered)	5 2
Square	4, 4mm	P (primitive)	3
Hexagonal	3, 3m, 6, 6mm	P (primitive)	5

In 3D there are 230 spatial groups, and to enumerate them all we have to consider the operations of chiral symmetry (rototranslations) only possible in 3D (without these the spatial groups would be 219).

By generalizing the theory of symmetry in 1, 2 and 3D Hilbert formulated the so called problem 18 (one of the 23 he proposed in 1900) on the existence of lattices larger than 3. In 4D there are 52 Bravais lattices. In 1970 it was shown that there are 4783 space groups in 4D.

Another extension of the theory of symmetry, more physically useful, is the one concerning the so called groups of colours or groups of Shubnikov, in which it is considered the possibility that, without moving the lattice, inversions are made of the different "colors" with which the bases are drawn. This has applications in the field of magnetism where the colors represents the different spin of the electrons.

Dimensions	Bravais lattices	Crystal systems	Crystallographyc point groups	Space groups	Color groups
1D	1	1	2	2	
2D	5	4	10	17	
3D	14	7	32	230	1651
4D	52			4783	

1.10 Physical consequences of symmetry

Pyroelectricity

When a material is heated, electrical charges are generated on both sides of it. Also, given a certain face, on that one always appears charge with the same sign. This means that the system is not symmetrical by inversion. For this reason, pyroelectric crystals have no inversion centre.

Piezoelectricity

Deforming a solid gives rise to a potential difference between its ends, which corresponds to a polarization charge that, like in the case of pyroelectricity, has always the same wign on a given face of the sample. Also in this case, therefore, there is no inversion symmetry.

Optical activity

Optical activity consists in the rotation of the polarization plane when light passes through a solution or a crystal. For this to happen, the material must be "chiral", i.e. its structure and mirror image must not be superimposable. Considering chiral molecules, for example, they exist in two distinct forms (called "enantiomers"), one of which is the mirror image of the other.

When polarized light passes through a solution of a chiral compound, there is a rotation to the right or left of the polarization plane, depending on which of the two enantiomers it is.

Molecules of this type can crystallize into chiral crystals. Crystals of this type are characterized by symmetries of rototranslation and give rise to phenomena of optical activity. Another well-known case of optically active material is that of quartz, which exists in two forms called alpha and beta quartz, with opposite chirality.

2. Crystal structures in the reciprocal space and diffraction

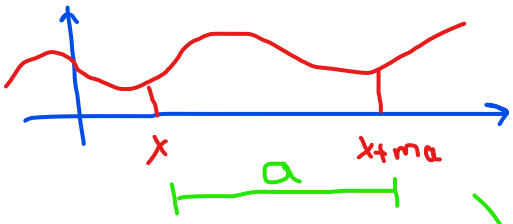
A concept of great importance in the study of periodic structures is that of reciprocal lattice. At first glance it is a rather abstract concept but in reality it has a very precise physical meaning, which we will try to highlight as soon as possible. The starting point of the whole discussion is the possibility of expressing a periodic function by means of a Fourier series. Let us first see the basic concepts in the 1D case.

2.1 Fourier series

Fourier series development of real functions

Let us consider a real function of a real variable f(x). We will say that it is periodic if:

$$f(x + l_m) = f(x)$$



for all the l_m values so defined:

$$l_m = ma$$

where m is an integer (with zero and sign, therefore m = 0, ±1, ±2, ...) and a is the period. The periodic functions in 1D are associated with a lattice in real space. In fact, given the periodic function f(x), let us choose an arbitrary point x: all the other points that are at a distance ma from x, where f(x) always takes the same value, evidently constitute a lattice.

Given a periodic function it is possible, and extremely useful, to express it as a summation of trigonometric functions, what is called Fourier series development in series. The expression generally used is the following:

$$f(x) = \frac{a_o}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{2\pi nx}{a} + b_n \sin\frac{2\pi nx}{a}\right)$$

$\frac{2\pi nx}{T}$ → $T = a$

This formula tells us that the function is expressed as the sum of many components, each of which is a trigonometric function of the sine or cosine type, that is, a wave in space if x represents a spatial coordinate. Not all the possible waves, however, appear in the development, only those whose period is equal to that of the function f(x) or a submultiple of it. This is why the argument of the trigonometric functions is written in the form 2πnx/a. It is also present a constant term that could be included in the summation by starting the index n from zero as is done in some texts. Here we will leave it in evidence because it is the most widespread practice.

A rigorous theory of the Fourier series is quite complicated. It is possible to demonstrate that all the functions that are not "strange" from a mathematical point of view can be expressed in this way, and in the physics of solids there are no strange functions therefore we can count on the fact that the development in series of Fourier always works.

The problem to be solved is to calculate the a_n and b_n coefficients of the development. To do this let us evaluate the following integrals:

$$\int_o^a f(x)dx; \quad \int_o^a f(x)\cos\frac{2\pi nx}{a}dx; \quad \int_o^a f(x)\sin\frac{2\pi nx}{a}dx$$

By inserting in place of f(x) its development in series, and remembering that the sine and cosine functions constitute a family of <u>orthogonal</u> functions, that is that they enjoy the following property, we obtain:

$$\int_o^a \cos\frac{2\pi n_1 x}{a}\ \sin\frac{2\pi n_2 x}{a}dx = 0$$

$$\int_o^a \cos\frac{2\pi n_1 x}{a}\ \cos\frac{2\pi n_2 x}{a}dx = \frac{a}{2}\delta_{n_1,n_2}$$

$$\int_o^a \sin\frac{2\pi n_1 x}{a}\ \sin\frac{2\pi n_2 x}{a}dx = \frac{a}{2}\delta_{n_1,n_2}$$

In this way, the values of the coefficients of the series expansion are immediately obtained:

$$\begin{cases} a_o = \dfrac{2}{a}\displaystyle\int_o^a f(x)dx \\ a_n = \dfrac{2}{a}\displaystyle\int_o^a f(x)\cos\frac{2\pi nx}{a}dx \\ b_n = \dfrac{2}{a}\displaystyle\int_o^a f(x)\sin\frac{2\pi nx}{a}dx \end{cases}$$

Fourier series development of real functions using complex coefficients

It is possible to generalize what has been seen up to now for the real functions using complex numbers. To do so, we start with the Euler formula:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

By reversing the sign of the exponent we obtain $e^{-i\theta} = \cos\theta - i\sin\theta$. Putting the two versions of the Euler formula into a system gives rise to:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

By replacing these expressions for the cosine and sine of an angle in the Fourier expansion seen previously for the real functions, one arrives at expressing in this way the function f(x):

$$f(x) = \frac{a_o}{2} + \sum_{n=1}^{\infty}\left(a_n \frac{e^{\frac{i2\pi nx}{a}} + e^{-\frac{i2\pi nx}{a}}}{2} + b_n \frac{e^{\frac{i2\pi nx}{a}} - e^{-\frac{i2\pi nx}{a}}}{2i} \right) =$$

$$\frac{a_o}{2} + \sum_{n=1}^{\infty} \frac{a_n - ib_n}{2} e^{\frac{i2\pi nx}{a}} + \sum_{n=1}^{\infty} \frac{a_n + ib_n}{2} e^{-\frac{i2\pi nx}{a}}$$

The two summations thus obtained can be combined into a single one. To do this we redefine the index n in the second summation by making it to vary from negative infinity to 1, in which case the single terms of the summation are no longer indexed by the value n but by the value −n. Moreover, to take into account the fact that instead of the index n we use −n, we change sign to the exponent and in this way in the two sumations the exponential terms become the same:

$$f(x) = \frac{a_o}{2} + \sum_{n=1}^{\infty} \frac{a_n - ib_n}{2} e^{\frac{i2\pi nx}{a}} + \sum_{n=-\infty}^{1} \frac{a_{-n} + ib_{-n}}{2} e^{-\frac{i2\pi nx}{a}}$$

At this point it is possible to collect the common exponential term and write a single summation, with an index n that runs from minus to plus infinity, including zero, in which each exponential term is multiplied by a coefficient that we indicate with A_n:

$$f(x) = \sum_{n=-\infty}^{\infty} A_n e^{\frac{i2\pi nx}{a}}$$

Comparing this way of expressing the series with the one seen immediately before and based on the a_n and b_n coefficients, we immediately obtain the link between these and the complex coefficients A_n:

$$A_n = \begin{cases} \frac{1}{2} a_o & n = 0 \\ \frac{a_n - ib_n}{2} & n > 0 \\ \frac{a_{-n} + ib_{-n}}{2} & n < 0 \end{cases}$$

There is a precise condition relative to the complex coefficients A_n so that the resulting function f(x) is real. Let us consider, for each value of the index n, the term corresponding to it and that corresponding to the value −n and add them together:

$$A_n e^{\frac{i2\pi nx}{a}} + A_{-n} e^{-\frac{i2\pi nx}{a}} = A_n \left[\cos\left(\frac{2\pi nx}{a}\right) + i\sin\left(\frac{2\pi nx}{a}\right) \right] + A_{-n} \left[\cos\left(-\frac{2\pi nx}{a}\right) + i\sin\left(-\frac{2\pi nx}{a}\right) \right]$$

Using the symmetry properties of the cosine (even) and sine (odd) functions we write:

$$A_n e^{\frac{i2\pi nx}{a}} + A_{-n} e^{-\frac{i2\pi nx}{a}} = (A_n + A_{-n})\cos\left(\frac{2\pi nx}{a}\right) + i(A_n - A_{-n})\sin\left(\frac{2\pi nx}{a}\right)$$

For the whole to be a real number the sum $(A_n + A_{-n})$ must be a real number while $(A_n - A_{-n})$ must be a pure imaginary. This is true if the two coefficients are one the complex conjugate of the other, a condition that is called <u>hermitian symmetry</u>:

$$A_n^* = A_{-n}$$

If we define the vector g_n:

$$\boxed{g_n = n\frac{2\pi}{a}}$$

The series development can be written in this even more compact form:

$$\boxed{f(x) = \sum_{g_n} A_n e^{ig_n x}}$$

In this way the function is expressed as the sum of many exponential functions $e^{ig_n x}$, each characterized by a g_n <u>wave vector</u> and weighed with an A_n coefficient.

<u>Fourier development of complex functions</u> ?

The above formula expresses real functions if the hermitian symmetry condition for the A_n coefficients is met. If this is not verified, the series still makes sense but obviously the function will, in general, be complex.

To obtain the value of the coefficients in the general case of a complex function, we proceed as already done in the case of real functions by calculating the integral:

$$\int_o^a f(x)e^{-ig_m x}dx = \int_o^a \left(\sum_{g_n} A_n e^{ig_n x}\right) e^{-ig_m x}dx$$

and exploiting the condition of orthogonality that in this case takes the following form:

$$\int_o^a e^{ig_n x} e^{-ig_m x}dx = a\delta_{n,m}$$

The expression for the complex coefficients is therefore the following:

$$A_n = \frac{1}{a}\int_o^a f(x)e^{-ig_n x}dx$$

To put it in the language of functional analysis, the function f(x) is a vector in a functional space and the coefficients A_n are the projections of the function f(x) on the functions $e^{ig_n x}$ that, all together, constitute a complete ortonormal set. Complete means that through them it is possible to represent any function f(x) that meets appropriate conditions, always verified in the cases that solid state ?

physics deals with. This set is the equivalent, in general functional spaces, of the orthogonal vector triads normally used to express vectors in the Euclidean space.
The generalization based on the use of complex coefficients allows to express the developments in series of Fourier in a more elegant and compact form, certainly more suitable for calculations. It also allows us to use Fourier series developments even when dealing with complex functions. Finally, it lends itself to using the Fourier method even with non periodic functions, leading to the definition of the Fourier integral, better known as the Fourier transform. This is possible by stretching the period a to infinity.

2.2 The reciprocal lattice in 1D

The set of possible wave vectors g_n evidently constitutes a further lattice in 1D. The lattice of the l_m points defined in real space is called the direct lattice. The g_n points lattice is called the reciprocal lattice of the direct lattice. Three properties are easily verified:

1) The reciprocal lattice is in turn a Bravais lattice (as seen from the formula that defines the vectors g_n, which is precisely the definition of a Bravais lattice in 1D). All the arguments about periodic structures are therefore valid, including the fact that it is possible to define the elementary cell of the reciprocal lattice. If the Wigner-Seitz cell is chosen it is called the Brillouin zone. If we consider the cell centered on the point k = 0 then we speak of the first Brillouin zone (1ZB) even if the adjective is often omitted and in this case we simply speak of the Brillouin zone. In 1D obviously the 1BZ is a segment of length equal to $2\pi/a$ centered on one of the g_n. (The fact that we are talking about "first" zone makes us think that there are others and in fact it is so and we will see it when we will study the electron states).

2) If we consider the lattice constituted by the points g_n, its reciprocal is constituted by the points l_m of the real lattice, as it is easily verified.

3) The distance (equal to $2\pi/a$) between the points of the reciprocal lattice is inversely proportional to the period a of the real lattice.

These properties continue to apply when we will generalize by considering periodic functions in 2D or 3D.

Another way, more elegant and useful for the complex case of 3D lattices, to define the vectors of the reciprocal lattice is as follows. The waves e^{igx} by means of which the periodic function f(x) is expressed, must obviously possess the same periodicity as the function f(x):

$$e^{ig(x + l_m)} = e^{igx}\, e^{igl_m} = e^{igx}$$

In this expression l = ma is any vector of the direct lattice. As a result, we have that $e^{igl_m} = 1$. The reciprocal lattice, in this way, is defined as the set of vectors g that satisfy the condition

$$e^{igl_m} = 1$$

for every possible value l_m of the direct lattice. We can verify that this definition is consistent with what we already know. For it to be satisfied it must be gl_m = integer * 2π. As we have seen before l_m = ma and g = n * $2\pi/a$ and therefore gl_m = (n * $2\pi/a$) * ma = nm * 2 π = integer * 2 π.

2.3 The reciprocal lattice in the general case

A function f(**r**) defined in the Euclidean space is periodic if:

$$f(\mathbf{r}+\mathbf{R}) = f(\mathbf{r})$$

for all vectors **R** constituting a Bravais lattice:

$$\mathbf{R} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$$

The m_i are integer numbers. The vectors **r** instead indicate the position of any point in space and we write them like this:

$$\mathbf{r} = r_1\,\mathbf{a}_1 + r_2\,\mathbf{a}_2 + r_3\,\mathbf{a}_3$$

Unlike in the previous case, the r_i coefficients are real numbers. We do not call them x, y and z because it is not said that the three vectors $\mathbf{a}_i$ of the basis constitute an orthogonal triad, in the general case they are oblique (as long as they are not complanar, of course).

Also in this case it is interesting to develop the periodic function in Fourier series:

$$f(\mathbf{r}) = \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}}$$

The problem is to understand which are the waves (or rather their wave vectors **g**) that have to enter in the development.

In the following we will extend what we have already seen for the one-dimensional lattice to the case of lattices in two or three dimensions, starting with the simplest case, that of the so called orthogonal lattice, in which the vectors of the basis are perpendicular to each other. We will then consider the more general case of a non orthogonal lattice.

<u>Orthogonal lattice</u>

For simplicity let us consider first of all the case of an orthogonal lattice, that is when the basis vectors that define the elementary cell of the Bravais lattice are mutually orthogonal. The development in series of f(**r**) is made by adding plane waves that propagate along the three directions of the vectors $\mathbf{a}_i$ and whose wave fronts are planes perpendicular to each of these directions.

Let us first express function f(**r**) as the sum sum of plane waves directed as $\mathbf{a}_1$, by reducing therefore to the 1D case. We get:

$$f(\mathbf{r}) = \sum_{g_{n_1}} A(g_{n_1}, r_2, r_3) e^{i g_{n_1} r_1}$$

Obviously, the function f(**r**) varies periodically along all three dimensions and therefore the waves of this development cannot be simple plane waves. For a fixed value of r_1, i.e. on the front of the wave moving in the direction of $\mathbf{a}_1$, the amplitude must vary according to the coordinates r_2 and r_3 that

identify the position on the front itself. This variability is highlighted by indicating that the A coefficients also depend on the other two coordinates, r_2 and r_3.

If we consider the development along the r_1 axis only, the problem is one dimensional and the corresponding wave vectors g_{n_1} can be expressed using the formula already obtained previously:

$$g_{n_1} = n_1 \frac{2\pi}{|\mathbf{a}_1|}$$

Where $|\mathbf{a}_1|$ is the period along direction $\mathbf{a}_1$.

However, the function f(**r**) must also be periodic in the other two directions, $\mathbf{a}_2$ and $\mathbf{a}_3$ Therefore the dependence of the A coefficients on r_2 and r_3 is in turn periodic and this allows to develop in series these A coefficients also along these other two directions:

$$f(\mathbf{r}) = \sum_{g_{n_1}, g_{n_2}, g_{n_3}} A(g_{n_1}, g_{n_2}, g_{n_3})\, e^{i g_{n_1} r_1} e^{i g_{n_2} r_2} e^{i g_{n_3} r_3} =$$

$$\sum_{g_{n_1}, g_{n_2}, g_{n_3}} A(g_{n_1}, g_{n_2}, g_{n_3})\, e^{i\left(g_{n_1} r_1 + g_{n_2} r_2 + g_{n_3} r_3\right)}$$

What appears in the exponent is the scalar product of two vectors. One, with components r_i, is the vector **r** that locates any point in space. The other one, we call it **g**, has components:

$$g_{n_1} = n_1 \frac{2\pi}{|\mathbf{a}_1|}\ ,\ g_{n_2} = n_2 \frac{2\pi}{|\mathbf{a}_2|}\ ,\ g_{n_3} = n_3 \frac{2\pi}{|\mathbf{a}_3|}$$

and it can be expressed in this way:

$$\mathbf{g} = n_1\, \mathbf{b}_1 + n_2\, \mathbf{b}_2 + n_3\, \mathbf{b}_3$$

The three $\mathbf{b}_i$ vectors are directed as the corresponding $\mathbf{a}_i$ and have modulus:

$$|\mathbf{b}_i| = \frac{2\pi}{|\mathbf{a}_i|}$$

These three $\mathbf{b}_i$ vectors form an orthogonal basis in the reciprocal space of the **g** vectors, just as the $\mathbf{a}_i$ are a basis in real space. The vectors **g** expressed by using the basis $\mathbf{b}_i$ that enter into the development in series of f(**r**) constitute, by definition, the reciprocal lattice of the direct lattice formed by the vectors **R**.

We can easily calculate the volume of the elementary cell of the reciprocal space by multiplying the three moduli of the $\mathbf{b}_i$ vectors:

$$V_{rec} = |\mathbf{b}_1| \cdot |\mathbf{b}_2| \cdot |\mathbf{b}_3| = \frac{2\pi}{|\mathbf{a}_1|} \cdot \frac{2\pi}{|\mathbf{a}_2|} \cdot \frac{2\pi}{|\mathbf{a}_3|} = \frac{(2\pi)^3}{|\mathbf{a}_1| \cdot |\mathbf{a}_2| \cdot |\mathbf{a}_3|} = \frac{(2\pi)^3}{V_{dir}}$$

As we can see, the volume of the two cells, in the direct and reciprocal space, are inversely proportional to each other (as in 1D where the periods of the two lattices are inversely proportional to each other).

Also in this case we can define the vectors **g** that enter in the development in series as those vectors that satisfy the condition:

$$e^{i\mathbf{g}\cdot\mathbf{R}} = 1$$

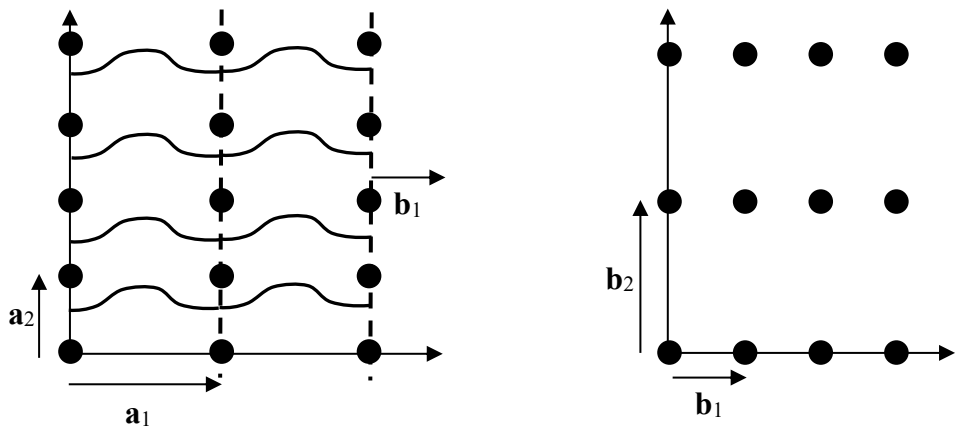
In fact, as in the 1D case, the waves entering into the series development must have the same periodicity **R** of the function f(**r**):

$$e^{i\mathbf{g}(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{g}\cdot\mathbf{r}}\, e^{i\mathbf{g}\cdot\mathbf{R}} = e^{i\mathbf{g}\cdot\mathbf{r}}$$

Hence $e^{i\mathbf{g}\cdot\mathbf{R}} = 1$. For this relationship to be verified the scalar product $\mathbf{g}\cdot\mathbf{R}$ must be a multiple of 2π, which can be immediately verified:

$$\mathbf{g}\cdot\mathbf{R} = n_1 \frac{2\pi}{a_1} m_1 a_1 + n_2 \frac{2\pi}{a_2} m_2 a_2 + n_3 \frac{2\pi}{a_3} m_3 a_3 = \text{integer} * 2\pi$$

To understand the relationship between the direct and the reciprocal lattice we consider a rectangular Bravais lattice as that on the left side of the figure:



If we move along a straight horizontal line, that is, along the direction of vector $\mathbf{a}_1$ of the direct basis, the functions describing the structure are periodic, being the period the modulus of $\mathbf{a}_1$. In the development of the function along this direction enter only waves whose wave vector is an integer multiple of the vector $\mathbf{b}_1$ which, as we now know, is one of those of the basis of the reciprocal lattice, and whose modulus is inversely proportional to that of $\mathbf{a}_1$. Moving instead along the other direction, the one identified by the vector $\mathbf{a}_2$ of the direct lattice, there will be waves whose wave vector is an integer multiple of $\mathbf{b}_2$, the second vector of the reciprocal basis whose modulus is inversely proportional to the one of $\mathbf{a}_2$. The corresponding reciprocal lattice is still a rectangular lattice in which the long side of the direct lattice corresponds to the short side, and vice versa.

Let us recall that the wave vectors represent the direction of propagation of the waves themselves and are always perpendicular to the wave front. The vector $\mathbf{b}_1$ indicates the direction of propagation of the waves whose wavefronts are planes perpendicular to the direction of $\mathbf{a}_1$ in real space. For this reason we can say that the vector $\mathbf{b}_1$, although a different physical quantity, can be considered as parallel to the vector $\mathbf{a}_1$. The same is true, in the case of the orthogonal lattice, also for $\mathbf{b}_2$ and $\mathbf{b}_3$.

<u>Non orthogonal lattice</u>

To define the reciprocal lattice in the more general case of a non orthogonal lattice, it is better to start from the definition seen above, a condition that allows to select only the waves that enter into the development in series of the periodic function f(**r**). These are the waves for which $e^{i\mathbf{g}(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{g}\cdot\mathbf{r}}$, and therefore $e^{i\mathbf{g}\cdot\mathbf{R}} = 1$.

The vectors **g** constitute a discrete set and therefore they can be expressed, as it is done with the direct lattice, with a formula of this type:

$$\mathbf{g} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2 + n_3\mathbf{b}_3$$

The three vectors $\mathbf{b}_i$ form the basis of the reciprocal lattice. The three basis vectors that meet the condition defining the reciprocal lattice vectors are:

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \; ; \; \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1)} \; ; \; \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)}$$

In some texts the factor 2π does not appear in the definition of the $\mathbf{b}_i$ but is introduced later. This does not change the essence of the discussion. It is immediate to verify that the case of the orthogonal lattice falls within these formulas.

The denominator term is a number that represents in all three cases the volume of the elementary cell in the direct lattice, V_{dir} defined by the vectors $\mathbf{a}_i$.

It is possible to check, with a little vector algebra, that vectors **g** of this type meet the condition $e^{i\mathbf{g}\cdot\mathbf{R}} = 1$. To proceed with the calculations first we observe that:

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\,\delta_{ij}$$

If i = j, it is sufficient to explicitly calculate the product by taking into account how the $\mathbf{b}_i$ are defined. Consider for example the scalar product of $\mathbf{b}_1$ and $\mathbf{a}_1$:

$$\mathbf{a}_1 \cdot \mathbf{b}_1 = \frac{2\pi}{V_{dir}}\,\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = \frac{2\pi}{V_{dir}} V_{dir} = 2\pi$$

If, on the other hand, $i \neq j$, the product is null because, for example if i = 2 and j = 1, $\mathbf{a}_2 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = 0$, since the prism defined by the three vectors of the mixed product is reduced to a flat figure and therefore its volume is equal to zero.

Using this property we evaluate the expression of the scalar product of **g** and **R**, as previously done for the orthogonal lattice, and we verify that it is equal to integer*2, and therefore $e^{i\mathbf{g}\cdot\mathbf{R}} = 1$.

Also in the case of a non orthogonal lattices, the vectors **g** serve to define the waves $e^{i\mathbf{g}\cdot\mathbf{r}}$ that enter into the series development of the periodic functions f(**r**). In this case the series development is carried out on all the **g** vectors of the reciprocal lattice:

$$f(\mathbf{r}) = \sum_{\mathbf{g}} A_{\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}}$$

The $A_{\mathbf{g}}$ coefficients are calculated by means of the following formula (which generalizes the one seen in 1D and which is valid in all cases, even that of the orthogonal lattice):

$$A_{\mathbf{g}} = \frac{1}{V_{dir}} \int_{cell} f(\mathbf{r}) e^{i\mathbf{g}\cdot\mathbf{r}} dV$$

2.4 Reciprocal lattice properties in 2D and 3D

The reciprocal lattice, also in 2D and 3D and with a non orthogonal basis, has properties that are a generalization of those already seen for the 1D lattice.

1) We have already seen that the reciprocal lattice is, in turn, a Bravais lattice, as is obvious from the formula that defines it.

2) The reciprocal lattice is a Bravais lattice, therefore it admits a reciprocal lattice, which coincides with the direct one. To verify this we can proceed by calculating the basis vectors $\mathbf{c}_i$ of the reciprocal of the reciprocal lattice using the formulas that define the reciprocal basis $\mathbf{b}_i$. It turns out that the $\mathbf{c}_i$ and the $\mathbf{a}_i$ coincide.

3) The volume of the elementary cell of the reciprocal lattice V_{rec} is equal to:

$$V_{rec} = \frac{(2\pi)^3}{V_{dir}}$$

Where V_{dir} is the volume of the direct lattice unit cell. We already verified this property for the 3D orthogonal lattice. For the demonstration in the non orthogonal case, starting from the formulas for the $\mathbf{b}_i$, it is sufficient to calculate the volume of the cell they define, which is given by the mixed vector product $\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3)$.

There are also other properties that in 1D do not make sense because they establish a close relationship between the <u>crystallographic planes</u>, that in 1D do not exist, and the vectors of the reciprocal lattice.

First of all, let us define what is meant by <u>crystallographic plane</u>. If we consider, in 3D, a plane on which lie three points of the lattice, given the symmetry of the system there will be infinite others that define on the plane 2D Bravais lattice. To better understand this, let's think to draw a line that joins two of the three points. Since we are considering a 3D Bravais lattice each of these points must have the same configuration of other points around, and therefore on that line there will be infinite

equidistant points. This must apply to any other pair of points, so the entire plane will contain a 2D lattice. If, instead, we work in 2D, given a straight line which contains two points of the lattice, there will be endless others that define on the line a Bravais lattice in 1D.
Given a crystallographic plane, always because of the translational symmetry there will be infinite others parallel to it, to form a family of crystallographic planes.

The vectors of the reciprocal lattice possess these additional properties (not demonstrated

4) Each vector of the reciprocal lattice is perpendicular to a family of crystallographic planes.

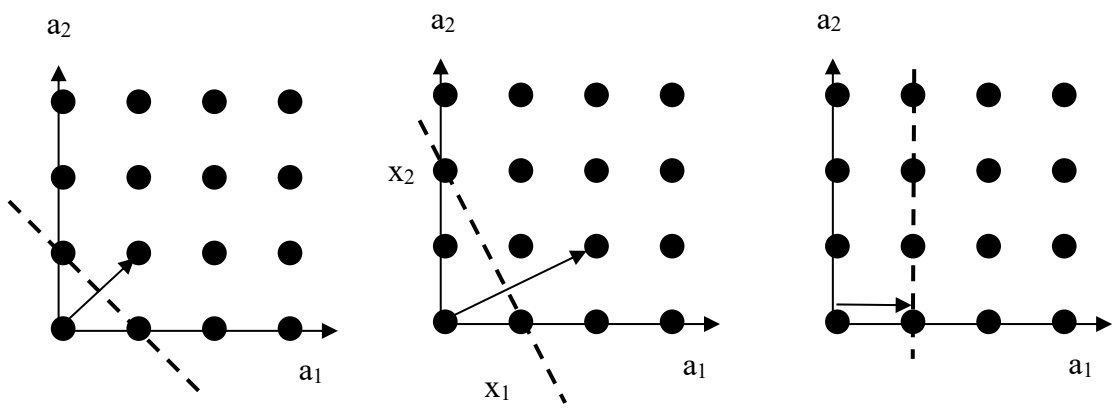
5) Given a vector **g** of the reciprocal lattice, there are infinite other parallel to it. Of all these, the one with the minimum modulus is that whose components do not have common factors. The modulus of this minimum modulus vector is inversely proportional to the distance d between the planes of the direct lattice to which it is perpendicular:

$$|\mathbf{g}| = \frac{2\pi}{d}$$

2.5 Miller indices

It follows from the properties of the reciprocal lattice that, in order to identify a family of crystallographic planes of the direct lattice, it is possible to use the minimum modulus vector of the reciprocal lattice perpendicular to them. The integer numbers used to express the three components of this vector are the Miller indices of that family of plans.

To understand the concept consider the figure that shows the case of a squared lattice in 2D, which can be seen as the projection of a cubic lattice. The lattice planes whose direction we want to specify with the Miller indices are indicated with dashed lines. The minimum modulus vector of the reciprocal lattice perpendicular to these planes is indicated with an arrow. The corresponding indices are indicated between two round brackets, as is usually done. The first two numbers indicate the components along the two directions a_1 and a_2, the third index, relative to the axis a_3, is equal to 0 in all three cases of the figure because the vectors have no components along this direction.



(1,1,0) (2,1,0) (1,0,0)

Similar considerations can be made for the lattices in three dimensions, as shown in the next figure.

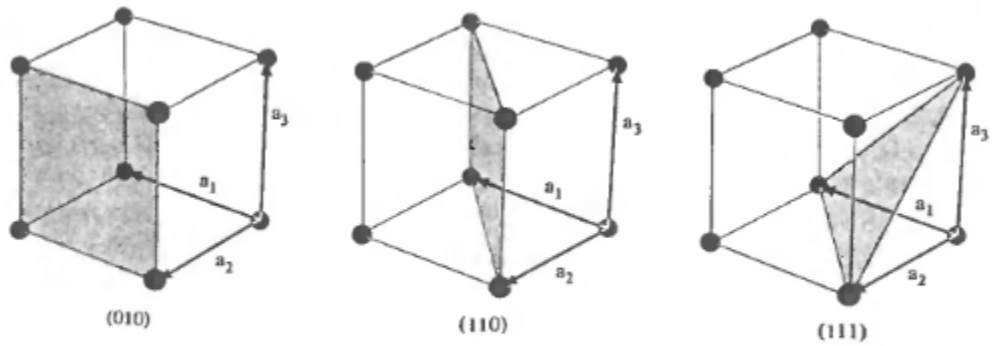


Figure 5.5
Three lattice planes and their Miller indices in a simple cubic Bravais lattice.

Miller indices can also be defined by referring, instead of the vectors of the reciprocal lattice, to the planes in the direct lattice. Let us limit ourselves to the case of a simple cubic lattice in which we consider one of these planes and call x_1, x_2 and x_3 the coordinates of the intersections of this plane with the three axes of the direct lattice (see figure). It is verified that these intersections are inversely proportional to Miller three indices:

$$x_1 = \frac{A}{h};\ x_2 = \frac{A}{k};\ x_3 = \frac{A}{l}$$

The proportionality coefficient A obviously depends on which plan of the family we are considering, or rather on its distance from the origin. It follows that it is possible to define Miller indices as a set of three integers, without common factors, proportional to the three intersections of one of these planes with the crystallographic axes:

$$h = \frac{A}{x_1};\ k = \frac{A}{x_2};\ l = \frac{A}{x_3};$$

This is actually the original definition of Miller, a 19th century mineralogist. The higher the value of Miller indices, the more "inclined" the corresponding planes are and, more importantly, the lower the density of atoms on them. In this case, we speak of "high index" plans.

2.6 General considerations on the reciprocal lattice

The concept of reciprocal lattice is undoubtedly abstract but, as we will see, it plays an important role in all the physics of crystalline systems and it is possible to attribute an immediate physical meaning to the vectors **g**. The first concrete example that we will see shortly is diffraction. First, however, let us make some very general comments on the reciprocal lattice.

We started from a periodic structure, a Bravais lattice, and we expressed the functions having the periodicity of this lattice as the sum of many waves. Each of these waves is characterized by a **g** wave vector. The Fourier analysis tells us that if the periodicity of the direct lattice is to be respected, then

the functions describing its physical properties must contain only Fourier components with particular wave vectors, the **g** vectors of the reciprocal lattice. To put it a little more physical way:

> *the reciprocal lattice is made up of all the waves (to be exact their wave vectors) physically significant for a given direct lattice.*

Physically significant means that in the physical processes involving the lattice only these waves play a relevant role, whereas all the others give rise to evanescent contributions. This is true both for the properties of the lattice as such, think for example to the function that describes the electronic density, but also for the phenomena in which the lattice interacts with electromagnetic waves or, more generally, with any material particles that, as we know, can also be considered a wave from a quantum mechanical point of view.

All these phenomena can be regarded as the interaction of two different kind of waves, those that enter into the series development of the properties of the lattice and those of the particles with which the lattice interacts. It is understandable that the interaction between two waves is maximum when the waves completely overlap, that is, if their wavelengths are equal. It follows, therefore, that the waves defined by the reciprocal lattice have a relevant role because they are the only ones for which these interactions are significant.

The first phenomenon of this type that is studied is the diffraction of waves by a crystalline structure. It is certainly not the only case. Another case of great importance concern the valence electrons of the crystal. These are objects with an undulatory nature and it is therefore evident that electronic waves characterized by a wave vector belonging to the reciprocal lattice will have a particular behavior. This will be seen very clearly when we will talk about the nearly free electron (NFE) model of electrons in a solid.

2.7 Diffraction from a Bravais lattice

The phenomenon of diffraction occurs when waves interact with a periodic structure giving rise to other waves, called diffracted waves, which are observed only along certain particular directions. In formulating a model for diffraction we will think in general terms without specifying if we are considering electromagnetic waves or other. The approximations that we will adopt are as follows:

- the incident radiation is elastically scattered by the individual atoms and the phase shift of the scattered waves is the same for all the atoms (coherent scattering);
- the waves scattered by the individual atoms are supposed to be spherical;
- absorption, reflection, refraction and interference of the waves inside the material are neglected.

To do better than this there is the so called <u>dynamic theory</u> of diffraction, very complex. The previous approximations, which are the basis of the <u>kinematic theory</u>, work well with waves that are poorly absorbed, with a refractive index close to unity and weak scattering. It is therefore good for X-rays from fairly large and poorly absorbing samples, and also for neutrons. For electrons, however, it is not adequate.

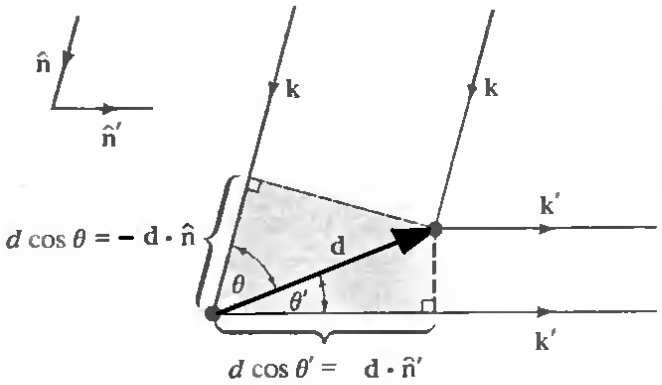
In the following we will consider the diffraction from a periodic structure proceeding in steps (a complete theory of diffraction is rather heavy and here we just want to introduce the fundamental concepts). We will begin by considering the diffraction from a Bravais lattice, that is, from a structure

with only one atom per cell. Then we will see what happens when there are several atoms per cell, whether they are all the same or different.

Formulation of von Laue

Let us first consider diffraction from a Bravais lattice (one atom per unit cell). From the figure, which shows two arbitrary points of the lattice separated by the vector **d**, we can clearly see that the interference condition can be formulated in this way, that is, by imposing that the difference in the optical path between the two rays is an integer multiple of the wavelength:

$$\mathbf{d}\cdot(\mathbf{n}'-\mathbf{n}) = m\lambda$$



The two versors **n** and **n'** are, respectively, parallel to the direction of the incident waves, identified by the wave vector **k**, and to the possible direction of the diffracted wave **k'**. Multiplying both members by $2\pi/\lambda$, and remembering that $k = 2\pi/\lambda$, we get:

$$\mathbf{d}\cdot(\mathbf{k}'-\mathbf{k}) = m2\pi$$

This condition must apply not only to **d** but to all vectors **R** of the direct lattice. Taking this into account and elevating to power, the interference condition is therefore written:

$$e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}} = 1$$

This formula is the same as the one used to define the vectors of the reciprocal lattice, provided that $\mathbf{k}' - \mathbf{k} = \mathbf{g}$. Having made this observation, we can formulate the condition of von Laue in these terms:

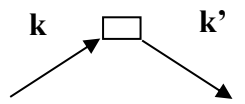
> *to have diffraction the difference between the wavevectors of the incident and of the diffracted waves must be equal to one of the vectors of the reciprocal lattice.*

We can interpret this condition by rewriting the previous formula:

$$\mathbf{k}' = \mathbf{k} + \mathbf{g}$$

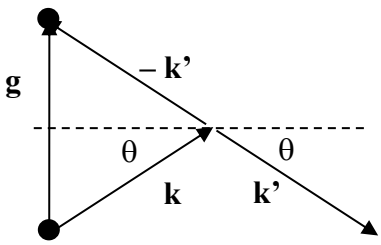
Written in this way, the condition of von Laue shows how the diffracted wave **k'** is equal to the sum of the incident wave **k** plus "something" that is given to it by the lattice. This something can only be a further wave with the periodicity of the lattice, so its wave vector can only be one of the vectors of the reciprocal lattice **g**.

Condition of von Laue



Also **k** − **k'** belongs to the RL (a Bravais lattice is a set of closed vectors with respect to sum and difference). We call this vector **g** = **k** − **k'.**

g = k + (− k') = k − k'



Vectors **k** and **k'** must be on the plane indicated by the dashed line (Bragg plane).

An important consequence of the condition of von Laue can be obtained in a purely geometric way, as shown in the figure: in order to have diffraction there must be at least one vector **g** of the reciprocal lattice such that the projection of the vector of the incident wave **k** on **g** is equal to half of the modulus of **g**.

We can also get there algebrically. We know that **g** = **k** − **k'**, so **k'** = **k** − **g**. The moduli of the two vectors **k'** and **k** − **g** must be the same, since we are considering elastic scattering, and so we write:

$$|\mathbf{k}'|^2 = (\mathbf{k}-\mathbf{g})\cdot(\mathbf{k}-\mathbf{g}) = |\mathbf{k}|^2 + |\mathbf{g}|^2 - 2\,\mathbf{k}\cdot\mathbf{g}$$

Knowing that $|\mathbf{k}'| = |\mathbf{k}|$ we have:

$$\mathbf{k}\cdot\mathbf{g} = \frac{1}{2}|\mathbf{g}|^2$$

We express **g** as the product of a unitary vector $\mathbf{n_g}$ multiplied by $|\mathbf{g}|$:

$$\mathbf{k}\cdot|\mathbf{g}|\mathbf{n_g} = \frac{1}{2}|\mathbf{g}||\mathbf{g}|$$

and therefore

$$\mathbf{k}\cdot\mathbf{n_g} = \frac{1}{2}|\mathbf{g}|$$

The projection of **k** on the direction of **g** must be $|\mathbf{g}|/2$. This is, again, the condition on vector **k** to have diffraction.

Bragg condition

Let us still consider the previous figure. The two vectors **k** and **k'** form with the plane indicated by the dashed line two angles that it is easy to verify are equal.

This plane is, as we have seen before, perpendicular to some vector **g** of the reciprocal lattice. We know that each of the **g** vectors of the reciprocal lattice is associated with a family of crystallographic planes, perpendicular to **g**. We can therefore think that the crystallographic planes act as mirrors for the incident waves. This is the ad hoc hypothesis of the Bragg formulation of diffraction from a lattice, which in the von Laue formulation becomes a consequence.

Now let us relate the quantities involved, which are the distance d between the crystallographic planes perpendicular to **g**, the wavelength and the angle of incidence. We call g the modulus of the reciprocal lattice vector and k the modulus of the incident and diffracted waves.

Let us first observe (figure) that:

$$g = 2k\sin\theta$$

In addition, it is known that g is bound to d by this relation:

$$g = n * 2\pi/d$$

If we put the two together, we get:

$$k\sin\theta = n * \pi/d$$

Remembering that $k = 2\pi/\lambda$ the Bragg condition is obtained:

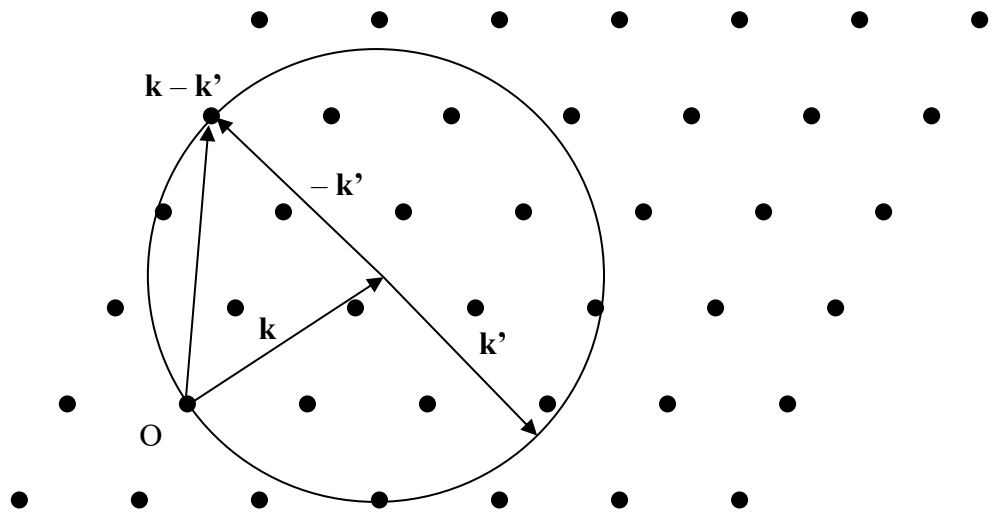
$$n\lambda = 2d \sin\theta$$

This condition can also be obtained by simply imposing that the difference in optical path between the waves reflected by two different planes is an integer multiple of λ. The result is the same but obtaining the Bragg condition from that of von Laue serves to demonstrate their equivalence.

The Ewald sphere

Another graphic way of expressing the diffraction condition has been proposed by Paul Ewald, a German physicist contemporary of von Laue, and represented in the figure. The points are those of the reciprocal lattice, whose origin is placed at point O. **k** is the vector of the incident wave and **k'** the vector of the diffused wave. The difference between the two (**k'** – **k**) and also its opposite (**k'** – **k**) must belong to the reciprocal lattice, as imposed by the condition of von Laue.

To draw the Ewald sphere we apply vector **k** to the origin O, then we trace a sphere centered on the tip of the arrow of **k** having radius equal to the modulus of **k**. If on the sphere there is another point of the reciprocal lattice then there will be diffraction and the vector **k'** of the diffracted wave will be the one that goes from the tip of the vector **k** to the point on the surface of the sphere.



In general, the sphere does not intersect points of the reciprocal lattice. For this to happen, the experimental parameters can be varied in various ways, corresponding to different diffraction techniques:

- Use polychromatic X-rays, so the **k** vector modulus will extend over a certain range. This will result in two spheres (those with the maximum and minimum values of k). The points of the reciprocal lattice between these two spheres will be those for which diffraction occurs. This is the method of von Laue

- Keeping **k** fixed and rotate the lattice (which means rotating the crystal). This is the method of the rotating crystal.

- Use a polycrystalline sample (or powder). This is equivalent to rotating the crystal in all possible ways. The method is known as the Debye-Scherrer method.

2.8 Diffraction from structures with basis

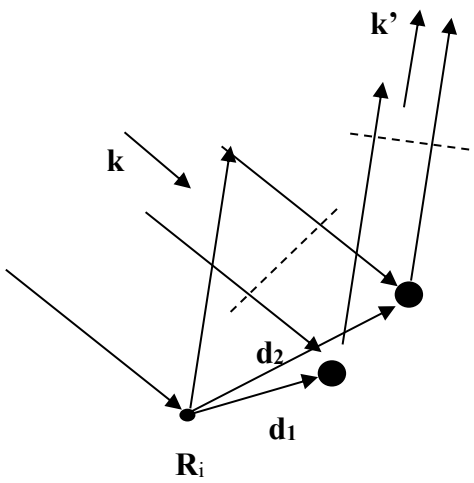
Talking about diffraction from a structure with a polyatomic basis the question becomes more complex but the substance of what happens can be summarized in a few words. We have seen that diffraction from a Bravais lattice produces figures consisting of peaks along the directions that meet the condition of von Laue. The whole discussion made so far has not taken into account the intensities of these peaks of diffraction. In reality, the relative intensities of the peaks of the diffraction figure strongly depend on the structure of the basis.

By exploiting this dependence of the intensity of the diffraction peaks on the basis, it is possible to trace the arrangement of the atoms in the cell. The use of suitable X-ray sources, such as synchrotrons, and the ability to manipulate huge amounts of data using powerful computers now allows us to use diffraction to determine the cell structure of the so called "protein crystals", i.e., crystals in which the cell consists of protein molecules, which can contain hundreds of thousands of atoms. The most important ingredients that come into play to understand the link between cell structure and diffraction figure are the geometric factor of structure and the atomic form factor.

The geometric factor of structure

The diffraction theory carried out up to this point refers to a lattice with one atom per cell. In the case of several atoms, the problem must be formulated in different terms. Consider the figure showing a basis fixed to a point of the Bravais lattice identified by an $\mathbf{R}_i$ vector (the smallest black point in the figure). The basis is formed by two equal atoms (the two largest black circles). The position in the cell of the two atoms is identified by vectors $\mathbf{d}_1$ and $\mathbf{d}_2$. A wave **k** is sento onto the lattice and a diffracted wave **k** is generated.

If there are several atoms per cell it is as if there were different and equals Bravais lattices shifted from each other, each of them having the same diffraction condition. In evaluating the intensity of the diffracted wave, it is necessary to add up the various waves diffused by the individual sublattices, taking into account the phase of each one, which will be different for each of the sublattices. We choose to express the phase of each sublattice by using one of the $\mathbf{R}_i$ as a reference point. This point, in turn, can be any point in space and not necessarily one of those where an atom is located.



Let us consider the sub-lattice corresponding to the atom that is in the $\mathbf{d}_j$ point. The phase of the wave diffracted by this sub-lattice will have a phase difference equal to $\mathbf{d}_j \cdot \mathbf{g}$ if measured with respect to point $\mathbf{R}_i$. The total diffracted wave is the sum of the waves diffused by all the sublattices j and therefore its amplitude S is determined by the coherent sum of the waves diffracted by each of them:

$$S_g = \sum_{i=1}^{n} e^{i\mathbf{g}\cdot\mathbf{d}_i}$$

This term is called structure factor. Its importance lies in the fact that, in some cases, there is completely destructive interference between the various waves and this leads to the cancellation of some peak of diffraction that would be present if the diffraction occurred from a single Bravais lattice and not from various intersecting sublattices.

The atomic form factor

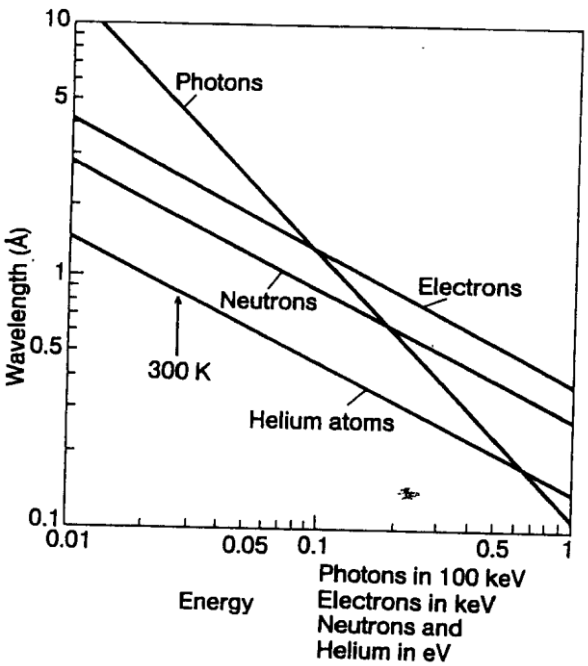
A further factor influencing the intensity of the individual peaks is the probability of wave diffusion, which obviously depends on the type of atom and also on the value of the wave vectors involved. This also determines the relative value of the diffraction peaks.

Temperature and the Debye-Waller factor

Until now, we have been thinking about a cell made up of motionless atoms while in reality the atoms move from their positions of equilibrium because of thermal agitation. This involves an attenuation of the intensity of the diffraction peaks and their enlargement due to the appearance of diffracted waves even at slightly different angles than those provided by the condition of von Laue. It is possible to calculate the reduction factor of the intensity of the diffraction peaks due to thermal agitation, the so called Debye-Waller factor.

2.9 The probes used for diffraction measurements

To have diffraction it is necessary to send waves on the structure under examination. The most used ones are X-rays, electrons and neutrons. Atoms or molecules are also used.



X-rays: electromagnetic waves of energy between a few KeV and a few tens of KeV are used. They have always been the most used probe for the ease with which they can be produced (X-ray tubes and synchrotrons) and for their ability to penetrate materials. The probability of X-ray diffusion depends on the number of electrons present, more precisely it is proportional to the square of Z.

Neutrons: thermal neutrons have wavelengths in the order of lattice parameters and can therefore be used for diffraction from solids. To produce them, it is necessary to have a nuclear reactor or a spallation source. They are penetrating and the probability of being diffracted is high even for low Z elements. In addition, they possess spin and this allows to obtain information on the magnetic ordering of samples.

Electrons: electrons of a few tens of eV have wavelengths suitable for producing diffraction from crystalline solids. The corresponding technique is called LEED (Low Energy Electron Diffraction). They do not prnetrate in matter (only a few atomic layers) and this makes them interesting for the study of surfaces.

3. Classical model of lattice oscillations

Ordinary condensed matter consists, as we have seen, of ions that do not undergo appreciable modifications with respect to the neutral atom and of the outermost electrons, which are generally in very different states from those they occupied in the isolated atom. Ultimately, a solid is a set of two types of particles, ions and electrons.

We have already said that the De Broglie wavelength of ions is usually much smaller than the lattice parameters. This makes it possible, by means of a superposition of waves, to construct objects of defined shape (ions) with a resolution smaller compared to the lattice parameter. The ions can therefore be treated as classical objects and the possibility of doing so is ensured by the correspondence principle according to which classical mechanics is the limiting case of quantum mechanics for very large quantum numbers (to which correspond small wavelengths). The same, as we will see, does not apply to the valence electrons for which the use of quantum mechanics is mandatory.

3.1 Approximations

Once we have chosen to represent a solid as a set of classic objects (ions) held together by a quantum glue (the valence electrons) we can proceed to the study of the oscillations of the ions around their equilibrium positions.

Ions are held in their positions by the chemical bond, i.e., by the electrons. When they move from this position, restoring forces are generated that bring the ions back to their original position and this causes oscillations in the entire structure.

The approximation generally adopted to treat an oscillating structure is the adiabatic approximation of Born-Oppenheimer, according to which in every instant the electron system is in the ground state corresponding to the instantaneous configuration of the ions. It follows that it is possible to speak of the force acting on the atoms of the lattice by assuming that this force only depends on the relative position of these atoms.

We will also assume that the oscillations are "small" and therefore that the restoring forces are proportional to the displacements. In this way the potential energy can be expressed only by means of quadratic terms of the atomic displacements. In these conditions we speak of harmonic approximation. This model is good in most cases. There are some important situations, however, where anharmonic terms must be taken into account (e.g. to explain thermal expansion or certain aspects of thermal conduction).

At first we might think to consider each atom as independent and treat it as an harmonic oscillator. On this assumption is based the simplest of the theories formulated to explain the thermodynamic properties such as the specific heat of a solid, the Einstein model. Apart from a few solid noble gases, however, this model does not accurately account for the observations and the reason is that the oscillations of the individual atoms are not independent of each other but are instead strongly correlated.

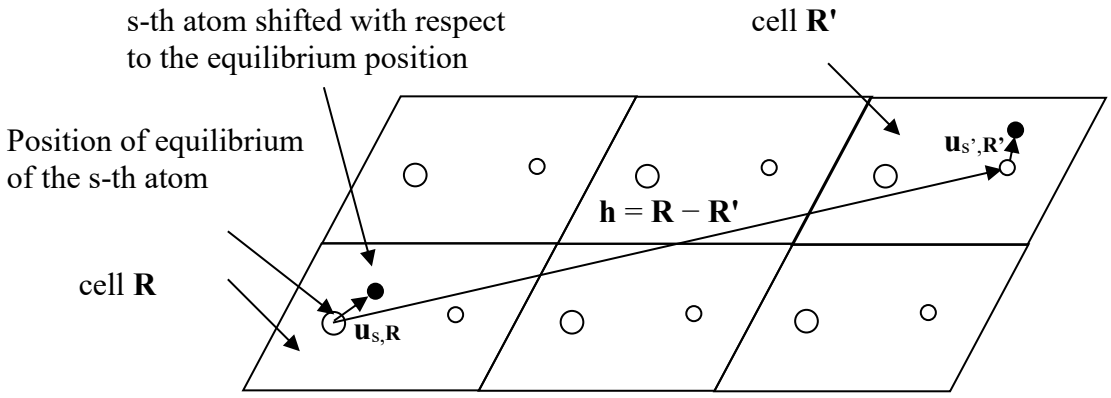
Taking these premises into account, we can briefly state that the philosophy used to study the oscillations of a lattice is the following:

- A classic model is formulated that allows the derive the correct characteristics of the oscillations, in particular the dispersion relations and the density of the states.

- Then we quantize the oscillations arriving at the concept of phonon. The quantization of the oscillations is indispensable to understand the thermodynamic properties of solids, in particular the specific heat, and the diffusion phenomena in which the lattice interacts with other particles.

3.2 General scheme of the lattice oscillations theory

Let us first consider the general case of a structure in which the basis is made of an arbitrary number of different atoms. The discourse is quite abstract and the formalism heavy, but it is worth starting in this way and then moving on to consider in detail some particularly simple cases that exemplify well the general formulation.



First we define $\mathbf{u}_{s,\mathbf{R}}(t)$, the displacement from the equilibrium position of the s-th atom of the cell identified by the vector **R** of the Bravais lattice of the structure. To understand what is meant, look at the figure showing a lattice with two atoms per cell whose positions of equilibrium are indicated with the empty circles.

The study will proceed according to this order:

1) Formulate the expressions of the kinetic energy and of the potential energy.

2) Write the equations of motion (using the lagrangian formalism).

3) Translational symmetry is exploited to simplify the equations and to make hypotheses on the form of the solutions.

4) The motion equations are solved by obtaining the so called <u>dispersion relations</u>.

Let us develop this scheme in detail.

3.3 The equations of motion

1) Expression of the kinetic and the potential energy

The kinetic energy is simply written using the classical expression:

$$E_k = \sum_{s,\mathbf{R}} \frac{1}{2} M_s \left|\dot{u}_{s,\mathbf{R}}\right|^2$$

M_s is the mass of the s-th atom. The sum includes all the atoms and runs on the index that identifies the cells (**R**) and also on the index that identifies the atoms inside them (s).

Let V be the function that expresses the potential energy of the system which, as we know, is a function of the displacements of all the atoms with respect to their equilibrium position (by assuming that it is null in the equilibrium configuration). Developing V in series of the atomic displacements we get:

$$V = \sum_{s,\mathbf{R},j} \left[\frac{\partial V}{\partial u^j_{s,\mathbf{R}}}\right]_o u^j_{s,\mathbf{R}} + \frac{1}{2} \sum_{ss',\mathbf{RR'},jj'} \left[\frac{\partial^2 V}{\partial u^j_{s,\mathbf{R}} \partial u^{j'}_{s',\mathbf{R'}}}\right]_o u^j_{s,\mathbf{R}} u^{j'}_{s',\mathbf{R'}} + \ldots$$

The index j refers to the different Cartesian components x, y and z. The first term is null because we are in equilibrium and therefore all the derivatives with respect to the displacements are equal to zero. The term that interests us, in the harmonic approximation, is the quadratic one, and the anharmonic ones of higher order are therefore neglected.

2) The lagrangian and the equations of motion

In order to formulate the equations of motion we first write the lagrangian:

$$L = E_k - V$$

By imposing that it satisfies the variational principle of minimum action (to be precise, the stationarity of the action defined as the integral of the lagrangian in time and by assuming appropriate boundary conditions), we arrive at the equations of Euler-Lagrange:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{u}^j_{s,\mathbf{R}}} - \frac{\partial L}{\partial u^j_{s,\mathbf{R}}} = 0$$

By replacing the expressions previously obtained for the kinetic and potential energy, finally we obtain the equations of motion:

$$M_s \ddot{u}^j_{s,\mathbf{R}} = -\sum_{s',\mathbf{R'},j'} \left[\frac{\partial^2 V}{\partial u^j_{s,\mathbf{R}} \partial u^{j'}_{s',\mathbf{R'}}}\right]_o u^{j'}_{s',\mathbf{R'}}$$

It is a set of many coupled second-order differential equations. The unknown are the functions that describe how the 3 components of the $u^j_{s,\mathbf{R}}(t)$ displacements vary in time. If there are n atoms per

cell and we consider N cells we have 3nN equations. We can streamline the notation by defining the tensor:

$$\left[\frac{\partial^2 V}{\partial u^j_{s,\mathbf{R}}\partial u^{j'}_{s',\mathbf{R'}}}\right]_o \equiv G^{jj'}_{s\mathbf{R},s'\mathbf{R'}}$$

In this way the equations are written in a more compact form:

$$M_s \ddot{u}^j_{s,\mathbf{R}} = -\sum_{s',\mathbf{R'},j'} G_{s\mathbf{R},s'\mathbf{R'}} u^{j'}_{s',\mathbf{R'}}$$

In order not to weigh down the notation, let us remove momentarily the index j and continue to treat the displacements **u** as vectors being understood that in reality each of these equations corresponds, in three dimensions, to three different equations, one for each of the components of $\mathbf{u}_{s,\mathbf{R}}$:

$$M_s \ddot{\mathbf{u}}_{s,\mathbf{R}} = -\sum_{s',\mathbf{R'}} G_{s\mathbf{R},s'\mathbf{R'}} \mathbf{u}_{s',\mathbf{R'}}$$

The physical meaning of this equation is very simple. The left member is the product of the mass of the s-th atom in the **R**-th cell by its acceleration, i.e., the force acting on this atom. The right member, on the other hand, is the sum of several terms each of which expresses the force due to the $\mathbf{u}_{s'\mathbf{R'}}$ displacement of the s'-th atom of the **R'**-th cell. It is a generalization of the equation of the elementary harmonic oscillator:

$$m\ddot{x} = -kx$$

Where x is the displacement, m is the mass and k is the elastic constant of the spring. Due to this analogy the term $G_{s\mathbf{R},s'\mathbf{R'}}$ is called <u>force tensor</u>.

The system of 3nN equations above is obviously unassailable for an amorphous solid, but in crystal structures the presence of translational symmetry allows to simplify it enormously.

3) Translational symmetry is used to simplify the equations

Let us first consider the form of the force tensor G. Since the structure is periodic the value of G will not depend on the absolute values of **R** and **R'**, but only on their difference **h** = **R'** − **R**. Remembering the third of the definitions of Bravais lattice given in the initial chapter we can say that the vectors **h** define in turn a Bravais lattice. By taking this fact into account the equations of motion assume the following form:

$$M_s \ddot{\mathbf{u}}_{s,\mathbf{R}} = -\sum_{s',\mathbf{h}} G_{s,s'}(\mathbf{h})\, \mathbf{u}_{s',\mathbf{R}+\mathbf{h}} \qquad \text{(EM)}$$

In this formulation, the cell **R** serves as the origin for the summation, which is in fact indexed with **h** because it is made on vectors **R** + **h**.

As a consequence of the fact that the force between two atoms depends only on **h** the different equations of motion for all s-th atoms in the different cells all contain the same coefficients. This

means that the equation of motion is the same for all s-th atoms regardless of the cell to which they belong.

If the equations for all equivalent atoms are the same, so are the solutions, which physically means that all equivalent atoms move according to the same law. As we will see, what makes the laws of motion different are the initial conditions of each one.

To define these initial conditions we will formulate some hypotheses on the solutions of the problem, always exploiting translational symmetry, and this will lead us to a drastic decrease in the number of equations that need to be solved.

3.4 The properties of the solutions

We are studying the excitations of a periodic structure and the equations of motion are a generalization of the equation of the harmonic motion: it is therefore reasonable to think that the solutions we are looking for are harmonic oscillations of the atoms. In addition, these excitations must reflect the translational symmetry of the structure. To understand what it means, let us consider first a single atom that is oscillating.

- Let us start with the assumption that the oscillations take place along a particular direction and with a fixed amplitude, identified by a vector $\boldsymbol{\varepsilon}_s$. For translational symmetry, both the direction and the amplitude must be the same for all s-th atoms that are in the lattice.

- Suppose that each atom oscillates sinusoidally with pulsation ω. The same, for symmetry, must do all other equivalent atoms in the other cells.

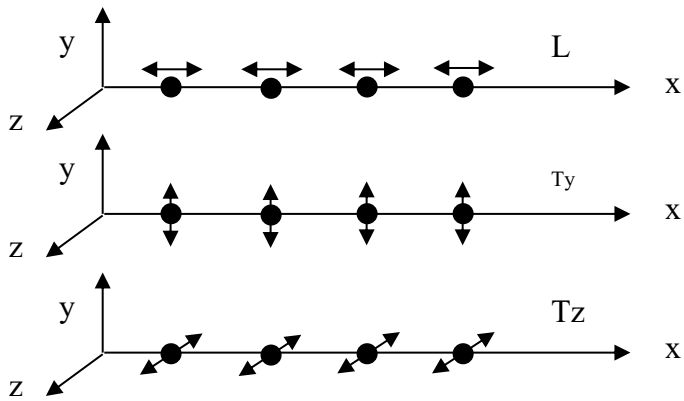
- If the equations of motion are the same and if the frequency of oscillation is the same, considering two homologous atoms in two different cells the difference in their motion will be constituted at most by a phase factor. This phase difference must obviously be, for the translational symmetry, the same among all the equivalent atoms placed in adjacent cells of the lattice. Therefore, the phase factor must be able to be expressed as $e^{i\mathbf{q}\cdot\mathbf{R}}$.

Ultimately, the $\mathbf{u}_{s,\mathbf{R}}$ must have an expression of this type (of which only the real part is considered):

$$\mathbf{u}_{s,\mathbf{R}}(t) = \boldsymbol{\varepsilon}_s e^{-i\omega t}\, e^{i\mathbf{q}\cdot\mathbf{R}} \qquad \text{(SOL)}$$

These are obviously plane waves in which the perturbation is directed as $\boldsymbol{\varepsilon}_S$, which therefore represents the polarization of the wave, and whose wave vector is $\mathbf{q}$, which indicates the direction of propagation and the wavelength. The possible directions of propagation of the perturbation depend on the symmetry of the lattice. In the simple case of a cubic lattice, and of a direction of propagation that coincides with one of the axes of symmetry, three different types of wave are possible, as the figure shows. The three corresponding oscillation modes are the longitudinal mode (L) and the two transverse modes (T_y and Tz). In general, however, waves are characterized by an arbitrary propagation direction with respect to the $\boldsymbol{\varepsilon}_S$.

By replacing the hypothesized solution (SOL) in the motion equations (EM) results in:



$$-M_S\omega^2\left(\boldsymbol{\varepsilon}_s e^{i\mathbf{q}\cdot\mathbf{R}}e^{-i\omega t}\right) = -\sum_{s',\mathbf{h}} G_{s,s'}(\mathbf{h})\ \left(\boldsymbol{\varepsilon}_{s'} e^{i\mathbf{q}\cdot\mathbf{h}}e^{i\mathbf{q}\cdot\mathbf{R}}e^{-i\omega t}\right)$$

$$\uparrow \qquad\qquad\qquad \uparrow$$

$$\mathbf{u}_{s,\mathbf{R}}(t) \qquad\qquad\qquad \mathbf{u}_{s',\mathbf{R}+\mathbf{h}}(t)$$

The common factor $-e^{i\mathbf{q}\cdot\mathbf{R}}e^{-i\omega t}$ can be eliminated to get:

$$M_S\omega^2\boldsymbol{\varepsilon}_s = \sum_{s'}\left(\sum_{\mathbf{h}} G_{s,s'}(\mathbf{h})\, e^{i\mathbf{q}\cdot\mathbf{h}}\right)\boldsymbol{\varepsilon}_{s'}$$

This is a very important step because it is the one that allows us to greatly simplify the problem:

- first of all, the unknowns are no longer the $\mathbf{u}(t)$ functions but are the algebraic variables $\boldsymbol{\varepsilon}_S$;
- secondly, it is no longer necessary to write an equation for each of the atoms of the structure. Instead, it is sufficient to write the equations only for the atoms of one cell because since we know that the solutions for the atoms in a cell displaced by a vector **R** are the same unless the phase factor $e^{i\mathbf{q}\cdot\mathbf{R}}$. This is a huge simplification as it reduces the number of equations to be solved from 3nN to 3n.

The solutions of the system of 3n equations depend on the particular value of the vector **q**. All the possible modes of oscillation of the structure are obtained by solving the systems of equations corresponding to all the possible values of **q**.

The coefficients of the system of equations are calculated, in general, by making sums that extend to the entire lattice and each of them is in turn a function of **q**:

$$G_{s,s'}(\mathbf{q}) = \sum_{\mathbf{h}} G_{s,s'}(\mathbf{h})\, e^{i\mathbf{q}\cdot\mathbf{h}}$$

As we can see, we can say that it is the Fourier transform of the tensor $G_{s,s'}(\mathbf{h})$. At the end, this is how we write the equations of motion:

$$M_s\omega^2\boldsymbol{\varepsilon}_s \; - \; \sum_{s'} G_{s,s'}(\mathbf{q})\,\boldsymbol{\varepsilon}_{s'} = 0$$

3.5 Solution of the problem

Let us move the terms and rewrite the equations of motion in this form:

$$\frac{1}{M_s}\sum_{s'} G_{s,s'}(\mathbf{q})\,\boldsymbol{\varepsilon}_{s'} = \omega^2\boldsymbol{\varepsilon}_s$$

This makes evident how the system of 3n equations is an eigenvalue equation that can be represented in matrix form: the matrix of the coefficients multiplies a column vector constituted by the components of the displacements $\boldsymbol{\varepsilon}_s$ and in doing so produces the same column vector multiplied by the constant term ω^2.

The condition for having non trivial solutions (all the unknowns equal to zero), is to satisfy <u>the secular equation</u>, in which the unknowns are the pulsations ω^2 (**q**, let us remember, has a fixed value). This equation has order 3n and therefore allows 3n solutions.

In this way we find 3n positive values of ω of such that, assigned the value of **q**, there are non trivial solutions. If we choose a different value of **q,** the values of ω, which are therefore to be considered functions of **q,** also change. The relations between ω and **q** are called <u>dispersion relations</u>:

$$\omega = \omega(\mathbf{q})$$

For a crystal with n atoms per cell there are 3n relationships of this type, not just one. It is possible to demonstrate that these relations are in general "regular" functions, i.e. they are single valued functions continuous in the **q** variable.

The solutions thus obtained are periodic in the reciprocal space:

$$\omega(\mathbf{q} + \mathbf{g}) = \omega(\mathbf{q})$$

To demonstrate this statement we consider the tensor that represents the coefficients of the equations of motion:

$$G_{s,s'}(\mathbf{q}) \;=\; \sum_{\mathbf{h}} G_{s,s'}(\mathbf{h})\, e^{i\mathbf{q}\cdot\mathbf{h}}$$

If we replace **q** with **q** + **g**, the term $G_{s,s'}(\mathbf{h})$ remains unchanged since it does not depend on **q** while the exponential term becomes $e^{i\mathbf{q}\cdot\mathbf{h}}e^{i\mathbf{g}\cdot\mathbf{h}}$. Let us think about the term $e^{i\mathbf{g}\cdot\mathbf{h}}$ by remembering that the vectors **h** are those of the direct lattice (they were in fact defined in this way: **h** = **R'** − **R**, where **R** and **R'** are vectors of the Bravais lattice). We have therefore that $e^{i\mathbf{g}\cdot\mathbf{h}} = 1$ for each **h**.

In other words, the coefficients of the system of equations are periodic in the reciprocal lattice and therefore their solutions, and the dispersion relations that derive from them, are necessarily so too.

As we have seen the 3nN equations that it is possible to write for all the atoms of the lattice are reduced to 3n equations by virtue of the translational symmetry. Until now, we have considered the **u** displacements of the individual atoms, which have then been reduced to the $\boldsymbol{\varepsilon}_s$ algebraic variables only. This has led us to write a system of equations in each one of them, in general, are presente all the $\boldsymbol{\varepsilon}_s$. Due to this reason it is a system of coupled equations. It would be better to be able to use as unknown variables certain linear combinations of the polarization vectors $\boldsymbol{\varepsilon}_s$ that make all the equations independent one from the others or, to use a different word, decoupled. In this way each equation would represent an independent harmonic oscillatore and we would have found the normal modes of oscillation of the structure. In this way then we would be able to break down any vibrational motion into the sum of many separate oscillators (each one described by one of the independent equations).

The difficult thing is to find a priori these coordinates corresponding to the normal modes of vibration. This can be done by considering the symmetries and using the group theory.

Having made these general considerations, it is advisable to examine closely two cases that are particularly simple but which contain all the essential elements for understanding the lattice dynamic from a classical point of view.

3.6 Chain of equal atoms

Physical model

Let us consider a linear chain of N atoms separated by a distance a and connected by means of equal springs of elastic constant K. In such a system, each atom only interacts with its first neighbours. In 1D, which is the case we will consider, the displacements can only occur along the line of the chain, in 3D the same system can also give rise to transverse waves.

Motion equations

It is possible to proceed by writing the lagrangian of the system and obtaining the equations of Euler-Lagrange. In this simple case however it is more immediate to write the Newton law for the n-th atom which is located at the point wit coordinate na. The force acting on the atom is the elastic force of the two springs to which the atom is attached, a force equal to K (the elastic constant, which plays the role of tensor of forces G in a very simplified form) times the variation of the spring length. Taking into account the signs of the displacements and that we are considering the force on the n-th atom, we write:

$$M\ddot{u}_n = K(u_{n+1} - u_n) + K(u_{n-1} - u_n) =$$
$$K(-2u_n + u_{n+1} + u_{n-1}) = -K(2u_n - u_{n+1} - u_{n-1})$$

Solutions

As we can see, the equation (which is obviously the same for all atoms) contains three variables, u_n, u_{n+1} and u_{n-1}. At this point, a solution of this kind is hypotesized:

$$u_n = ue^{i(qna - \omega t)}$$

Where na represents the position, along the chain, of the n-th atom and u the amplitude of the oscillations. Inserting it in the equation of motion and remembering the formula for the sine of the half angle ($1 - \cos\alpha = 2\sin^2(\alpha/2)$) we obtain:

$$M\frac{d^2}{dt^2}\left(ue^{i(qna - \omega t)}\right) = -Ku\left[2 - e^{-iqa} - e^{iqa}\right]e^{i(qna - \omega t)} =$$

$$-2Ku(1 - \cos qa)e^{i(qna - \omega t)} = -4Ku\sin^2\left(\frac{qa}{2}\right)e^{i(qna - \omega t)}$$

Calculating the derivative with respect to time:

$$-\omega^2 ue^{i(qna - \omega t)} + \frac{4K}{M}\sin^2\left(\frac{qa}{2}\right)ue^{i(qna - \omega t)} = 0$$

By eliminating the common exponential factors:

$$-\omega^2 + \frac{4K}{M}\sin^2\left(\frac{qa}{2}\right) = 0$$

we immediately obtain the <u>dispersion relation</u>:

$$\omega(q) = 2\sqrt{\frac{K}{M}}\left|\sin\frac{1}{2}qa\right|$$

We put the modulus because the unknowns are ω^2, always positive, and moving to ω we are only interested in the positive roots not having physical meaning the negative ones.

<u>The characteristics of the dispersion relation</u>

The dispersion relation found is evidently periodic in the reciprocal space, with period $2\pi/a$. This is correct because this is the periodicity that must have the reciprocal lattice of a direct lattice where the lattice parameter is equal to a. The portion between $-\pi/a < q < \pi/a$ is the first Brillouin zone.

For small wave vector values with respect to the zone edge ($q << \pi/a$, and therefore $\lambda = 2\pi/q >> a$) we approximate the sine with its argument, $\sin\frac{1}{2}qa = \frac{1}{2}qa$, from which:

$$\omega(q) = a\sqrt{\frac{K}{M}}|q|$$

Let us rewrite it in this form:

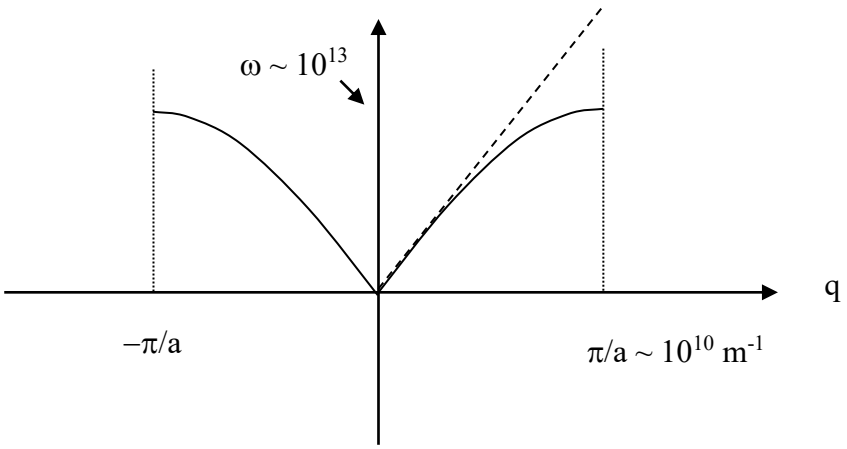
$$\frac{\omega(q)}{|q|} = a\sqrt{\frac{K}{M}}$$

Remembering that $f = \omega/2\pi$ and that $\lambda = 2\pi/q$:

$$\frac{\omega}{q} = f\lambda = a\sqrt{\frac{K}{m}}$$

The product of the wave frequency times its wavelength represents the propagation speed. For the waves we are considering, characterized by a small value of q, and therefore by a wavelength much greater than the lattice parameter a, the propagation speed is the same regardless of the value of q. These modes of oscillation are called acoustic because they correspond to sound waves, oscillations of great wavelength that all propagate with the same speed.

$$\frac{d\omega}{dt} = \text{speed} \approx 10^3 \text{ m/s}$$



The typical value of the lattice parameter is about $3*10^{-10}$ m and this allows to fix the order of magnitude of the wave vector q at the zone edge, 10^{10} m^{-1}. To define the order of magnitude of the oscillation frequencies we can assume that it is $\omega = vq$ in the whole Brillouin zone. In solids, the speed of sound v is in the range of 10^3 m/s. In this way we obtain, for q on the edge of the zone, $\omega = 10^{13}$, and therefore $f = 10^{12}$ Hz, a frequency that in the case of electromagnetic waves corresponds to the infrared region of the spectrum.

A packet constructed by means of waves of this type, all travelling at the same speed, propagates without changing shape because the phase speed and the group speed ($d\omega/dq$ calculated for the central value of q of the packet) coincide.

If the frequency is not simply proportional to q, as is the case for values of q close to the zone edge, then each component of the packet propagates with different phase speeds and this produces two effects:

- The centre of gravity of the packet travels with a group speed $v = d\omega/dq$ calculated for the central q value of the packet which is different from that of the other component waves.

- The packet spreads during its propagation and we have <u>dispersion</u>.

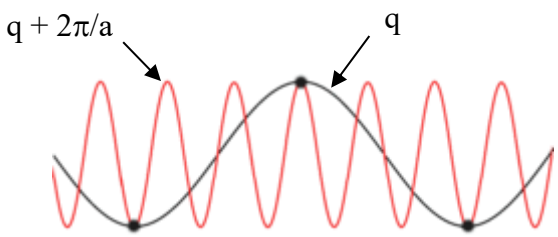
Low q waves correspond to waves whose length is much longer than the lattice parameter and in which the phase shift (equal to qa) between one atom and the other is very small. Vice versa, at the zone edge ($q = \pi/a$) the wavelength is comparable to the lattice parameter ($\lambda = 2\pi/q = 2a$) and there is a phase shift $qa = \pi$. This means that adjacent atoms vibrate in phase opposition and the whole lattice can be seen as the set of two intersecting lattices oscillating 180° out of phase one with respect to the other. Moreover, at the edge of the zone the derivative of the dispersion relation vanishes and this indicates that the speed of the corresponding waves is zero, as it is in the stationary modes.

An interesting aspect of the dispersion relationship just examined is the physical meaning of the wave vectors that lie outside the first Brillouin zone. Let us consider a vector q within the zone, to which corresponds a given value of the frequency ω. If we consider the vector $q + g = q + 2\pi/a$, that is translated by a vector of the reciprocal lattice, we obtain the same value of the frequency since the dispersion relation is periodic. The corresponding solution is therefore:

$$u_n = ue^{i[(q + \frac{2\pi}{a})na - \omega t]} = ue^{i[qna - \omega t]}e^{i2\pi n} = ue^{i[qna - \omega t]}$$

since $e^{i2\pi n} = 1$. This means that the oscillations of a given atom corresponding to the two wave vectors q and q + g are physically identical, i.e. with the same amplitude and in phase with each other. This seems strange because the wave described by the vector $q + 2\pi/a$ has a shorter wavelength than that described by the wave vector q.

In reality we are considering the oscillations of the atoms, for each of which the position can be made to coincide with the respective center of mass. It is therefore a discrete distribution and therefore different from zero only in correspondence of the positions occupied by the atoms (or their centre of mass, to be pedantic). Of the formula that represents the oscillations of the lattice, therefore, only the points corresponding to the positions occupied by the atoms are affected, i.e. x = na. What happens between one point and the next is irrelevant.



Of course, things change when we are dealing with a physical system defined at all points in real space. In this case, the shorter wavelength of the spatial component of the wave assumes a real physical meaning. An example of a quantity defined on a continuous and not on a discretized basis is the electronic wave function Ψ(x), defined for all position values and not just for those where x = na. In this case, we will see that the question changes and becomes more complicated. Another example of a continuous distribution is represented by the electromagnetic waves.

3.7 Chain with two atoms per cell

Imagine a system consisting of two atoms of different mass (m_1 and m_2) connected by equal springs of elastic constant k. The distance between atoms when they are stationary and in equilibrium is equal to a (and therefore the period is 2a).

We use as coordinates the positions of the atoms measured with respect to their positions of equilibrium. In cell n-th the two solutions will be called u_n^1 and u_n^2, the indices 1 and 2 serve to distinguish the two different atoms of the cell. At this point we write the equations for the two atoms and the hypothesized solutions are inserted:

$$u_n^i = u^i e^{i(qn*2a - \omega t)}$$

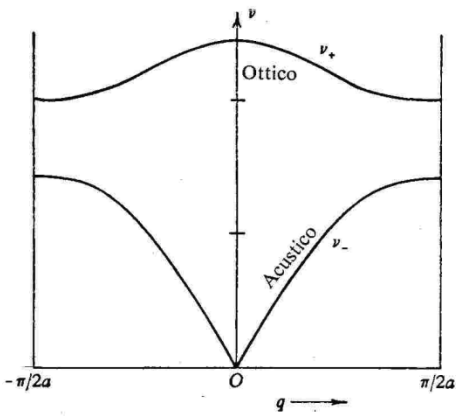
Thus, proceeding as already done for the single equation of the linear chain, we arrive at a system of two algebraic equations, where the variables are the amplitudes of the oscillations u^i :

$$\begin{vmatrix} 2k - m_1\omega^2 & -2k\cos qa \\ -2k\cos qa & 2k - m_2\omega^2 \end{vmatrix} \begin{vmatrix} u^1 \\ u^2 \end{vmatrix} = 0$$

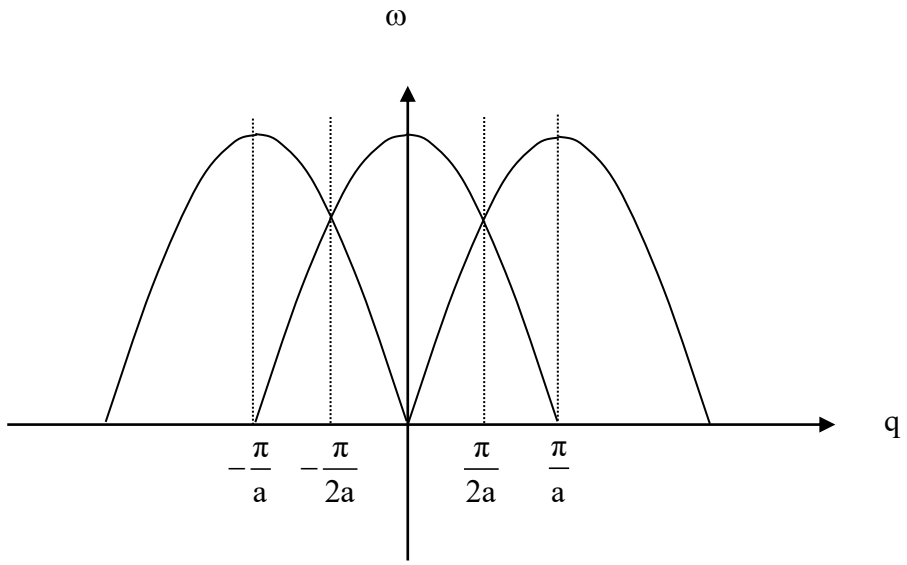
This system allows for non trivial solutions if the determinant of the system coefficients is null. Solving the secular equation that translates this condition gives two solutions for the frequency:

$$\omega_\pm^2 = k\left(\frac{1}{m_1} + \frac{1}{m_2}\right) \pm k\sqrt{\left(\frac{1}{m_1} + \frac{1}{m_2}\right)^2 - \frac{4\sin^2 qa}{m_1 m_2}}$$

In these two solutions the wave vector q appears as a parameter and so we can say that $\omega = \omega(q)$. Therefore we have two dispersion relationships, the trend of which is shown in the figure.



The previous formula tells us that the period of the dispersion relations is the one of function $\sin^2 qa$, i.e., π/a, and for this reason the two dispersion relation are plotted in a Brillouin zone extending between $-\pi/2a$ and $\pi/2a$. This is consistent with the fact that the lattice we are considering has, in the real space, a period equal to 2a and therefore the period of its reciprocal lattice must be $2\pi/2a$.



The solution ω_- tends to zero for small values of q and the corresponding modes are the acoustic ones seen for the chain of equal atoms. The ω_+ curve instead assumes a value different from zero when q = 0 and describes a mode of oscillation that does not have an equivalent in the chain of all equal atoms.

Once the values of the frequencies are known, it is possible to insert them in the motion equations, solve the system and obtain the relative solutions. If we limit ourselves to small values of q, in correspondence of the two different branches of the dispersion relations we obtain two different types of solution:

For the ω_- relation, which describes <u>acoustic modes</u> of oscillation, $u^1 = u^2$, which means that the movements of the different atoms of the cell are of the same sign and therefore their oscillations are in phase (furthermore let us remember that for small values of q the phase shift between two equivalent atoms in adjacent cells is very small).

For ω_+ we have $u^1 = -\frac{m_2}{m_1}u^2$, therefore the two atoms in the cell always oscillate in phase opposition. If they are ionized atoms, this corresponds to having oscillating electric dipoles and therefore to the possibility of absorbing or emitting electromagnetic radiation. In other words, it is an optically active mode of oscillation, which is why it is called <u>optical mode</u>.

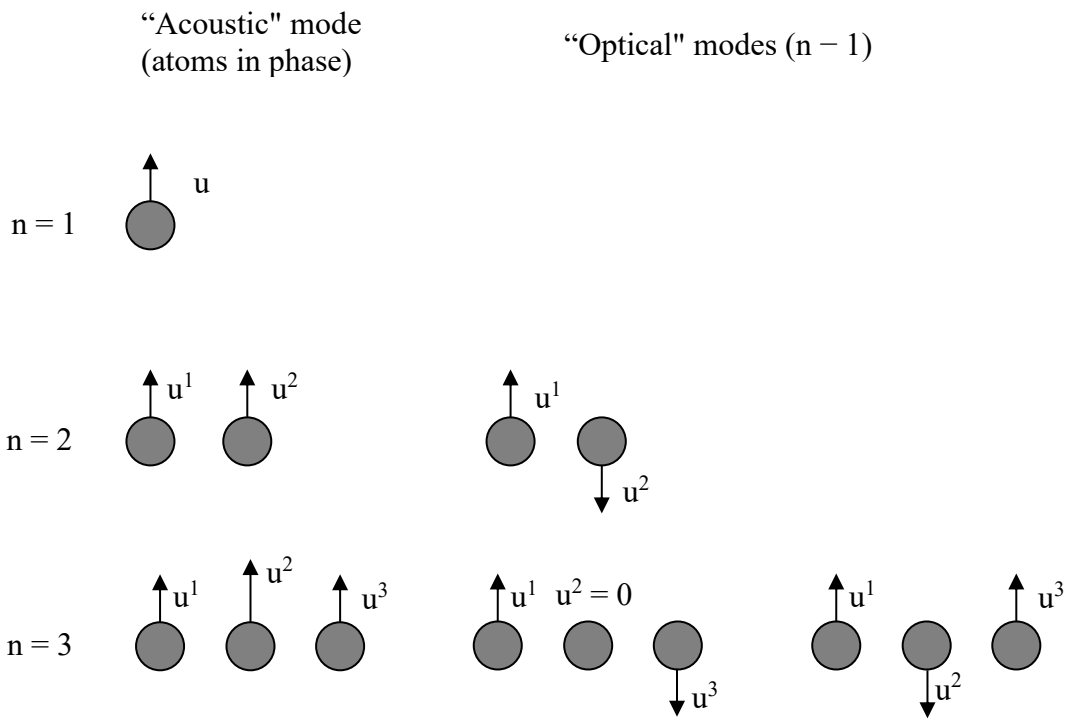
Let us imagine now that everything said up to now is valid but that the masses of the two atoms are equal ($m_1 = m_2$). In this case, the two dispersion relations assume the same value at the edge of the zone as is easily verified by noting that the term under root is cancelled for $q = \pi/2a$.

Let us consider the two dispersion relations only within the Brillouin zone which, as mentioned, extends between $-\pi/2a$ e $\pi/2a$. Since the dispersion relations are periodic in the reciprocal space, this figure must be repeated periodically. At this point we observe the figure that is determined between the points $-\pi/a$ and π/a: we find the dispersion relation of the chain of equal atoms! Obviously this must be the case, because we are now considering atoms that are all the same, separated by a distance a which is half the distance considered before, so in the reciprocal space the lattice parameter doubles.

In the chain of equal atoms previously studied, the parabola centered on the vertical axis is not present because there is only one dispersion relationship with period $2\pi/a$. In the case in which there is a slight difference in mass between the two species of atoms, the shape of the dispersion relations remains roughly the same. However, the figure that represents the dispersion relations within the first Brillouin zone of the chain of equal atoms must be repeated but with a periodicity equal to π/a and this determines the appearance of more than one dispersion curve.

In the case of the two equal masses, the two dispersion relations meet at the edge of the zone. The fact that the two masses are different therefore is responsible for opening a gap between the two dispersion relations at the edge of the area, which separates the curves of the optical branch and the acoustic branch.

Another observation that should be made is that the presence of more than one different object (two atoms) in the elementary cell of the direct lattice produces more curves (the two dispersion relations) in the Brillouin zone, as was already foreseen in the general treatment of the lattice oscillations. If the cell were made up of n different atoms separated by a distance a we would have n distinct curves and the thing can be understood still considering the extreme case of equal atoms but with an elementary cell of na length. Also in this case we have to find the dispersion relation of the chain of equal atoms, whose period is $2\pi/a$, which however must repeat itself with a period $2\pi/na$, hence the presence of n curves in the Brillouin zone.



3.8 Chains with more than two atoms per cell

Periodicity, as seen, allows to study the problem of the oscillations of an infinite structure by reducing to the oscillations within a single cell. We have seen it in the general case and then in the mono and diatomic chains. Let us consider again the characteristics of the solutions found.

In the case of the chain of equal atoms we have assumed solutions of the type $u_n = ue^{i(qna - \omega t)}$. By replacing this solution in the motion equation, the dispersion relations have been obtained. In turn these solutions are simply characterized by the value of the amplitude u_o which is the same for all the atoms and therefore in the motion equation it does not appear being able to assume arbitrary values. Turning to two atoms per cell we have seen that the algebraic unknowns u^1 and u^2 admit two different types of solution:

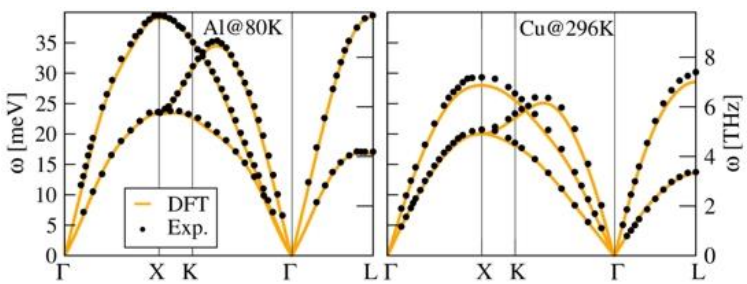
$$u^1 = u^2$$

$$u^1 = -\frac{m_2}{m_1} u^2$$

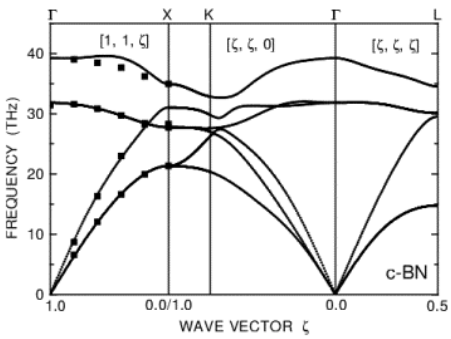
Moving on to three atoms we know that there will be three solutions and we can imagine that the solutions are of the type shown in the figure. Moving to an arbitrary number of atoms, and generalizing this discourse, we will have n modes of oscillation, one of acoustic type and (n − 1) of the optical type. In three dimensions we will have 3n modes of oscillation, 3 of which acoustic and 3(n − 1) optical.

3.9 Experimental dispersion relations

The simplest case is that of a cubic lattices with one atom per cell, such as Al or Cu shown in the figure. In this case the dispersion relations are like those of the 1D lattice with the only difference that there are three acoustic oscillation modes for each symmetry direction, one longitudinal (LA) and two transverse (TA).



In general, longitudinal modes have a greater slope than transverse ones (i.e. the propagation speed is higher). The graphs of the dispersion relations (theoretical or experimental) are represented moving



in the reciprocal space along different directions. Greek letters indicate points of high symmetry of reciprocal space. The point is conventionally the center of the first Brillouin zone.

In anisotropic lattices (not cubic) but with only one atom per cell there are always 3 modes of oscillation for each direction **q** but the polarizations can have arbitrary directions.

The 3D case with more atoms per cell (equal or different atoms, the same thing does not change since even if the atoms are equal the fact of being in equivalent sites makes them different) is the most complicated case. If there are n atoms per cell then there are 3n modes of oscillation, three of acoustic type, all the others of optical type. The figure shows the dispersion relationships of boron nitride (BN) where we can see both acoustic and optical modes.

4. Specific heats of solids

4.1 Phenomenological aspects

Let us consider the molar specific heat of some elementary solids, that is, consisting only of atoms of the same kind. Molar means that the amount of material referred to contains the same number of atoms, the Avogadro number $N_A = 6.022 \times 10^{23}$. The specific heat we will consider is the constant volume heat because in solids the volume variations with temperature are negligible.

Table 6-4

Dulong and Petit's Data on the Molar Heat Capacities of Solid Elements[a,b]

Element	Gram heat capacity ($J\ deg^{-1}\ g^{-1}$)	Atomic weight	Molar heat capacity ($J\ deg^{-1}\ mole^{-1}$)
Bi	0.120	212.8	25.64
Au	0.125	198.9	24.79
Pt	0.133	188.6	25.04
Sn	0.215	117.6	25.30
Zn	0.388	64.5	25.01
Ga	0.382	64.5	24.60
Cu	0.397	63.31	25.14
Ni	0.433	59.0	25.56
Fe	0.460	54.27	24.98
Ca	0.627	39.36	24.67
S	0.787	32.19	25.30

[a]Reproduced by permission from J. B. Conant, *Harvard Case Histories in Experimental Science*, Harvard University Press, Cambridge, 1957, Vol. 1, p. 305.

[b]All data are the original values (with atomic weights changed to a scale in which O = 16.0, and calories converted to joules). If you use modern atomic weights rather than Dulong and Petit's as given in this table, does this make the molar heat capacities more nearly constant?

At room temperature the specific molar heat is almost constant for very different solids, as can be seen from the table. This experimental observation was raised to the rank of law for the first time in 1819 by Dulong and Petit. Today we know that this law is justified by the principle of energy equipartition. This principle states that each of the degrees of freedom connected with the different forms of energy is associated with an average energy equal to kT/2. Each atom of a crystal has three degrees of freedom and therefore, both considering the kinetic energy and the potential energy, we can say that it has an average energy equal to 6(kT/2) = 3kT.

The total energy of a solid formed by N atoms is therefore equal to $3kTN = 3kTn_mN_A$ (N_A = Avogadro's number, n_m = number of moles). The specific heat is the derivative of the internal energy with respect to T and therefore, by remembering that R, the constant of the perfect gases, is defined as $R = kN_A$ we can write:

$$c_v = 3kn_mN_A = 3n_mR$$

This shows that the specific heat referred to a single mole of a monatomic solid is a constant, independent from the atomic species, as stated by the law of Dulong and Petit:

$$c_v = 3R$$

The law can also be generalized to the case of polyatomic substances, as already did the same Dulong and Petit. To do this we observe that in n_m moles of a polyatomic substance the number of atoms is equal to $n_a(n_m N_A)$ where n_a is the number of atoms present in the stoichiometric formula of the compound. In the case of NaCl for example $n_a = 2$, in the case of water $n_a = 3$. By applying the principle of energy equipartition as done above we get that the specific heat is equal to $n_a 3R$. In the case of NaCl, for example, we observe that $c_V = 50\ JK^{-1}mol^{-1}$, about two times that of elementary solids. In the case of water, however, the value rises to $75\ JK^{-1}mol^{-1}$, about three times.

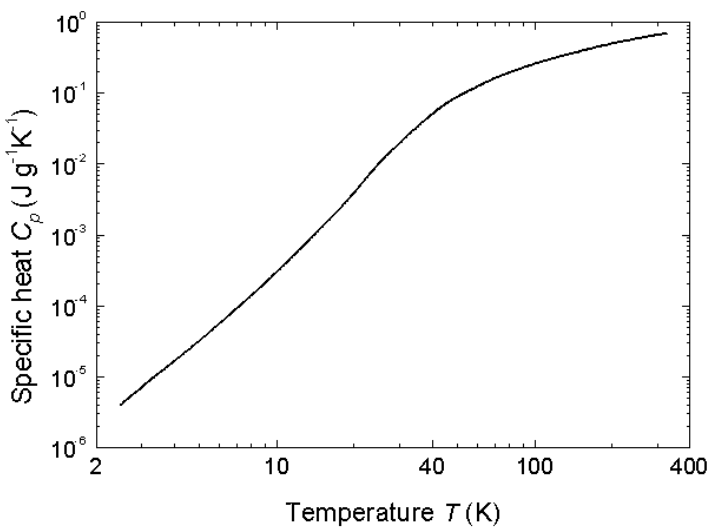
In reality, things do not work exactly in this way for at least two reasons.

- At low temperature the specific heat decreases to zero for T tending to zero (see the figure showing experimental data for Si). This specific heat behavior is also observed in the case of polyatomic gases. This behaviour can be interpreted by thinking that at high T the solid is able to receive energy by varying its temperature to a small extent while at low T this capacity is lost, as if the possibilities of storing energy decreased with decreasing T. Moreover, it is also in agreement with what follows from Nernst theorem, according to which entropy S ie equal to zero when the temperature is equal to the absolute zero. The entropy of a solid at temperature T can be expressed in this way:

$$S(T) = \int_0^T \frac{dQ}{T} = \int_0^T \frac{c_V}{T} dT$$

  If, for T = 0, the specific heat c_v assumed a value different from zero, the integral would diverge and it would not be possible to respect the Nernst theorem. Therefore, when the temperature goes to zero, the specific heat must do the same

- At high temperatures, the specific heat is not exactly a universal constant, but depends to some extent on the material.



To solve the first discrepancy we need to completely change the paradigm and formulate more sophisticated theories, based on quantum mechanics. For the second one we have to take into account the fact that in a real crystal the expression of the total energy is not a simple quadratic form of its coordinates and speed, and therefore the principle of energy equipartition takes a different form from

that previously stated. In other words, we must consider the anharmonic effects, which are also responsible for other important phenomena such as thermal expansion and the finite value of thermal conductivity.

In the following we will deal only with the first of these two aspects and we will show how the use of quantum concepts and statistical physics allows us to understand the behavior of specific heat at low temperature.

4.2 The solid as a gas of oscillators

The model used to explain the thermodynamic properties of a solid is based on the idea that it can be represented as a set of many harmonic oscillators, each of which corresponds to one of the normal modes of oscillation.

The first quantum theory of lattice oscillations was formulated by Einstein, and was based on the idea that each atom oscillates independently from the others and that the possible energy values of these oscillations are integer multiples of a quantum, whose value is proportional to the frequency of oscillation. This theory extended to the lattice oscillations the ideas that had led to explain the black body radiation, i.e. the oscillations of the electromagnetic field in a cavity. Einstein's theory, however ingenious it may have been, was inadequate and contained unjustified ad hoc hypotheses.

In order to improve the Einstein's model we have to take into account the fact that the real oscillations of a lattice are a collective phenomenon and cannot be interpreted in terms of independent atoms (as we know from the classical theory of lattice vibrations) and this led to the formulation of the Debye model. A central concept in this model is the <u>density of states</u>, meant as the number of possible vibrational states as a function of their energy.

In the following we will formulate a suitable model to explain the trend of the specific heat as a function of temperature. The steps we will take are as follows:

- The quantization of the normal modes of oscillation.

- The quantization of the possible values of the wave vectors **q**.

- The density of vibrational states and its link with the dispersion relations.

- The average value of the energy connected with each normal mode of oscillation.

4.3 Quantisation of the oscillator energy and quasi-particles

As we have seen, in the case of the linear chain of atoms solutions are waves: $u_n(t) = u_o e^{i(qna - \omega t)}$. For each of the possible value of q, it is possible to write an independent equation:

$$\ddot{u}_n(t) + \frac{4K}{M}\sin^2\left(\frac{qa}{2}\right)u_n(t) = 0$$

This equations describes the behavior of a single harmonic oscillator. Since the possible values of q are all those within the first Brillouin zone, this leads us to represent the solid as a set of many

independent oscillators, each characterized by a particular value of q and frequency ω. Moreover, since the set of all the possible solutions of the problem constitutes a complete and orthogonal set, we can say that any oscillation of the lattice can be expressed as the sum of many plane wave. For this reason these collective oscillations are calle elementary excitations.

Until now, in the study of lattice oscillations we have used classical mechanics because the individual atoms have an associated wavelength very small compared to the lattice parameter and therefore can be treated as classical particles. This allowed us to conclude that the oscillations of a lattice can be seen as harmonic waves with wave vector q and pulsation ω. As known, in quantum mechanics a harmonic oscillator characterized by a pulsation ω has energy levels given by the formula:

$$E_n = \left(\frac{1}{2} + n\right)\hbar\omega$$

The separation between the levels is constant and its value is $\hbar\omega$, with $\hbar = 10^{-34}\ \mathrm{J \cdot s}$. Let us first consider a classic oscillator, for example a child on a swing that oscillates at 1 Hz, then $\omega = 2\pi$. The separation of the levels in this case is in the order of 10^{-33} J, or 10^{-14} eV. It is a very small value compared to the energy involved in any physical phenomenon in which this oscillator can be involved, so we can overlook the effects of quantization by treating the energy spectrum as a continuum and safely use classical mechanics.

Let us consider instead the lattice oscillations, where the oscillation frequencies are in the order of 10^{13} Hz. The separation of the levels in this case is about 6×10^{-21} J, or 30 meV. The order of magnitude is comparable with the thermal energy available for each atom at room temperature (25 meV) and therefore quantum phenomena become important. This is the same reason why, when studying the oscillations of individual molecules, it is necessary to take into account the quantization of the levels and the specific heat of the gases, calculated with the classic formula based on energy equipartition, is no longer good at low temperatures.

Quantized oscillators can only exchange energy in form of discrete "packets" $\hbar\omega$. This exquisitely quantum behavior is observed in all excitations that take the form of waves (such as lattice oscillations or electromagnetic waves): it is the well-known wave-particle duality. In the case of electromagnetic waves, these packets, or "particles", are called photons. In the case of lattice oscillations on the other hand we speak of phonons and instead of the term particles we prefer to use the term quasi-particles.

In addition to possessing energy, elementary excitations are characterized by a wave vector **q**. In quantum mechanics, for free particles, the product $\hbar\mathbf{q}$ represents the momentum. In the case of quasi-particles such as phonons, this interpretation cannot hold. Let us just think about the fact that a lattice that is oscillating has a momentum (in the classical sense) equal to zero.

But there is also a more fundamental argument to say that $\hbar\mathbf{q}$ can't really be the momentum. The argument is that the momentum of an isolated physical system is preserved if, as a result of the homogeneity of space, the system is invariant for any translation (the famous Noether theorem). In the case of lattices, instead, there is invariance only for discrete translations, therefore the momentum cannot be one of the constants of the motion. For this reason, quantity $\hbar\mathbf{q}$ is called crystal momentum Although it does not coincide with the usual momentum, the crystal momentum has a very important role because for it a conservation law can be formulated valid for the scattering events where quasi-particles are involved or when the solid interacts with external fields (electromagnetic waves, particles, ...). It is a law that resembles the well-known law of conservation of momentum, but it is not exactly the same, it is almost the same, as we will see.

Basic quantum mechanics teaches us that what we are normally used to consider as particles are actually also waves. In the present discussione we are proceeding in the opposite direction, we started from waves (the lattice oscillations) and we are sticking on them some particle characteristics (the quantization of energy levels that leads to the concept of quasi-particle). This way of "quantizing" classical objects has given rise to a formalism called second quantization, widely used in theoretical physics and applied to fields of all kinds, from the electromagnetic field to other fields associated with the elementary particles of the standard model.

In solid state physics there are many elementary excitations. Besides the phonons there are the collective excitations of the electron gas, and the corresponding quasi-particles are called plasmons (which have energies in the order of some eV). Then there are the excitations of the spin distribution, known as magnons whose energies are intermediate between those of the phonons and those of the plasmons. Obviously, as with phonons, each of these quasi-particles is characterized by characteristic dispersion relationships that link the frequency of oscillation to the wave vector.

To sum up, an oscillating lattice can be considered as a set (a "gas") of many quasi-particles, the phonons, and its state can be described by assigning the number of phonons n(**q**,s) for each of the possible modes of oscillation. Each oscillator is identified with its wave vector **q** and the mode of oscillation s. For now we will assume that there is no interaction between the quasi-particles, which follows from operating in the harmonic approximation.

The first observation to make on this hypothetical gas is that the number of particles present n(**q**,s) can assume any value, and therefore the phonons must be treated like bosons. Furthermore, the number of phonons present in a lattice is not necessarily constant, since it depends on temperature. Finally, it should be noted that even in the absence of phonons the energy of the system never goes to zero due to the presence of the so called zero point energy (the term ½ in the formula of the energy levels of the harmonic oscillator).

4.4 The boundary conditions

Up to this point we have assumed that the crystals we are dealing with are of infinite size even if, of course, all real solids have finite dimensions. Making this clarification is not very important if the size of the solid is large enough to make surface effects negligible. However, it is important from another point of view, which we shall now clarify.

When considering a finite system, boundary conditions must be defined. The first observation to make is that the choice of these conditions must not affect the results that descend from them.

We are considering excitations of a periodic structure. The simplest case we have dealt with is the chain of atoms separated by a distance a. Now we are saying that we are only interested in a finite sample formed by N atoms, therefore of total length L = Na. An easy way to impose a boundary condition on the functions that define the excitations of the system is to set to zero, regardless of time, the value of the displacement at the ends of the chain. In the case of displacement u of the atoms means:

$$u(1) = u(N) = 0$$

This hypothesis is equivalent to saying that the excitations of the structure are standing waves. This, however, is not adequate if we want to study excitations that can propagate within the structure. To do this, the so called Born-Von Karman (or cyclic) conditions are adopted. In this case it is assumed

that the sample of finite length repeats itself periodically in space (or that it closes on itself, which is why we also speak of cyclic conditions) and that the excitations present within the structure are periodic with periodicity equal to the length of the sample:

$$u(i, t) = u(i + Na, t)$$

In order to realize that we are considering a real excitation, we have highlighted the dependence on time.

From the point of view of this discussion it is not particularly important which of these boundary conditions we use, in some contexts however it may be more convenient to adopt the conditions of Born-Von Karman as for example when we will talk about the theorem of Bloch in relation to the electron states.

4.5 Quantization of vectors q

Up to this point we have assumed that **q** can assume, within the first Brillouin zone, any value. This is true if the structure is infinite and therefore the waves can have an arbitrary wavelength $\lambda = 2\pi/q$. If, on the other hand, we imagine that the structure is made up of a finite number of cells, then <u>the number of possible values for **q** is finite and equal to N</u>.

To demonstrate this statement in one dimension let us suppose that the finite size sample consists of N cells of width a and has total length L = Na. By imposing the above boundary conditions, only waves with a length equal to L or one of its submultiples can be observed:

$$\lambda_n = \frac{L}{n}$$

From the point of view of the wave vectors this means that the values of q must satisfy the condition:

$$q_n = \frac{2\pi}{\lambda_n} = n\frac{2\pi}{L} = n\frac{2\pi}{Na}$$

Obviously, assigned N, the possible values for n are only 0, 1, 2, ... N − 1. In fact, if we place n = N we get the value $q = 2\pi/a$, which is the first vector of the reciprocal lattice. We can therefore say that, by varying n between 0 and N, an entire cell of the reciprocal lattice is covered. We know that the dispersion relations are periodic in the reciprocal lattice and therefore that all possible vibrational states can be traced back to those present in a single cell of the reciprocal lattice, typically the first Brilluoin zone.Therefore, if there are N cells in the direct lattice, there are also N possible distinct values of q in the reciprocal space.

The separation Δq between two contiguous values of q is:

$$\Delta q = q_{n+1} - q_n = (n+1)\frac{2\pi}{L} - n\frac{2\pi}{L} = \frac{2\pi}{L}$$

The possible values of q are therefore <u>equally spaced</u> in the reciprocal space. To give an idea of the orders of magnitude, if L = 1 cm and the lattice parameter a = 3 Å, the Brillouin zone extends up to $q = 2\pi/a = 2.1 \times 10^{10}\ m^{-1}$ and $\Delta q = 2\pi/L = 628\ m^{-1}$. As we can see, the separation between the levels

in a macroscopic sample is enormously small compared to the extension of the zone and therefore the distribution of q vectors can safely be considered as continuous. However, the fact that we have quantized the values of q allows us to do something very important: count them.

More than spacing, we are interested in the density of the possible values of q in reciprocal space:

the density of the vectors q is the inverse of q and is obviously constant and equal to $L/2\pi$.

The fact that the density of vibrational states is constant in the reciprocal lattice allows us to immediately calculate the number of states present in a known portion of the reciprocal space.

Let us now turn to the 3D case of a macroscopic crystal formed by $N = N_1N_2N_3$ cells, where N_i represent the number of cells along each of the three fundamental directions. Repeating the same reasoning done in 1D for all three directions, it is easy to verify that in the reciprocal lattice there are N possible distinct vectors **q**. These vectors are uniformly distributed within the first Brillouin zone and therefore each of them has at its disposal a portionq of volume in the reciprocal space equal to:

$$\Delta\mathbf{q} = \frac{V_{rec}}{N}$$

Where V_{rec} is the volume of the elementary cell of the reciprocal lattice. We call V_{dir} the volume of the elementary cell of the direct lattice. When we studied the reciprocal lattice, we saw that this relation holds:

$$V_{rec} = \frac{(2\pi)^3}{V_{dir}}$$

If in the volume V_{rec} there are N vectors **q** we can express in this way the volume $\Delta\mathbf{q}$ in the reciprocal space available to each of the vectors **q**:

$$\Delta\mathbf{q} = \frac{V_{rec}}{N} = \frac{1}{N}\frac{(2\pi)^3}{V_{dir}} = \frac{(2\pi)^3}{V}$$

Where $V = NV_{dir}$ is the volume of the sample. As we can see, each vector **q** occupies a portion of the reciprocal lattice inversely proportional to the volume of the crystal, in analogy with the 1D case where for each q there is at disposal a length inversely proportional to the total length L. In the 3D case the density of the vectors **q** in the reciprocal space, that is the inverse of **q**, is $V/(2\pi)^3$.

4.6 Density of states

The vibrational energy of a crystal is the sum of the energies of all the oscillators that make it up:

$$E = E_o + \sum_{\mathbf{q},s}\left(\frac{1}{2} + n(\mathbf{q},s)\right)\hbar\omega(\mathbf{q},s)$$

E_o is the energy corresponding to the equilibrium position of the atoms, it is constant and its value can be set to zero. The second term expresses the total energy of a set of oscillators each characterized by the two indices: s (the mode of oscillation) and **q**. Each oscillator contributes to the total energy

with n quanta each one of them havint energy $\hbar\omega$. The frequency ω and the wave vector **q** are not independent but are linked by the dispersion relations.

If in the expression already seen above we neglect the constant terms, and we omit the index s for simplicity, we can write:

$$E = \sum_{\mathbf{q}} n(\mathbf{q},s)\hbar\omega(\mathbf{q},s)$$

Each possible oscillator, characterized by a particular value of **q**, contributes to the total energy with a multiple n of the corresponding quantum of energy $\hbar\omega$. Let us think about the structure of this formula by distinguishing different aspects.

- First of all, let us consider a single wave vector oscillator **q** and frequency ω. Its contribution to the total energy depends on ω, which in turn is a function of **q**, because its energy is a multiple of $\hbar\omega$.

- Always considering a single oscillator, its contribution to the total energy is n times this amount. One must therefore find the way to know n, and this is the task of statistical physics. We can already anticipate that n depends only ω on and is the same for all oscillators having the same frequency. This explains why ω is the most physically significant variable for energy considerations.

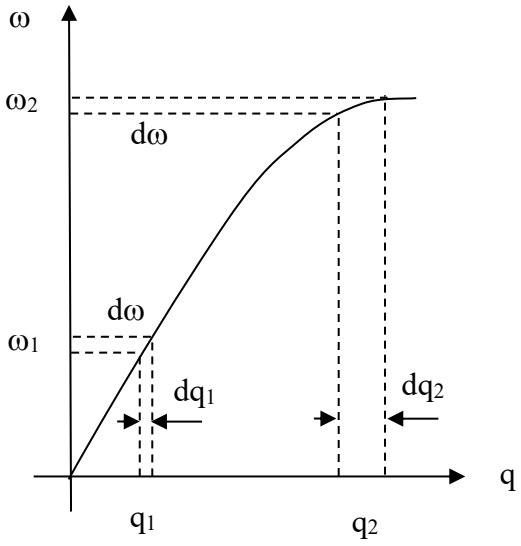
- The contribution to the total energy of the oscillators with frequency ω is therefore equal to the average energy of each of them (i.e. $n\hbar\omega$) times the number of oscillators of this type present in the crystal. It is therefore useful to have a function that expresses the number of oscillators for each frequency or better, a function g(ω) such that g(ω)d ω represents the number of oscillators in the interval dω centered on a particular vaule ω of the frequency. The function g(ω) is called <u>density of states</u>.

For the moment we know how to calculate, or measure, the dispersion relations $\omega_s(\mathbf{q})$. Now we are interested in moving from these to the density of states g(ω). Let us see how to proceed with reference to the simple case of the dispersion relation valid for a linear chain of atoms represented in the figure.

Let us consider an interval dω centered around a value of ω_1 of the frequency. We are interested in how many oscillators there are within this range. To do this it is better to consider the corresponding interval dq_1 and the reason for this is that in the reciprocal space the possible oscillators constitute a discretized set with constant density and this obviously makes them easy to count. The number of oscillators within dq_1 (and therefore within dω) is equal to the product of dq_1 times the density of the possible values of q which, as we know, in one dimensio is equal to L/2π. If we refer to a sample with a unit length (L = 1) the density is simply 1/2π. In this case then the number of oscillators is:

$$\text{Number of oscillators within } d\omega = g(\omega)d\omega = \frac{1}{2\pi}dq_1$$

The number of oscillators within the same frequency range depends on the value of the frequency around which it is centered. Consider in fact the case in which dω is centered around the frequency ω_2. The corresponding range dq_2 is wider, which means that the number of oscillators contained in it is higher.



As seen, the number of oscillators $g(\omega)d\omega$ is:

$$g(\omega)d\omega = \frac{1}{2\pi}dq$$

From which:

$$g(\omega) = \frac{1}{2\pi}\frac{dq}{d\omega}$$

In the expression appears the derivative of q with respect to ω and the result is a function of ω. To calculate this function we have to start from the dispersion relation $\omega(q)$ and derive its inverse function $q(\omega)$. Alternatively, the density of states $g(\omega)$ can be expressed directly in terms of the derivative of the dispersion relationship:

$$g(\omega) = \frac{1}{2\pi}\frac{dq}{d\omega} = \frac{1}{2\pi}\frac{1}{\frac{d\omega}{dq}} = \frac{1}{2\pi}\frac{1}{\omega'(q)} = \frac{1}{2\pi}\frac{1}{\omega'[q(\omega)]}$$

In this case the result obtained is a function of q which, however, we know to be a function of ω and this allows us to express the density of states in terms of frequency.

These formulas are important because they relate the density of states to the dispersion relations. We will come back to this subject later to talk about the properties of the density of states that can be deduced from this formula.

In the 3D case we proceed in the same way. What we need to know is the number of oscillators present in the portion of space between the surface $\omega(\mathbf{q})$ and the surface $\omega(\mathbf{q}) + d\omega$, which we can formally express in this way.

$$g(\omega)d\omega = \int_{\omega(\mathbf{q})}^{\omega(\mathbf{q})+d\omega} \frac{d\mathbf{q}}{8\pi^3}$$

If different modes of oscillations are present, then the corresponding density of the states must be calculated for each of them; the overall density will be the sum of all the individual ones.

In the previous expression d**q** represents the volume of an element of the reciprocal space that, multiplied by the density of the vectors **q** in the reciprocal space in three dimensions which is equal to $\frac{1}{8\pi^3}$, gives precisely the number $\frac{d\mathbf{q}}{8\pi^3}$ of vectors **q** present in the element d**q**.

To proceed, we express the volume of element d**q** as the product $dSdq_\perp$ of an element dS of the surface where ω is constant for an element $dq_\perp$ perpendicular to the surface:

Moving in the direction perpendicular to the surface, the frequency varies and we can write:

$$d\omega = \left|\mathrm{grad}_{\mathbf{q}}\omega\right| dq_\perp$$

In this way the volume of the element d**q** assumes the following expression:

$$dSdq_\perp = \frac{dSd\omega}{\left|\mathrm{grad}_{\mathbf{q}}\omega\right|}$$

and then, by integrating over the entire surface:

$$g(\omega) = \int_S \frac{1}{\left|\mathrm{grad}_{\mathbf{q}}\omega\right|} \frac{dS}{8\pi^3}$$

4.7 Van Hove singularities

The previously defined density of states g(ω) is used to express the number of states in each particular frequency range. In order to express it correctly, it is necessary to search in the reciprocal space for all the states characterized by the same frequency, whatever the value of their wave vector **q**.

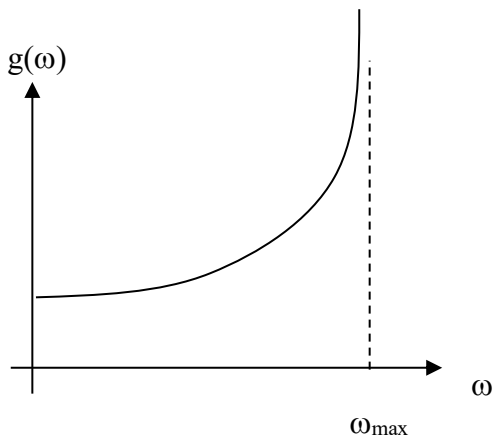
The functional relationship between density of states and frequency depends on the dimensionality of the system. If we consider only acoustic excitations, where the dispersion relation is linear, we can easily demonstrate that in 1D g(ω) is constant, linear in ω in 2D and quadratic in ω in 3D. The reality is different, however, since the dispersion relations of the acoustic modes are linear only for small values of **q** and this is reflected in the form of the g(ω) which is never a regular curve as in the simplified models but presents singularities known as Van Hove singularities, named after the Belgian physicist who first dealt with them in the 1950s. To understand the origin of these singularities it is worth considering the case of a linear chain of equal atoms whose dispersion relations is, as is known, the sine functions:

$$\omega(q) = 2\sqrt{\frac{K}{M}}\left|\sin\frac{1}{2}qa\right|$$

Previously we have obtained the relation that binds the density of the states $g(\omega)$ to the dispersion relations $\omega(\mathbf{q})$ and therefore we can calculate it for the case of the linear chain:

$$g(\omega) = \frac{1}{2\pi}\frac{1}{\omega'(q)} = \frac{1}{2\pi}\frac{1}{\cos qa}$$

This formula expresses the density of states as a function of q. However, we have seen that by inverting the dispersion relation and expressing q as a function of ω, we obtain the desired formula that relates the density to frequency. The graph of the density obtained is the one shown in the figure. For small values of ω the function is approximately constant, as it should be in the case of a linear dispersion relationship in 1D. For larger values, however, the trend changes until it reaches, at the edge of the zone, a singularity. This is a simple example of a Van Hove singularity.



The situation in 3D is more complex but even in this situation singularities in the densities of states are still present. They exist in different forms and occur at particular values of the wave vector **q** called critical points.

4.8 A few remind of statistical physics

At this point we are able to "count" the different types of oscillators present in a crystal. What we must now learn to do is to understand how the globally available energy is distributed among them. This aspect is the subject of statistical physics of which we will summarize the essential concepts useful for the study of specific heats.

A crystal is a thermodynamic system in thermal balance with the surrounding environment. To study its properties, therefore, the so called formalism of the canonical ensemble is used. The energy of the system can fluctuate and its value can assume any of the eigenvalues of the energy of the Hamiltonian describing the system itself, which we call E_n. The statistical nature of the problem manifests itself in the fact that we are not able to say, moment by moment, what is the value of the energy, we can only say what is the probability w_n that the energy is equal to E_n. The corresponding statistical distribution is the Gibbs distribution:

$$w_n = Ae^{-\frac{E_n}{kT}}$$

This law applies both in classical and quantum mechanics. The important point to note is that the energies E_n are not single particle energy levels, they are the possible values of the energy of the system intended as a whole (many-body). For this reason, the distribution is valid in general, also for systems of interacting particles. If, on the other hand, reference is made to the possible energy values of the individual particles in a gas of non interacting particles, then we are talking about the Boltzmann's distribution.

The constant A is obtained by imposing the normalization condition:

$$\sum_n w_n = A\sum_n e^{-\frac{E_n}{kT}} = 1$$

Hence:

$$A = \frac{1}{\sum_n e^{-\frac{E_n}{kT}}}$$

The distribution function is then written:

$$w_n = \frac{e^{-\frac{E_n}{kT}}}{\sum_n e^{-\frac{E_n}{kT}}} = \frac{1}{Z}e^{-\frac{E_n}{kT}}$$

Z is called the <u>partition function</u>:

$$Z = \sum_n e^{-\frac{E_n}{kT}}$$

The distribution function is used to calculate the average value of any physical quantity f that can assume the possible values f_n:

$$\langle f \rangle = \sum_n f_n w_n = \frac{1}{Z}\sum_n f_n e^{-\frac{E_n}{kT}}$$

This formula allows to calculate the value of any thermodynamic quantity, such as internal energy, entropy, and all other thermodynamic potentials. If we consider the free energy of Helmoltz $F = E - TS$, which represents the energy exchanged during a transformation where T and V are kept constant, working with the formalism of the canonical ensemble it is possible to demonstrate that:

$$F = -kT\ln Z$$

This formula is very important because it binds a thermodynamic quantity, the free Helmoltz energy, to the partition function which instead contains information on the microscopic states of the system.

The fact that the microscopic properties of the system are present in the partition function is easily understandable considering that all possible energy levels E_n appear in it, and therefore in order to be able to write its expression we must know all the possible microscopic states of the system and the corresponding energy values.

4.9 The general formula for the lattice specific heat

To begin let us consider a <u>single</u> quantum oscillator. Its energy levels are:

$$\varepsilon_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

The partition function of a single oscillator is therefore:

$$Z_{oscill} = \sum_{n=0}^{\infty} e^{-\frac{\varepsilon_n}{kT}} = \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega}{kT}\left(n+\frac{1}{2}\right)}$$

If the value corresponding to n = 0 is supposed to correspond to the zero of the energy, then the formula is simplified and can be written:

$$Z_{oscill} = \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega}{kT}n}$$

It is a series whose limit is:

$$Z_{oscill} = \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega}{kT}n} = \frac{1}{1 - e^{-\frac{\hbar\omega}{kT}}}$$

From which the free energy of a single oscillator is obtained:

$$F_{oscill} = -T\ln Z_{oscill} = T\ln\left(1 - e^{-\frac{\hbar\omega}{kT}}\right)$$

Remembering that the entropy $S = -\left(\frac{\partial F}{\partial T}\right)_V$ and that F = E − TS, we get an expression for the average energy of a single oscillator:

$$E_{oscill} = \frac{\hbar\omega}{\left(e^{\frac{\hbar\omega}{kT}} - 1\right)}$$

As we can see, the average energy of a single oscillator is equal to the value of the quantum of energy $\hbar\omega$ weighed for the term $1/(e^{-\hbar\omega/kT} - 1)$ which therefore represents the average number of quanta present in the oscillator. Put in another way, the number of particles present in the energy state $\hbar\omega$.

Passing to a real solid, the average energy due to the oscillations of the lattice is calculated by summing up all the possible oscillators identified by the two parameters **q** and s:

$$E = \sum_{\mathbf{q},s} \frac{\hbar\omega(\mathbf{q},s)}{\left(e^{\frac{\hbar\omega}{kT}} - 1\right)}$$

This formula tells us that the contribution of each oscillator depends only on its frequency, and that to calculate the total energy it is necessary to know how many oscillators are present in correspondence of each of the possible values of the frequency. In order to provide this type of information, the density of states g(ω) has been introduced previously, which in the development of the calculations is more convenient to use instead of the summation on the **q** and s indices.

Once the formula for the vibrational energy of the lattice has been found, the specific heat is calculated by deriving the expression obtained with respect to temperature. In performing this calculation in general it is necessary to know, or make some assumptions, concerning the density of the states. However, there are two limiting cases for which it is not necessary to know in detail g(ω). These are the cases where temperature assumes very high values, and we will specify with respect to what, and values close to the absolute zero. We will begin just dealing with these two particular cases, immediately after we will deal with a model valid at all temperatures, the model of Debye.

4.10 Specific heat for large and small values of T

<u>High T</u>

For large temperature values, or better for large values of the kT energy with respect to the quantumenergy ħω, the exponential term in the expression of the energy of a single oscillator is small and therefore it is possible to expand the formula in series of ħω/kT stopping at the first order term:

$$E_{oscill} = \frac{\hbar\omega}{\left(e^{\frac{\hbar\omega}{kT}} - 1\right)} \simeq \frac{\hbar\omega}{\frac{\hbar\omega}{kT}} = kT$$

This result is nothing more than the theorem of energy equipartition according to which the average energy per oscillator is equal to kT (half due to kinetic energy and half due to potential energy). If the solid is formed by N cells, each containing n atoms, then there are 3nN oscillators, each of which has an average energy equal to kT. The total energy of the solid is therefore equal to 3NnkT and the corresponding specific heat is equal to 3nNk. The fact that we are assuming a high temperature value means assuming that all these oscillators are excited. The result obtained is also in accordance with the empirical law of Dulong and Petit, as we saw at the beginning of this chapter.

Small T

In this case the calculations are more complex and are based on the assumption that the only modes of oscillation excited are those of the acoustic type, a reasonable hypothesis since they are the only ones for which there are always possible modes of oscillation even if the values of the quanta of energy is close to zero. The important result is that, for a 3D system, the specific heat dependence at very low temperatures is proportional to T^3.

4.11 The Einstein Model

Based on the idea of quantized lattice oscillations Einstein, in 1907, proposed a model which traced the Planck model of the oscillations of the electromagnetic field in a cavity. The procedure adopted was the same. The initial assumption was that the solid is made up of many equal oscillators. To put it in a more actual language, this corresponds to a density of states $g(\omega)$ which is a delta centered on a particular value of frequency.

This assumption stemmed from the fact that diamond, to which the original article referred and of which c_V down to 50 K was known, had a net absorption line in the IR that was associated with a fixed frequency oscillator. In its original formulation the Einstein model contained two errors, but they compensated each other. The first was that the contribution due to the zero point energy was neglected, the second was that the Boltzmann statistics were used. These two errors actually compensate each other for a fortuitous reason and therefore Einstein obtained a result that describes the experimental data quite well. The same pair of mutually compensating errors was present in the deduction of the Plank formula for black body radiation. The correct formulation of both problems had to wait for the full development of quantum mechanics and of quantum statistical physics.

4.12 The Debye model

A more refined model was proposed by Debye in 1912. Instead of a set of oscillators all equal, the model is based on hypotheses more appropriate to the reality, even if it contains in turn many rather drastic approximations.

- First of all, it is assumed that there are only acoustic phonons and approximating the dispersion relationship with a straight line. The slope of the line is the speed of sound, assumed to be constant.

- We approximate, in the reciprocal space, the first Brilluoin zone wih a sphere of radius q_D (called the Debye sphere whose radius is the Debye wave vector). The volume of the Debye sphere is therefore $\frac{4}{3}\pi q_D^3$. To obtain the value of q_D we keep in mind that within the first Brillouin zone, here approximated with the sphere of Debye, must find place all the possible N wave vectors **q**. N, as we know, is equal to the number of cells of which the lattice is constituted. We also know that **q** vectors are equally spaced in the reciprocal space and that the volume available for each of them is $\Delta\mathbf{q} = \frac{(2\pi)^3}{V}$. From this we obtain:

$$N\frac{(2\pi)^3}{V} = \frac{4}{3}\pi q_D^3$$

Bearing in mind that $V/N = V_{dir}$, the volume of the elementary cell, we can write:

$$q_D = \left(\frac{6\pi^2}{V_{dir}}\right)^{\frac{1}{3}} \simeq \frac{4}{a}$$

Where a is the lattice parameter (in a simple lattice in 1D, π/a is the boundary of the Brillouin zone).

To the Debye wave vector correspond a frequency ω_D, called the Debye frequency, a Debye energy $\hbar\omega_D$ and a temperature of Debye Θ_D defined by the relation:

$$k\Theta_D = \hbar\omega_D$$

Under these assumptions it is possible to calculate the DOS which comes out to be proportional to ω^2. To understand this type of dependence we return to the definition of density of states $g(\omega)$. This represents the number of possible states for an assigned value of frequency. We are moving in the reciprocal space where the possible **q** vectors are distributed with constant density. In 3D this is equivalent to choosing a sphere of assigned radius q, to which corresponds a precise frequency $\omega = v_s q$, being v_s the speed of sound in the materials. The number of oscillators at this frequency is proportional to the surface of the sphere of radius q, that is to say to q^2 and therefore, given the linear link between q and, to ω^2. The same reasoning leads us to say that, in 2D, the DOS is proportional to ω and that in 1D it is constant.

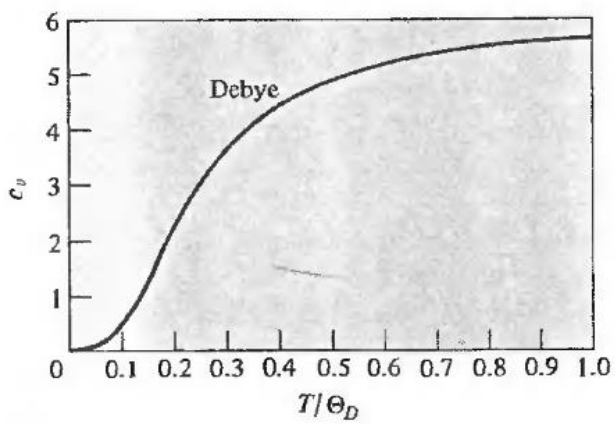


Figure 23.3
Specific heat in the Debye approximation (in cal/mole-K) vs. T/Θ_D. (From J. de Launay, *op. cit.*; see Table 23.2.)

In the development of the Debye model, once calculated the DOS, we write the expression of the energy density that, derived with respect to T, provides the specific heat. The result obtained correctly describes the trend of c_V at all temperatures.

Above the Debye temperature the specific heat varies little, and this indicates that all modes of oscillation are excited. Below that, on the other hand, there is their progressive freezing. Under these conditions, the same amount of energy is distributed among a smaller number of oscillators and therefore produces a greater increase in temperature. The same thing that happens if the same amount of energy is supplied to a smaller mass of material.

4.13 Lattice and electronic specific heat

The specific heat of a metal, aside from the phonon specific heat c_{ph}, must also take into account the contribution of the electrons c_{el}:

$$c_{tot} = c_{ret} + c_{ele}$$

In reality, the specific electronic heat is completely negligible at ordinary temperatures and the reason is to be sought in the fact that the electrons are fermions that obey different statistical laws with respect to the phonons, which are bosons. Bosons can move freely between energy levels, regardless of whether they are already occupied. Electrons, on the contrary, can only move on empty levels and this makes the whole system much more rigid and reluctant to absorb energy. The specific electronic heat at room temperature actually observed is about 100 times smaller than we might expect by applying the energy equipartition to electrons too.

Things change when temperature becomes very low, where the electronic specific heat varies linearly with T, while the lattice one scales as T^3. This means that, below a certain temperature value, the specific electronic heat becomes more prevalent than the lattice heat.

It is shown that the ratio of the two specific heats is:

$$\frac{c_v^{el}}{c_v^{ph}} = \frac{5}{24\pi^2} Z \frac{\Theta_D{}^3}{T^2 T_F}.$$

Z is the number of free electrons for each lattice cell and T_F is the Fermi temperature. From here we obtain that the phononic contribution prevails for temperatures above:

$$T_0 = 0.145 \left(\frac{Z\Theta_D}{T_F}\right)^{1/2} \Theta_D.$$

The Fermi energy ($E_F = kT_F$) is of the order of some eV, and therefore T_F is of the order of tens of thousands of K. In conclusion we can say that the electronic contribution to the specific heat manifests itself only at very low temperatures, of the order of $10^{-4}\,\Theta_D$, which approximately means some tens of mK.

5. Lattice thermal conductivity

5.1 Phenomenology and macroscopic description

A crystal lattice is capable of transporting thermal energy by means of its vibrations. The heat flow in a material is described at a macroscopic level by the so called Fourier law:

$$\mathbf{j} = -\,K\ \mathrm{grad}T(\mathbf{r})$$

j is the density of heat flow, i.e. the amount of energy that flows in the unit of time through a unitary surface element perpendicular to **j**. T is the temperature, which depends on the **r** position, and K is a phenomenological parameter called thermal conductivity.

Thermal Conductivity of Common Materials (at 25° C)

Material	Conductivity (Watts/meter-°C)
Acrylic	0.200
Air	0.024
Aluminum	250.000
Copper	401.000
Carbon Steel	54.000
Concrete	1.050
Glass	1.050
Gold	310.000
Nickel	91.000
Paper	0.050
PTFE (Teflon®)	0.250
PVC	0.190
Silver	429.000
Steel	46.000
Water	0.580
Wood	0.130

Extreme cases

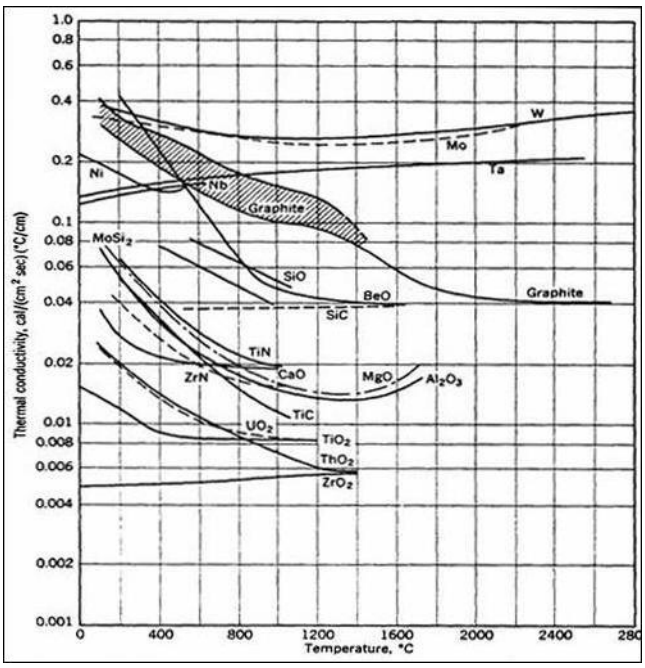
Diamond 2000 W/m°C

Silica Gel: 1.7×10^{-5} W/m°C

Unlike specific heat, thermal conductivity is highly dependent on the material, as can be seen from the table. In metals the conductivity is usually higher because of the presence of the electron gas, which in the case of thermal conductivity is important while we know that the specific electronic heat prevails over the lattice one only at very low temperatures and is negligible at room temperature. Although conductivity tends to be higher in metals, the material with the highest thermal conductivity is an insulator, the diamond.

In metals, as can be seen from the figure, thermal conductivity is weakly dependent on temperature. In insulators, on the other hand, the behaviour is more complex and the variations very large. Our aim is to explain this trend in the case of crystalline insulators.

In the following figure it is possible to observe the comparison between a crystalline material, the same material with many radiation-induced defects and the amorphous form of the material. This comparison allows us to make an important observation immediately: K is much greater in the ordered systems and this proves that the lattice oscillations certainly play a relevant role in carrying thermal energy.



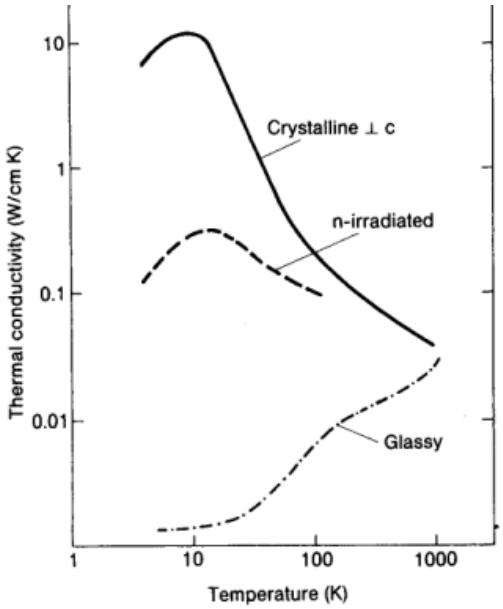


Fig. 5.7. Thermal conductivity of simple-crystal SiO_2 (quartz) perpendicular to the crystal *c*-axis. Also shown are the corresponding curves for the same crystal with defects induced by neutron bombardment, and for quartz glass [5.5, 5.6]

5.2 Modelling

In the following we will deal with thermal conductivity in insulating crystals, in which thermal conduction is due exclusively to the lattice oscillations. We will proceed step by step by introducing a model based on the following hypotheses.

A) We start by describing the propagation of vibrational energy as if there were particles, or rather packets of spatially localized waves, that propagate inside the crystal. We will use the term localized phonons for these particles.

B) The equation describing the propagation of these particles is written. In the more general case we have to write the Boltzmann equation, here we will limit ourselves to a version of it based on the approximation of the relaxation time τ. It is assumed that the localized phonons propagate undergoing scattering with a frequency $1/\tau$. This leads to write an expression for the conductivity K as a function of various parameters, the most uncertain of which is precisely τ.

C) The possible scattering mechanisms of the localized phonons will be studied in order to estimate the value of τ, and therefore of K. We will be particularly interested in their temperature dependence. There are many possible mechanisms. There are scattering with the impurities of the material. Then there are those with the sample surfaces. These are very effective mechanisms, but we will not deal with them in detail to talk instead about the collisions between phonons, much more interesting and instructive, which will force us to abandon the idea of the harmonic lattice on which we have based the study of the lattice oscillations.

Let us proceed according to this scheme, considering in detail the three previous steps.

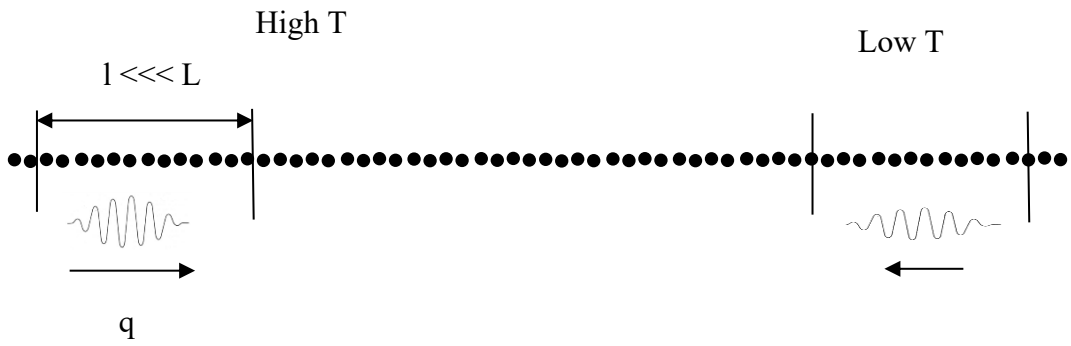
5.3 Localized packets of phonons

To model thermal conduction we can divide the crystal in which there is a temperature gradient in many small regions. Each of these regions, which we will identify by means of a coordinate x in the one-dimensional case considered first, must have the following characteristics:

1) Be small enough in relation to the size of the body so that inside it the temperature can be considered constant by making it possible possible to assign the temperature T as a function of position: $T = T(x)$.

2) Be extended enough to contain a large number of atoms and thus make it possible to use the tools of statistical physics, starting from the same possibility of defining the local temperature.

What this may mean in practice is understood by the following considerations. Refer to the figure showing a linear chain of atoms.



The transport of heat in a solid insulator is due to lattice oscillations, i.e. waves that propagate while transporting energy. In an infinite solid, and in equilibrium conditions, for each wave with wave vector **q** there is another one with wave vector −**q** and therefore there is no net energy transport.

To have a net flow of heat we must have a temperature gradient and this means that the system must be in a situation of non equilibrium. Having to deal with a system characterized by spatial

inhomogeneity, it is easy to understand that the use of completely delocalized plane waves to model physical phenomena is not the most appropriate choice.

In order to formulate a model of the phenomenon it is preferable to be able to attribute the transport to objects spatially localized but nevertheless characterized by a precise value of the wave vector **q**. We want to operate with objects that possess the requirements of the classical particles, for which it is always possible to assign the position and the momentum. We know that it is possible to construct objects with these characteristics by superimposing different waves whose vector **q** is within a certain interval $\Delta\mathbf{q}$. In the case of thermal conduction we will build localized packets of lattice waves and call them localized phonons.

Let us see what it means and what constraints it involves to build a packet with a defined value of q, thinking in one dimension. As known, a wave packet has an uncertainty both in the position x and in the value of q, and the two are linked by the known relation:

$$\Delta q \Delta x \approx 2\pi$$

To say that the value of q must be "well defined" means that the amplitude Δq of the packet must be small compared to the central value of q of the packet. Let us imagine we want to build localized phonons with a value of q in the order of the Debye wave vector, i.e. vectors that are near the edge of the first Brillouin zone. In 1D the maximum value of q is equal to π/a, being a the lattice parameter. Translated into formula the condition becomes:

$$\Delta q \ll \frac{\pi}{a}$$

By replacing this inequality in the uncertainty relationship we get:

$$\Delta q = \frac{2\pi}{\Delta x} \ll \frac{\pi}{a}$$

Hence the condition:

$$\Delta x \gg 2a$$

In other words, in order to be considered with good approximation as classical particles with an assigned position and momentum, localized phonons must extend over several lattice parameters. If, for example, we want an uncertainty $\Delta q \approx 1\%$ for packets where $q = \pi/a$ means that they must extend over 200 lattice parameters, i.e. something like 50 nm.

If we consider a macroscopic body whose size is in the centimeters range, on the basis of the above reasoning let us consider small regions within which T can be considered as constant. If we assume regions 10 μm wide the condition 1) defined bove is certainly satisfied. If, as we have hypothesized, the phonons extend for 50 nm, condition 2) is also fulfilled.

In the study of the specific heats it was seen that the whole lattice can be modeled as a set of N distinct oscillators, each characterized by a particular value of q and, through the dispersion relations, by a precise value of the oscillation frequency ω. The packets we are considering are characterized by a value of q defined with sufficient accuracy, and this because we impose the condition that $\Delta q \ll q$. This allows us to treat each of these packets as a harmonic oscillator with a well defined frequency ω

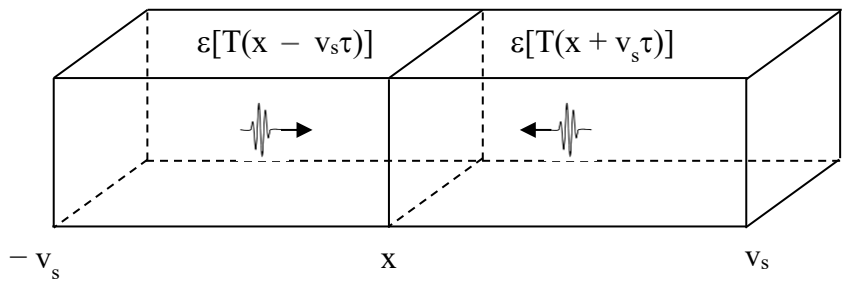
and then we can use the same approach with the only caveat that each oscillator of the ensemble is not associated with a completely delocalized wave but with a spatially localized packet.

5.4 The transport equation

The localized phonons defined above are the particles for which the propagation equation must be written. The easiest thing we can do is to use the same method used to calculate the electronic thermal conductivity in the Drude model, the only difference is that in that case the particles are electrons. The basic ingredients of this model are those of the so called relaxation time approximation:

- Localized phonons undergo diffusion processes every τ second; τ it is the relaxation time that is assumed to be independent on the position and speed of the particles. Between one scattering and another the phonons propagate freely, like the molecules of a gas.

- Scattering is the physical mechanism by which the local thermal equilibrium is maintained. This means that the energy distribution of the phonons coming out of the collisions that occur in a particular position is that of equilibrium and depends solely on the local temperature. This also means that the energy of the phonons before the scattering events has no effect on their energy distribution after the scattering took place. By virtue of this hypothesis it is possible to talk about the local value of temperature T(x) and the "relaxation" that gives the name to the model is that towards the state of local thermal equilibrium.



Imagine that the propagation of packets is one-dimensional along an x-axis. We can think that each of these localized packets carries a quantity of energy, which obviously depends on the temperature: $\varepsilon = \varepsilon(T)$. Since we are considering each of these packets as a single harmonic oscillator, the possible values of the energy it carries are:

$$\varepsilon = \left(\frac{1}{2} + n(T)\right)\hbar\omega$$

n is the average quantum number of energy in each packet, which obviously depends on the temperature T. The temperature, however, depends on the position x, and therefore $n = n[T(x)]$ from which it descends that $\varepsilon = \varepsilon\,[T(x)] = \varepsilon\,(x)$.

Let us suppose that the scattering events bringing to thermal equilibrium occur in a point in position x. The phonons that are scattered at point x are those that have undergone the previous scattering at a distance $l = v_s\tau$ (coming from both directions), being v_s the packet propagation speed, supposed to be constant and equal to the speed of the sound in the material (acoustic phonons).

The equation we are trying to write must express the energy flow j(x), i.e. the amount of energy flowing in the unit of time through a surface of unit area, perpendicular to the direction of propagation, at an arbitrary point x. The j(x) flow is calculated by adding together the contribution of the packets coming from the hottest areas of the sample, supposed to be on the left of x, and those of the packets coming from the right, where the coldest areas are located. Each of these two contributions is obtained by multiplying two terms:

1. The number of packets that pass through (i.e. the flow of phonons regardless of their energy). To understand what it is about we first refer to those coming from the left, similar considerations apply to the packets coming from the right. Since all packets travel at speed v_s, through the unitary surface perpendicular to x in the unit time will pass all the phonons contained in the parallelepiped of unitary basis and height v_s. This number is obtained by multiplying the volume of the parallelepiped (which is equal to v_s since the surface of the section we are considering is unitary) by the density ρ of phonons divided by two (this is because among all the phonons inside the parallelepiped, half go in one direction and half in the opposite one).

   The density ρ depends, as we know, on temperature. In the case we are considering however, within a small distance (such the height of the parallelepiped)from point x we can assume that ρ is constant and that the effects due to the dependence on T are of a higher order and therefore negligible. In the development of the calculations then we will see that it will be possible to eliminate the term ρ in a physically significant and clarifying way.

2. The energy carried by each packet, that is $\varepsilon(T[x - v_s\tau])$ for those coming from the left and $\varepsilon(T[x + v_s\tau])$ for those coming from the right. The energy of each packet obviously depends on its frequency of oscillation. Since there are packets distributed over all possible frequency values, it is intended that the energy we are talking about represents the average value of the packets energy.

Ultimately, the flow of energy from the hottest areas is $\frac{1}{2}\rho v_s \varepsilon\left[T(x - v_s\tau)\right]$. This is a flow to which we arbitrarily assign a positive value. In the same way, we calculate the flow from the left, which will be negative. The total flow is the sum of the two:

$$j(x) = \frac{1}{2}\rho v_s \left\{\varepsilon\left[T(x - v_s\tau)\right] - \varepsilon\left[T(x + v_s\tau)\right]\right\}$$

Now let us suppose that the distance between two scattering events is small from the point of view of temperature variations so that we can make the following developments in series:

$$T(x - v_s\tau) \approx T(x) - \left(\frac{\partial T}{\partial x}\right)_x v_s\tau \quad ; \quad T(x - v_s\tau) \approx T(x) + \left(\frac{\partial T}{\partial x}\right)_x v_s\tau$$

and so:

$$\varepsilon\left[T(x - v_s\tau)\right] \approx \varepsilon(x) - \frac{\partial\varepsilon}{\partial T}\left(\frac{\partial T}{\partial x}\right)_x v_s\tau \quad ; \quad \varepsilon\left[T(x + v_s\tau)\right] \approx \varepsilon(x) + \frac{\partial\varepsilon}{\partial T}\left(\frac{\partial T}{\partial x}\right)_x v_s\tau$$

from which:

$$j(x) = \frac{1}{2}\rho v_s\left[-2\frac{\partial\varepsilon}{\partial T}\left(\frac{\partial T}{\partial x}\right)_x v_s\tau\right] = -\ \rho v_s^2\tau\frac{\partial\varepsilon}{\partial T}\left(\frac{\partial T}{\partial x}\right)$$

To move on to the 3D case and obtain the vectorial flow **j** we remind that, if the distribution of the velocities is isotropic, the average square value of the velocity modulus can be expressed in this way:

$$\left\langle v_{s,x}^2\right\rangle = \left\langle v_{s,y}^2\right\rangle = \left\langle v_{s,z}^2\right\rangle = \frac{1}{3}\left\langle v_s^2\right\rangle$$

Now let us eliminate the phonon density and replace it with a quantity with a more immediate physical meaning, the specific heat:

$$\rho\frac{\partial\varepsilon}{\partial T} = \frac{N}{V}\frac{\partial\varepsilon}{\partial T} = \frac{1}{V}\frac{\partial E}{\partial T} = c_V$$

and so:

$$\mathbf{j}(x) = -\frac{1}{3}c_V v_s^2\tau\nabla T = -K\nabla T$$

where:

$$K = \frac{1}{3}c_V v_s^2\tau$$

At this point, we have achieved two results:

- We have derived from microscopic considerations the phenomenological equation of Fourier.
- We have linked the value of K to other physical quantities.

Of the three parameters that appear in the expression of K, the one that is clearly less known is precisely the relaxation time, which obviously depends on the details of the diffusion mechanisms.

5.5 Scattering mechanisms

Localized packets that in our model are responsible for heat conduction are subject to numerous types of interaction during their propagation. Among these, it is worth mentioning the diffusion due to impurities and defects in the material. In addition to these, there are also effects produced at the edges of the sample and the interfaces that separate the different parts of the material from each other.

In addition to these mechanisms, there are also those by which the lattice waves interact with the other particles present in the material. If we are in a metal we can certainly not ignore the presence of conduction electrons. In addition, it is possible to think that there is also an interaction between the same packets of waves, what we will call scattering between phonons, or phonon - phonon scattering, and which we will deal with diffusely. This does not mean that all other mechanisms are not

important, but from a conceptual point of view there is much to learn by dealing with the interaction between phonons.

5.6 The phonon - phonon interaction

The harmonic lattice

In the harmonic lattice, the plane waves, each characterized by its wave vector **q**, are normal modes of the system and therefore, once created, they remain indefinitely without there being any transfer of energy to other modes.

From this it follows that packets constructed by superimposing waves of different **q** propagate without attenuating. Moreover, if the dispersion relation is linear, the speed is the same for all the components and the packet propagates maintaining its own shape. If there is dispersion, the packet changes shape but there is no energy dissipation.

If the localized packets propagate without encountering any resistance we can have an energy flow even without a temperature gradient. To say that we can have a heat flow without a temperature gradient is to say that the thermal conductivity is infinite. In fact, the heat flow is proportional, through the thermal conductivity, to the temperature difference, so if its value is finished without there being a temperature difference it means that K must be infinite. This, however, is not physically meaningful and is not what is observed experimentally.

More, let us think of two packets of waves that travel one against the other. The two will cross each other and continue to propagate without one condition at all the other. This stems from the fact that, in the harmonic approximation, the principle of superposition of the effects applies and therefore the sum of two distinct solutions of the equations of the system is still a solution. Put in another way, there is no interaction between the two packets.

To summarize:

- The conductivity of a harmonic lattice is infinite.

- In a harmonic lattice there is no scattering between phonons.

All this leads us to think that, in order to improve the model and to introduce the possibility of collisions between phonons, the harmonic approximation must be at least correct by introducing terms of a order higher than the second in the expression of the potential energy.

The phonon-phonon interaction

Anharmonic terms, i.e. of an order of more than two, however important they may be in order to overcome the difficulties described above, must not be such as to upset what has been said up to now. The reason is that the theory developed so far, based on the harmonic approximation, works very well and makes it possible to predict the existence of phonons and to correctly calculate the dispersion relations $\omega(\mathbf{q})$.

It is therefore physically justified to add anharmonic terms to the hamiltonian that in any case we will assume to be small compared to the harmonic terms:

$$H = H_{arm} + H_{anharm}$$

The hypothesis that the additional term is small compared to the harmonic one allows us to use the theory of time-independent perturbarions to evaluate the effects produced by the anharmonic term.

The presence of this term leads to tiny corrections of the energy values that do not interest us. What interests us most is to verify that, thanks to the presence of this term, the interaction between phonons becomes possible.

First of all, it is advisable to describe the states of the system using the so called <u>occupation number representation</u>, i.e. specifying how many phonons are present in each of the oscillators that make up the system (each identified by the wave vector **q** and the mode index s). In other words, a particular state of vibrational of the lattice is described by the set of occupation numbers $n_{s,\mathbf{q}}$ of all the oscillators.

Starting from a description of the states of this type, it is necessary to evaluate the probability of transition between different states, characterized by different occupation numbers. In doing so, the time-dependent perturbation theory is used and matrix elements of this type must be evaluated:

$$\left|\langle n_f(\mathbf{q},s)\left|H_{anharm}\right|n_i(\mathbf{q},s)\rangle\right|^2$$

The calculations are very laborious and the formalism used is that of the second quantization, but the good news is that it is possible to demonstrate a simple selection rule:

> *if we are considering an an anharmonic term of order k, during a transition induced by this term (passing from the initial state to the final one) only k occupation numbers n_i can change (considering both those that vary in the initial state and those that vary in the final state).*

For example, if we are considering the cubic term it means that three occupation numbers have to change value. Only two physical processes are possible in this case:

- A phonon decays into two phonons: the occupation numbers of the incident phonon and those of the two outgoing phonons (which were initially equal to zero).

- Two phonons join together to form another phonon (the occupation numbers of the incoming phonons go to zero and the occupation number of the outgoing phonon, which was initially zero, changes).

On the other hand, it is not possible to carry out processes in which three phonons disappear or three phonons are born from nothing because they would violate the conservation of energy even though they are compatible with the theorem just stated.

5.7 High temperature thermal conductivity

Previously, based on simple kinetic theory, we obtained the following expression for the conductivity K:

$$K = \frac{1}{3} c_V v_s^2 \tau$$

At high temperatures, the specific heat c_v tends to a constant value, as we are told by the law of Dulong and Petit, which is a consequence of the theorem of energy equipartition. Even v_s, the speed of sound, is constant if we consider only acoustic phonons with low **q**.

It remains to be seen how the relaxation time can depend on temperature. The hypothesis that can be made is that it is inversely proportional to the number of phonons present (i.e. that the frequency τ^{-1} of the scattering events is proportional to the number of diffusion centers present represented by the phonons).

At high temperatures, i.e. when $kT >> \hbar\omega$, the Bose-Einstein factor which represents the number of existing phonons can be developed in series:

$$n(T) \propto \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} \approx \frac{kT}{\hbar\omega} \propto T$$

$\hbar\omega$ is the highest phonon energy, typically in the order of a few meV.

The number of phonons is therefore proportional to temperature and therefore:

$$\tau \propto \frac{1}{T} \rightarrow K \propto \frac{1}{T}$$

In reality, the trend experimentally observed is of the type $1/T^{\alpha}$, with $1 < \alpha < 2$. This behavior can be explained more precisely by more sophisticated theories.

5.8 Conservation laws in scattering between phonons

To understand what happens at low temperatures, it is necessary to distinguish the different scattering mechanisms between phonons, taking into account the conservation laws to which these processes must be subjected.

The calculations (which we do not report) show that in the diffusion processes, due to the interaction between phonons two conservation laws are respected:

- Energy conservation. Even without calculation, it is easy to see what is meant.

- Conservation of vector **q** (plus an arbitrary vector **g**). This means that the difference between the sum of the wave vectors of the phonons involved in the diffusion before and after the scattering must be equal to a vector of the reciprocal lattice. If we multiply all the wave vectors by the reduced constant ħ then we obtain the crystal moments of the phonons, for which, as already observed above, there is a law of conservation "almost" equal to the usual law of conservation of the momentum. The difference is that the crystal momentum before and after diffusion is not necessarily the same but can vary by an amount equal to ħ**g**, being **g** an arbitrary vector of the reciprocal lattice (including the particular case where **g** = 0, for which there is an effective conservation of the crystal momentum). This law is quite general and regards all the quasi-particles scattering. Another example that we have seen is the von Laue condition for diffraction.

The calculations to derive this conservation law are complex and intuition does not help. Let us limit ourselves to say, as already done before, that this behaviour follows from the fact that the momentum is not a constant of the motion because the system does not possess full translational symmetry.

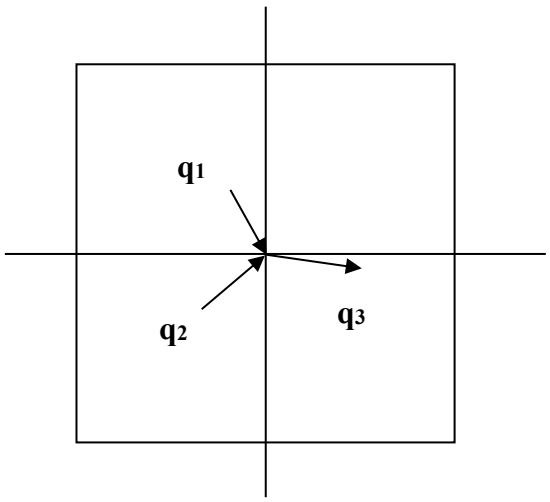
5.9 Normal and Umklapp processes

The law of conservation of the crystal momentum says that, in a collision between phonons, the sum of their wave vectors is conserved but one vector of the reciprocal lattice. In the case of the collision of two phonons that melt to form a third one, one has:

$$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}_3 + \mathbf{g}$$

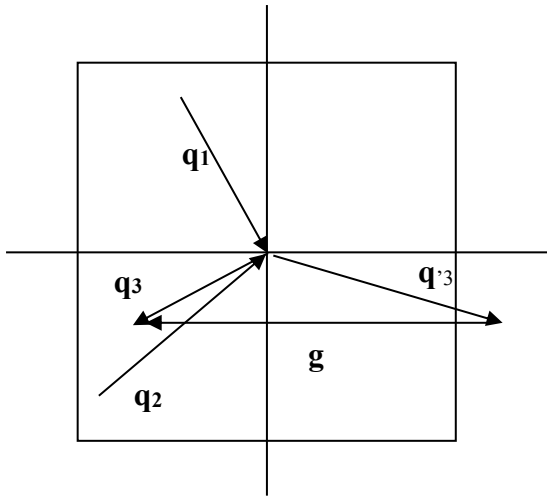
g can be any vector of the reciprocal lattice, including the case **g** = 0.

Normal Process



$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}_3$ ($\mathbf{g} = 0$)

Umklapp process



$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}'_3 - \mathbf{g}$ ($\neq 0$)

In the collisions between phonons that occur in a lattice, taking into account the law of conservation just stated, we can distinguish two cases:

<u>Normal scattering processes</u>

We talk of normal scattering (or N-type scattering) when **g** = 0. In this case the vector **q** is strictly conserved (and therefore also the crystal momentum ħ**q**). The formula is:

$$\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}_3$$

Phenomena of this kind occur if the values of the wavevectors of the phonons involved are small and are all within the first Brillouin zone (1BZ). This is obviously prevalent at low temperatures where the phonons present in the lattice have low values of **q**.

Umklapp processes

When the sum of the wave vectors of the two incident phonons leaves the 1BZ then a strange thing happens related to the fact that the oscillations described by a wave vector **q** are equivalent to oscillations described by a vector **q** + **g**, being **g** any vector of the reciprocal lattice. Let us suppose to add a vector **g** in such a way as to bring back the phonon within the 1BZ as in the figure. The resulting phonon, physically equivalent to the one outside the 1 BZ, is however directed in the opposite direction with respect to the two initial phonons. In this case we speak of processes of the Umklapp type (also called U-type). The term was introduced by Rudolph Peierls and derived from a German word meaning "to fold back".

From a physical point of view, an Umklapp process like the one shown in the figure indicates that two phonons directed to the right add up but that the resulting vector, by virtue of the subtraction of a vector of the reciprocal lattice, moves to the left. This behaviour seems rather strange and in fact does not occur for free particles, where $\hbar\mathbf{q}$ represents the real momentum, because this would violate the law of its conservation. In the case of phonons, on the other hand, it must be remembered that there is a lattice, understood as a single entity, which can act as a source of the missing momentum (or rather, of the crystal momentum).

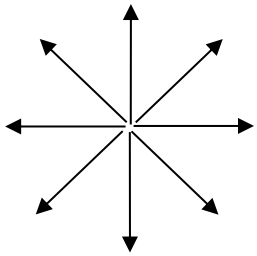
5.10 Thermal equilibrium and the role of Umklapp processes

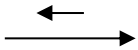
At thermal equilibrium, as it is known, for each phonon **q** there is a phonon −**q** and the sum of all **q** is equal to zero. Now let us look at the figure and think of a material in which there is a temperature gradient and therefore a heat flow occurs.

In the two extreme points of the sample there is thermal equilibrium at two different temperatures, and locally for each excitation **q** there is the corresponding −**q**. At an intermediate point phonons from both directions arrive and, given the gradient of T, from the left arrive phonons that are "warmer", i.e. with values of **q** on average higher than those from the left. This would mean that, in the points along the direction of heat propagation, the condition of thermal equilibrium is lost. However, this does not correspond to physical reality, because experiments tell us that thermal equilibrium is ensured locally and the temperature T(x) varies progressively as we move inside the material. Phenomena that reestablish local equilibrium must therefore occur.

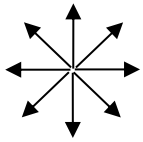
Suppose that only N type processes occur, which do not vary **q**. This would make it impossible to restore the local equilibrium. In practice it would mean that the flow of hot and cold phonons would take place independently and that at all points of the material there would be a situation of imbalance corresponding to a net flow of energy. That is to say, we would have an infinite thermal conductivity (unless of course other types of diffusion from impurities, ...). Since this is not the case, it means that the Umklapp processes act and are responsible for the finite value of thermal conductivity.

Going down with temperature, the thermal conductivity undergoes a sudden increase, with a much greater speed than that provided by the law valid at high T, according to which K is proportional to 1/T. In order to understand this abrupt rise, it is necessary to start from the fact that equilibrium is locally established by the Umklapp processes and therefore we must find a way to evaluate how many phonons are present with **q** large enough to be able to give rise to processes of this type. As a general criterion to evaluate the number of these phonons we decide to use the number of phonons whose wave vector is equal to that of Debye, q_D. It is a fairly large wave vector and therefore it is representative of the phonons capable of giving rise to Umklapp processes. The corresponding energy, the Debye energy, is $\hbar\omega_D = k\Theta_D$.





Hot and cold phonons
$\Sigma \mathbf{q}_i \neq 0$!!



T high, local equilibrium, $\Sigma \mathbf{q}_i = 0$

T low, local equilibrium, $\Sigma \mathbf{q}_i = 0$

For values of $T << \Theta_D$ the number oh phonons is:

$$n(q_D) = \frac{1}{e^{\frac{\hbar\omega}{kT}} - 1} = \frac{1}{e^{\frac{k\Theta_D}{kT}} - 1} = \frac{1}{e^{\frac{\Theta_D}{T}} - 1} \approx e^{-\frac{\Theta_D}{T}}$$

The relaxation time is inversely proportional to the number of phonons and therefore:

$$\tau \propto e^{\frac{\Theta_D}{T}}$$

This explains the exponential increase of the conductivity K, proportional to τ, as temperature decreases.

For even lower values of temperature it becomes important that the specific heat also decreases as T^3 and in fact also the conductivity K, proportional to c_v, begins to decrease for low temperatures. This is not a phonon related phenomenon, but a specific heat related one, and therefore also occurs in amorphous materials.

6. Electron states

6.1 Introduction

Solids are made up of atoms linked to each other by the outermost electrons. This is why all their physical properties depend to a large extent on the electron states, the study of which is a very complex and constantly evolving subject. The properties of solids linked to the electron states can be divided into two categories.

- Those that depend on the ground state of the electrons. These include cohesive energy, crystal structure, phase transitions between different states of aggregation or between different structures, electronic charge density, magnetic properties, effects produced by static electric and magnetic fields, lattice vibrations, ...

- Those that depend on excited states of the electron system. Low energy excitations, i.e. excitations of the order of kT, determine most of the properties of metals: specific electronic heat, electrical resistivity, ... Higher energy excitations determine the optical and transport properties of semiconductors and insulators, the effects of addition or subtraction of electrons as occurs in many spectroscopic investigation techniques, and many others.

We will deal with electron states in solids within the so called band theory based on the independent particles approximation. Better methods have been developed, and research is constantly evolving, but the band theory of independent particle still forms the basis of many of the more sophisticated methods. Besides the fact that it is able to explain many phenomena.

6.2 The formulation of the problem in general terms

The first choice to make is the type of physics to be used to study electrons in solids. In the case of lattice vibrations we have seen that classical mechanics is adequate because the wavelengths of De Broglie associated with ions are much smaller compared to the lattice parameters.

For electrons instead things are different: we know that electrons at the Fermi level in a metal have energies of the order of some eV and their wavelength is of the order of the lattice parameter. So, to study their behavior in structures where they are subject to a potential that varies on this dimensional scale, we must use quantum mechanics by describing the state of the system through wave functions that are obtained by solving the Schrodinger equation.

The state of the system is described by means of a wave function that depends on all the $\mathbf{R}_i$ coordinates of the ions, the $\mathbf{r}_i$ coordinates of the electrons and their spin σ_i:

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{R}_1, \mathbf{R}_2, ..., \sigma_1, \sigma_2, ...)$$

A function of this type is called a many body wavefunction.

To calculate the stationary states let us write the Hamiltonian (not relativistic) of the physical system, which we write in this form:

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 \; - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} \; + \frac{1}{2}\sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \; - \sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 \; + \frac{1}{2}\sum_{I \neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}$$

m_e and M_I are, respectively, the masses of electrons and ions, **r** and **R** their positions, e the absolute value of the charge of the electron and Ze the charge of the ions (with its proper sign). The summations extend to all the particles present. The five terms represent, in order:

- The kinetic energy of the electrons
- The interaction between electrons and ions
- The interaction between electrons
- The kinetic energy of the ions
- The interaction between ions

The potential between electrons and ions is attractive, the potential between electrons and between ions is repulsive and in both cases there is the term ½ not to count twice the same pairs of electrons and ions. This formula is rather general because it includes everything, from the hydrogen atom, to molecules, to solids.

6.3 Approximations

In order to formulate a theory and develop models of the electron states one has to make numerous approximations, often difficult to justify if not on the basis of the argument that, given the complexity of the problem, that is the only possibility to proceed. In the meantime let us start by saying, without worrying about justifying this decision, that from this point on we will ignore the spin of the electrons.

<u>Born-Oppenheimer approximation</u>

The overall wave function that describes the system consisting of ions and electrons must contain only the **r** coordinates of the electrons and the **R** coordinates of the ions since we decided to neglect the spin. Suppose we start with an assigned ion configuration defined by their **R** positions and know the electron eigenstates for that particular configuration. Let us make the hypothesis that the motion of the ions takes place very slowly compared to the electron dynamics and that the effects connected with their movement can be considered perturbatively. In this case the adiabatic theorem of Born and Fock allows us to affirm that if the energy of the state in which the electrons are found is separated from the rest of the spectrum, that is if it is a discrete level, then the system remains in the same state and the induced variations are not such as to produce transitions to other states. Of course, the state in which the electrons are located, along with its energy, undergoes some variations following the modification of the configuration of the ions, without however that an abrupt transition to other states occurs.

From the point of view of the wave function this means that it is possible to factorize it in two terms:

$$\Psi(\mathbf{r}, \mathbf{R}) = \Psi_{\mathbf{R}}(\mathbf{r})\, \Psi(\mathbf{R})$$

The factorisation of the wave function leads to writing two different equations. One is the equation that describes the dynamics of the ions in which the variable is the wave function $\Psi(\mathbf{R})$. In reality, the dynamics of ions, as we have seen when studying lattice oscillations, can be treated with classical

instruments and rarely a completely quantum treatment is necessary. However, it should be noted that the interaction between the ions must still take into account how the electrons are distributed among their quantum states.

The other equation that is obtained by separating the variables is that relative to the electron states described by the function $\Psi_{\mathbf{R}}(\mathbf{r})$, where the ion configuration enters as a parameter. One way to proceed is to calculate the electron states for all possible ion configurations, or at least a selected set of them, and obtain the property of the system by making appropriate averages on the states thus obtained.

If we want to proceed in an even simpler way, we can assume that the ions are immobile in their equilibrium positions. This means assuming that the electron states are not particularly sensitive to the ion displacements produced by thermal agitation. In the following we will rest on this assumption and we will indicate the electronic wave function as $\Psi(\mathbf{r})$ by omitting to explicitly indicate the parameter **R** because we will consider the ions fixed in their equilibrium position.

Effective potential

In the equation describing the electron states, the various terms expressing interaction potentials (electrons-ions and electrons-electrons) are a very difficult problem to deal with. What is done in the first approximation is to suppose that, being the structure periodic, they can be grouped into a single potential, for now not better identified, which has the periodicity of the structure. It is also possible to think of incorporating in this potential the effects due to the oscillatory motion of the ions. In this way we arrive at a Schrodinger equation for the many body wave function in which an average periodic potential $V(\mathbf{r})$ appears:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E\,\Psi(\mathbf{r})$$

Independent particles

The most drastic approximation is to suppose that electrons can be treated as independent particles and therefore their wave function can be factorized into many single particle functions:

$$\Psi = \Psi(\mathbf{r}_1)\Psi(\mathbf{r}_2)...\Psi(\mathbf{r}_N)$$

It is impossible to justify such a choice, except a posteriori and with very sophisticated arguments (the Landau theory of the Fermi liquids). Very roughly we can say that what makes the electrons relatively independent from each other is the principle of exclusion of Pauli, by virtue of which every electron must position itself in a different state from those of the others. We will justify this statement a little better when, talking about conduction, we deal with the collisions between electrons inside a solid.

A first step forward with respect to an independent particle treatment leads to the so called self-consistent field method for the calculation of the electron states (also called Hartree method) in which it is assumed that each electron is subject to an average potential due to all the others. This potential, a priori unknown, in turn depends on the wave functions of the electrons. Then we start by making an assumption on the form for of potential, with which we calculate the wave functions, which are then used to recalculate a more accurate potential, with which we calculate again the wave functions ... and so on, until convergence. If in the calculations we also take into account the fact that, being

electrons, their wavefunctions must be antisymmetrical by exchange of two particles, then we speak of the Hartree-Fock method.

6.4 Electrons in a periodic potential: preliminary considerations

The physics of solids we are dealing with is based on the idea that electrons in a solid can be treated as single particles moving in a periodic potential. From this hypothesis a lot of consequences can be drawn even without really solving any particular equation. This is what we are going to do.

The solutions of the Schrodinger equation in the case of free electrons (constant potential) are plane waves. In a plane wave the probability of finding the electron (which is proportional to the charge density) is constant:

$$|\Psi(\mathbf{r})|^2 = e^{i\mathbf{k}\cdot\mathbf{r}} \cdot e^{-i\mathbf{k}\cdot\mathbf{r}} = \text{costante}$$

If a periodic potential is present the system is made by a cell that repeats itself indefinitely and therefore the probability distribution to find the electron, that is the density of charge, must be the same in all the cells (for the stationary states). Reasoning in one dimension the condition just stated becomes:

$$|\Psi(x + ma)|^2 = |\Psi(x)|^2$$

where a is the lattice parameter and m an arbitrary integer. It must be observed that the considerations we are going to make remain valid even if we consider a true many body functions. A trivial way of meeting this condition is to assume that Ψ is periodic with period a. However, this is a condition too restrictive. A more general condition is that, passing from x to x + ma, the wave function undergoes at most a phase variation:

$$\Psi(x + ma) = e^{ikma}\,\Psi(x)$$

The verification that in this case $|\Psi|^2$ is periodic with period a is trivial. Obviously, the phase shift between adjacent cells must be the same throughout the lattice, for symmetry reasons. The proposed form for the wave function clearly satisfies this condition. It is a situation completely similar to that of the lattice oscillations where, to describe the state of the system, instead of the wave function there are the displacements of the atoms with respect to the positions of equilibrium, $u_n = u_o e^{i(qna - \omega t)}$. Equivalent atoms in adjacent cells oscillate in the same way but with a phase shift e^{iqa}. The parameter k in the formula for electrons has the same physical meaning as the parameter q we used for lattice oscillations. In both cases it is a wave vector but we use two different symbols because in many texts it is done in this way.

To give a more immediate physical meaning to the wave function of the electrons in the periodic potential, whose squared modulus we suppose to be periodic, we proceed as follows. Let us start by rewriting the previous equation in this way:

$$\Psi(x) = e^{-ikma}\,\Psi(x + ma)$$

If we multiply both members by e^{-ikx} we get:

$$e^{-ikx}\Psi(x) = e^{-ik(x+ma)}\Psi(x+ma)$$

Which shows that the expression $e^{-ikx}\Psi(x)$ is periodic with period a. We refer to this periodic function as $u(x) = e^{-ikx}\Psi(x)$. Moving the exponential results in:

$$\Psi(x) = e^{ikx}u(x)$$

This property of the wave functions in a periodic potential constitutes the Bloch theorem, for now stated in the simple one dimensional case and justified in a heuristic way by imposing the periodicity of the square modulus of the wave function. Later the theorem will be demonstrated in two different ways based on much more general arguments. This fundamental theorem, as we have seen, can be formulated in the two different ways:

$$\Psi(x) = e^{ikx}u(x)$$
$$\Psi(x+ma) = e^{ikma}\,\Psi(x)$$

To understand the physical meaning of these functions, called Bloch waves or Bloch states, it is worth considering two extreme cases, one in which the potential $V(\mathbf{r})$ is weak compared to the electron energies and one in which it is strong, in practice a deep potential well.

1) Be V(x) small. In the extreme case in which V is constant the solutions of the Schrodinger equation are plane waves, whose square modulus (i.e. the charge density) is constant. The Sommerfeld model of free electrons in a metal is based on this simplification. With a weak potential we expect the solutions to be still, approximately, plane waves. However, to take into account the presence of the potential, the waves amplitude will be modulated by a periodic term u(x).

2) Let V(x) instead consist of a series of deep potential wells centered on the sites where the atoms are located. The electrons are confined and in each cell their state is described by a wave function completely independent from those of the other cells. This is certainly the case for internal electrons, also called core electrons. Let us suppose that in each of the cells the electron is in a state described by a wave function $\Psi_i(\mathbf{r} - \mathbf{R}_i)$ located on the i-th site identified by the vector $\mathbf{R}_i$. The overall wave function can be expressed as:

$$\Psi = \sum_{\mathbf{R}_i} \Psi_i(\mathbf{r} - \mathbf{R}_i)$$

i.e. the sum of many individual wave functions, each centered on a different lattice point. The relative phases of the different wave functions can be arbitrarily assigned since it is assumed that there is no interaction between the electrons in different sites. Taking advantage of this freedom we can decide that the phase is the same in all sites and in this way the total wave function is periodic and can be made to coincide with the function u(x). In reality, the interaction that leads the atoms to condense forming the solid still has the effect of linking the phases of the different wave functions together. This results in the appearance of the term $e^{i\mathbf{k}\cdot\mathbf{r}}$ which guarantees the same phase relationship between adjacent cells throughout the lattice. Physically, this situation is

that of a solid whose atoms strongly maintain the characteristics of free atoms, for example, solid noble gases.

The Bloch Theorem, enunciated for electrons in a lattice, is generally valid for the solutions of a problem where a periodic potential is present. This is the case of phonons, but also of all the other excitations in a lattice (spin waves, ...). In the case of electrons it is not even necessary to assume the independence of the particles, it also applies to many body wave functions.

The form of the electronic wave functions in a periodic potential, established by the Bloch theorem, is completely equivalent to that of the lattice oscillations and this is not surprising because the Bloch theorem is valid in both cases.

	Lattice oscillations	**Electronic wave functions**
Description of the state of the system.	The $u_{i,n}(x = na, t)$ displacements of the i-th atom of the n-th cell (n.b., variable x is discretized).	$\Psi(x,t)$ (in this case the variable x is continuous).
Form of solutions.	$u_{i,n} = \varepsilon_i e^{ikx} e^{-i\omega t}$	$\Psi(x) = u(x)\, e^{ikx} e^{-i\omega t}$

As we can see the polarization vectors ε_i have the same role of function u(x). The difference is that the former are discretized while the u(x) function is continuous in x.

Translational symmetry allows us to consider only what happens in a cell, the Bloch theorem tells us what happens in nearby cells. When we face the problem, as we have already done for the lattice oscillations, we fix the value of k and we solve the algebraic equations that descend from it. In the case of oscillations these are in finite number and this depends on the fact that the system is discretized. This means that even the possible solutions are finite in number, for example in the case of the linear chain of two different atoms there are two solutions, one acoustical and one optical. In the case of electrons, on the other hand, the variable is continuous, the equations are infinite and so are the solutions.

6.5 The Bloch theorem in the general form

The things previously said with reference to a one-dimensional system have however a very general character and the whole discourse can be reformulated for the case of a three-dimensional lattice. There are two equivalent ways of enunciating the Bloch theorem:

$$1)\ \Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r})$$

$$2)\ \Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\ \Psi(\mathbf{r})$$

The wave functions Ψ are the solutions of the Schrodinger equation in the periodic potential, i.e. they are the eigenstates of the system, and each corresponds to a particular value of energy. **R** are the vectors of the crystal lattice, u(**r**) is periodic for translations of the vectors **R**: $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$.

In order to demonstrate that the two previous formulations are equivalent, it must be demonstrated that the each one of them can be derived from the other.

Let us suppose first of all that 1) is true and let us show that in this way we derive 2). To do this, we calculate 1) by replacing **r** with the translated coordinate **r** + **R**:

$$\Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot(\mathbf{r} + \mathbf{R})}\, u(\mathbf{r} + \mathbf{R})$$

u(**r**) is periodic and therefore:

$$\Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\, e^{i\mathbf{k}\cdot\mathbf{r}}\, u(\mathbf{r})$$

Since we know that, by hypothesis, $e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r}) = \Psi(\mathbf{r})$ then we get 2).

Now let us suppose that 2) is valid and let us proceed as we have already seen in the case in 1D, i.e. multiplying and dividing the term on the right by $e^{i\mathbf{k}\cdot\mathbf{r}}$:

$$\Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\, \Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}e^{i\mathbf{k}\cdot\mathbf{R}}\, \Psi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})}\, \Psi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}$$

The wavefunction Ψ, here calculated at point **r** + **R**, is expressed as the product of two terms. The first one is already calculated at the point **r** + **R**. The second, $\Psi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}$, instead is calculated at point **r**. For the expression to be true, therefore, the value of this second term calculated in **r** must be equal to its value calculated in **r** + **R**. This term must therefore be a periodic function that we indicate with $u(\mathbf{r}) = \Psi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}$, with u(**r**) = u(**r** + **R**). The result is 1), $\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r})$.

6.6 A few consequences of the Bloch theorem

1) Let us consider a Bloch state defined in form 2) and let us see what happens if instead of vector **k** we consider a vector **k'** defined in this way:

$$\mathbf{k}' = \mathbf{k} + \mathbf{g}$$

where **g** is a vector of the reciprocal lattice. Replacing **k'** in 2) we can see that nothing changes:

$$\Psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}'\cdot\mathbf{R}}\, \Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}e^{i\mathbf{g}\cdot\mathbf{R}}\, \Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}\, \Psi(\mathbf{r})$$

Since $e^{i\mathbf{g}\cdot\mathbf{R}} = 1$. This means that, as far as the wave function Ψ(**r**) is concerned, whether we use vector **k** or vector **k** + **g** it is the same thing, which means that, in space **k**, the wave functions are periodic and the periodicity is that of the reciprocal lattice. To highlight this property we write:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \Psi_{\mathbf{k}+\mathbf{g}}(\mathbf{r})$$

Obviously, if the wave functions remain the same after a translation defined by a **g** vector, the same applies to the energy eigenvalues of these functions (remember that we are talking about the eigenstates of the periodic hamiltonian):

$$E(\mathbf{k}) = E(\mathbf{k} + \mathbf{g})$$

The relations that binds together the energy of the electron states with the vector **k** that characterizes them are the dispersion relations of the electrons, completely analogous to those that we have also seen for the phonons and, in general, for all the elementary excitations.

2) Up to this point we have talked about a family of functions $\Psi_{\mathbf{k}}(\mathbf{r})$ that depend on the parameter **k**. Now let us learn something more by replacing the form of the $\Psi_{\mathbf{k}}(\mathbf{r})$ provided by the Bloch theorem in the Schrodinger equation. What is obtained is another equation, also this one at the eigenvalues but different from the Schrodinger equation, where the unknown is u(**r**):

$$\left(\frac{\hbar^2}{2m}\left(\frac{1}{i}\nabla + \mathbf{k}\right)^2 + V(\mathbf{r})\right)u_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}}u_{\mathbf{k}}(\mathbf{r})$$

In this new equation the vector **k** enters as a parameter, so the solutions also depend on **k**, and for this we add a subscript: $u_{\mathbf{k}}(\mathbf{r})$.

We know that $u_{\mathbf{k}}(\mathbf{r})$ must be periodic for **R** translations of the direct lattice, and therefore the solutions of the eigenvalues equation are a translated replica in the real space of the portion of the function defined within the unitary cell. In other words, the problem is reduced to the calculation of the u(**r**) eigenfunctions in a single lattice cell (and this is the enormous simplification that comes with translational symmetry). These eigenfunctions are numerable (we are in a system of finite dimensions as it could be a string that oscillates and therefore the possible ways of oscillation are infinite but numerable) and for this reason we label them, in addition to **k**, also with an index n, called band index:

$$u_{n,\,\mathbf{k}}(\mathbf{r})$$

Each of these eigenfunctions can be used to generate a family of Bloch states:

$$\Psi_{n,\,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, u_{n,\,\mathbf{k}}(\mathbf{r})$$

The wave function is therefore characterized by both the band index n and the value of **k**. For these wave functions, the periodicity properties in the reciprocal space seen above must apply:

$$\Psi_{n,\,\mathbf{k}}(\mathbf{r}) = \Psi_{n,\,\mathbf{k}+\mathbf{g}}(\mathbf{r})$$

$$E_n(\mathbf{k}) = E_n(\mathbf{k} + \mathbf{g})$$

It can be shown that the $E_n(\mathbf{k})$ functions are regular, as well as periodic, functions of **k**. These functions represent the so called band structure of the solid.

3) If we make the calculation, rather complicated, of the expectation value of the electron velocity **v** (defined as the ratio between its momentum **p** and its mass m) in a state $\Psi_{n,\mathbf{k}}$ we get:

$$v_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}E_n(\mathbf{k})$$

This formula shows two very important things:

a) In the stationary states of a periodic potential, electrons have an average velocity different from zero. In other words, these are states in which the electron moves perpetually in a

particular direction and therefore without encountering any resistance. We are therefore in a completely opposite situation with respect to the Drude model, where the electrons move and collide continuously. In an ideal lattice, electrons can move indefinitely while maintaining the characteristic speed of the eigenstate in which they are located.

b) The speed of the electron depends on the band structure since it is the derivative with respect to **k** of the function that describes E as a function of **k**. This shows that the band structure contains the information necessary to understand also the dynamic properties of electrons in a periodic potential.

4) The vector **k** that appears in the Bloch waves has a precise meaning that is not the one it assumes in the case of free electrons where, multiplied times ħ, it represents the electron's momentum (**p** = ħk).

In a periodic potential this is not the case. The general argument is that, since there is no complete translational symmetry, the momentum is not a constant of motion (operators $\hat{p}$ and $\hat{H}$ do not commute) and therefore applying the operator $\hat{p}$ to the eigenfunctions Ψ does not simply produce the eigenfunction multiplied by the corresponding eigenvalue. In other words ħ**k** cannot represent the momentum of an electron in a periodic potential.

ħ**k** is called <u>crystal momentum</u>, just like the analogous vector already defined for phonons. Its physical meaning is that, in processes involving electrons, we continue to have a law of conservation of the crystal momentum completely similar to that already stated for the case of phonons.

6.7 Demonstration of the Bloch theorem in real space

To demonstrate the Bloch theorem in real space, we start by defining <u>the translation operator</u> $\hat{T}_{\mathbf{R}}$ who performs a translation **R** on a state that we indicate, using Dirac's notation, as $|\Psi(\mathbf{r})\rangle$:

$$\hat{T}_{\mathbf{R}}|\Psi(\mathbf{r})\rangle = |\Psi(\mathbf{r}+\mathbf{R})\rangle$$

The translated state is indicated by $|\Psi(\mathbf{r}+\mathbf{R})\rangle$ and **R** is one of the vectors of the Bravais lattice.

Now let us evaluate the following operator expression, remembering that we are considering a <u>periodic Hamiltonian</u> in which the potential V(**r**) is periodic: V(**r**) = V(**r** + **R**):

$$\hat{T}_{\mathbf{R}}\hat{H}|\Psi(\mathbf{r})\rangle = \hat{T}_{\mathbf{R}}\left|\hat{H}(\mathbf{r})\,\Psi(\mathbf{r})\right\rangle = \hat{H}(\mathbf{r}+\mathbf{R})|\Psi(\mathbf{r}+\mathbf{R})\rangle = \hat{H}(\mathbf{r})|\Psi(\mathbf{r}+\mathbf{R})\rangle = \hat{H}\hat{T}_{\mathbf{R}}|\Psi(\mathbf{r})\rangle$$

This shows that the two operators commute between them:

$$\hat{T}_{\mathbf{R}}\hat{H} = \hat{H}\hat{T}_{\mathbf{R}}$$

We know that the translations commute between themselves, that is:

$$\hat{T}_{\mathbf{R}}\hat{T}_{\mathbf{R}'} = \hat{T}_{\mathbf{R}'}\hat{T}_{\mathbf{R}}$$

This means that <u>all</u> possible translations and the Hamiltonian are a set of commuting operators. All these operators have associated an eigenvalues equation:

$$\hat{H}\,|\Psi\rangle = E\,|\Psi\rangle$$
$$\hat{T}_{\mathbf{R}}\,|\Psi\rangle = c(\mathbf{R})\,|\Psi\rangle$$

According to a fundamental theorem of quantum mechanics, if two operators commute they have the same eigenfunctions (obviously not the same eigenvalues). In the case we are dealing with here, this means that by translating a function that describes ad energy eigenstate (i.e. one of the Hamiltonian eigenfunctions) we must obtain the same function multiplied by an eigenvalue c(**R**), and this because it is also an eigenfunction of the translation operator. This is the essence of the Bloch theorem that descends from the commutivity of the operators $\hat{H}$ and $\hat{T}$.

What we have to do now is to derive the eigenvalues c(**R**) and to this end we will rely on some very general considerations.

First of all, we observe that the sum of two translations, obviously made on the same state and to which correspond two different translation operators, is in turn a translation, to which it is possible to associate an additional operator:

$$\hat{T}_{\mathbf{R}}\hat{T}_{\mathbf{R}'} = \hat{T}_{\mathbf{R}+\mathbf{R}'}$$

We now perform two separate translations **R** and **R'**, which correspond to a single translation **R** + **R'**.

$$\hat{T}_{\mathbf{R}+\mathbf{R}'}\,|\Psi\rangle = c(\mathbf{R}+\mathbf{R}')\,|\Psi\rangle$$

Then we carry out the two translations separately, remembering that, in consideration of their commutative nature, the order in which the two translations are carried out is irrelevant. :

$$\hat{T}_{\mathbf{R}'}\left|\hat{T}_{\mathbf{R}}\,\Psi\right\rangle = \hat{T}_{\mathbf{R}'}\,|c(\mathbf{R})\Psi\rangle = c(\mathbf{R}')c(\mathbf{R})\,|\Psi\rangle$$

This proves that the eigenvalues of the sum of two translations multiply each other:

$$c(\mathbf{R}+\mathbf{R}') = c(\mathbf{R})c(\mathbf{R}')$$

The translation operator is not Hermitian. Without rigorously demonstrating this statement, we observe that the hermitian operators are those that correspond to observable physical quantities and that therefore their eigenvalues must be real numbers. In the case of translations it is not clear what the associated observable physical quantity can be and therefore there is no reason to request their hermiticity.

Not being Hermitian operators, their eigenvalues c(**R**) can be, in general, complex numbers. To ensure the normalization of the states obtained by making the translation operator act (these states are still eignestates of the system), the square modulus of the eigenvalues must be unitary:

$$|c(\mathbf{R})|^2 = 1$$

Complex numbers of unit modulus have the form:

$$e^{i2\pi x}$$

Therefore, any complex number that has this shape can be an eigenvalue of the translation operator.

Now let us consider a translation **R** and let us break it down using the fundamental vectors of the direct lattice:

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

Let us start from a translation of a vector $\mathbf{a}_i$ to which corresponds an eigenvalue $c(\mathbf{a}_i)$ that must have, as seen before, a form of this type:

$$c(\mathbf{a}_i) = e^{i2\pi x_i}$$

where x_i can be any number. In principle it could be a complex number but it can be shown that, by imposing to the edges of the wavefunction the boundary conditions of Born-Von Karman, the x_i must be real. We have, of course:

$$c(n_i\mathbf{a}_i) = \left|c(\mathbf{a}_i)\right|^{n_i} = e^{2\pi i x_i n_i}$$

Put in in another way it means that a translation of n steps along the direction of $\mathbf{a}_i$ introduces a phase shift $e^{i2\pi n x_i}$ in the wave function, and therefore:

$$\hat{T}_{n\mathbf{a}_i}\left|\Psi(\mathbf{r})\right\rangle = \left|\Psi(\mathbf{r} + n\mathbf{a}_i)\right\rangle = e^{i2\pi n x_i}\left|\Psi(\mathbf{r})\right\rangle$$

We can write for the translation **R**:

$$c(\mathbf{R}) = c(n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3) = c(n_1\mathbf{a}_1)c(n_2\mathbf{a}_2)c(n_3\mathbf{a}_3) = e^{2\pi i x_1 n_1}e^{2\pi i x_2 n_2}e^{2\pi i x_3 n_3} = e^{2\pi i(x_1 n_1 + x_2 n_2 + x_3 n_3)}$$

Different values of x_i have been chosen for the different directions $\mathbf{a}_i$ because the corresponding phase shifts are independent of each other.

The term in brackets at the exponent allows us to express everything as a scalar product:

$$c(\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot(n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3)}$$

k being a vector whose components are the terms x_i referred to the basis of the reciprocal lattice formed by the vectors $\mathbf{b}_i$:

$$\mathbf{k} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3$$

To convince ourself that this is actually the case, it is sufficient to calculate the product $\mathbf{k}\cdot\mathbf{R} = (x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3)\cdot(n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3)$ by remembering the property of the reciprocal basis:

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{i,j}$$

In conclusion we can write that:

$$\hat{T}_{\mathbf{R}} \left|\Psi(\mathbf{r})\right\rangle = \left|\Psi(\mathbf{r}+\mathbf{R})\right\rangle = c(\mathbf{R})\left|\Psi(\mathbf{r})\right\rangle = e^{i\mathbf{k}\cdot\mathbf{R}} \left|\Psi(\mathbf{r})\right\rangle$$

Which is precisely the Bloch theorem in one of the two equivalent forms set out above.

6.8 Possible values of k

Let us consider the case of a one-dimensional array of points with lattice parameter a and let us suppose that it has a finite size and is made up of N cells. To this structure we apply the boundary conditions of Born-Von Karman, also called cyclical conditions because they are equivalent to closing the chain of N cells on itself:

$$\Psi(x) = \Psi\,(x + Na)$$

In this situation the wave functions are periodic with period L = Na and therefore also the e^{ikx} waves entering the expression of the Bloch states must have a period submultiple of L and this limits the possible values of the vectors k that appear in the Bloch waves. By imposing the condition of Born-Von Karman on the states of Bloch we get:

$$\Psi(x+Na) = e^{ikNa}\Psi(x) = \Psi(x)$$

and so:

$$e^{ikNa} = 1$$

from which:

$$kNa = m*2\pi,\ \ m = 0, \pm1, \pm2, \ldots \quad \rightarrow \quad k = m*\frac{2\pi}{Na}$$

The possible k values are therefore infinite but numerable. If we consider only the 1BZ $\left(-\frac{\pi}{a} < k < \frac{\pi}{a}\right)$ it is easy to verify that inside it there are N possible values of k, separated by a constant $\Delta k = \frac{2\pi}{Na}$ distance. If L = Na is the length of the sample we can write that the separation between the possible values of k as:

$$\Delta k = \frac{2\pi}{L}$$

Moving on to the case in 3D, it is easy to verify that in the first zone there will be N possible values of **k**, where in this case N represents the number of cells in the crystal. We will see later that the way of filling the N available states by the valence electrons determines many of the physical properties of the solid, in particular the fact that it can be a solid or an insulator.

The **k** vectors are uniformly distributed in the reciprocal space. An important quantity is the volume Δk available to each of the vectors **k**. For now we know that in the 1BZ there are N vectors **k**, so each of them corresponds to a volume equal to V_{rec}/N, where V_{rec} is the volume of the 1BZ. We also know (from the study of the reciprocal lattice) that $V_{rec} = 8\pi^3/ V_{dir}$, where V_{dir} is the volume of the cell in the direct lattice. In this way we write:

$$\Delta k = \frac{V_{rec}}{N} = \frac{1}{N}\frac{8\pi^3}{V_{dir}} = \frac{8\pi^3}{V}$$

Where V is the volume of the crystal. This relation, that generalizes in 3D the case seen before in 1D, is the same relation that we had found for the vectors **q** of the phonons.

The fact that in a sample of finite size the set of possible values of **k** is discrete offers the possibility, very convenient, to count them.

6.9 Demonstration of the Bloch theorem in the reciprocal space

Let us consider the Schrodinger equation for an electron described by a wave function $\Psi(\mathbf{r})$ in a periodic potential $V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R})$:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\Psi(\mathbf{r}) = E\,\Psi(\mathbf{r})$$

The periodicity allows to express the potential $V(\mathbf{r})$ as a series in which enter only the waves identified by the vectors **g** of the <u>reciprocal lattice</u>:

$$V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}}\, e^{i\mathbf{g}\cdot\mathbf{r}}$$

The vectors of the reciprocal lattice are infinite, in the real calculations obviously only a finite number is used that we indicate with N_g.

The solutions $\Psi(\mathbf{r})$ of the equation instead, being in general non periodic, must be developed as sums of plane waves by considering all the possible values of **k**, that is the whole <u>reciprocal space</u>. We have seen that in the case of a sample of finite dimensions, the possible **k** constitute a numberable set and this allows us to work with a summation and not with an integral. Obviously, in a real calculation only N possible **k** values will be used:

$$\Psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}}$$

By replacing the two expansions in series in the equation we get:

$$\sum_{\mathbf{k}} \frac{\hbar^2 |\mathbf{k}|^2}{2m} C_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}} + \sum_{\mathbf{k}'} C_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{r}} \sum_{\mathbf{g}} V_{\mathbf{g}}\, e^{i\mathbf{g}\cdot\mathbf{r}} = E \sum_{\mathbf{k}} C_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}}$$

Two of these summations run on all the possible N values of the vectors **k** used in the calculations. The intermediate one instead is a double summation over two indices. The first is the one on the N vectors **k**, which we indicate here with **k'** for a reason that will be clearer later (remember that what counts is not the symbol that is used to indicate the index but the set of values that this can assume, which in this summation is obviously the same as the other two of the above espression). The second index is that on the N_g vectors **g** of the reciprocal lattice adopted in the calculations.

Free electrons

In the case of a null potential, the summation with two indices is not present and this equality is obtained:

$$\sum_{\mathbf{k}} \frac{\hbar^2 |\mathbf{k}|^2}{2m} C_{\mathbf{k}}\ e^{i\mathbf{k}\cdot\mathbf{r}} = E \sum_{\mathbf{k}} C_{\mathbf{k}}\ e^{i\mathbf{k}\cdot\mathbf{r}}$$

which is obviously verified when all the terms referring to the same value of **k** are equal, which leads to writing, for each term:

$$\frac{\hbar^2 |\mathbf{k}|^2}{2m} C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} = E\, C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}$$

Each of these equations already explicitly contains the eigenfunction, which is a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$. The C_k coefficient can assume an arbitrary value but must be the same for all waves so that they are all normalized in the same way (or rather relatively normalized with respect to each other since plane waves cannot be normalized). Simplifying the previous equation by eliminating the common terms we obtain the well-known dispersion relation of free electrons that binds energy to the wave vector:

$$\frac{\hbar^2 |\mathbf{k}|^2}{2m} = E$$

Periodic potential

If a periodic potential V(**r**) is present then in the equation appears the product of the potential with the wave function Ψ(**r**), whose development in series is a summation on two indices, one that identifies the vectors of the reciprocal space, the other the vectors of the reciprocal lattice:

$$\sum_{\mathbf{k}'} C_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{r}} \sum_{\mathbf{g}} V_{\mathbf{g}}\ e^{i\mathbf{g}\cdot\mathbf{r}} = \sum_{\mathbf{k}'}\sum_{\mathbf{g}} C_{\mathbf{k}'} V_{\mathbf{g}}\ e^{i(\mathbf{k}'+\mathbf{g})\cdot\mathbf{r}}$$

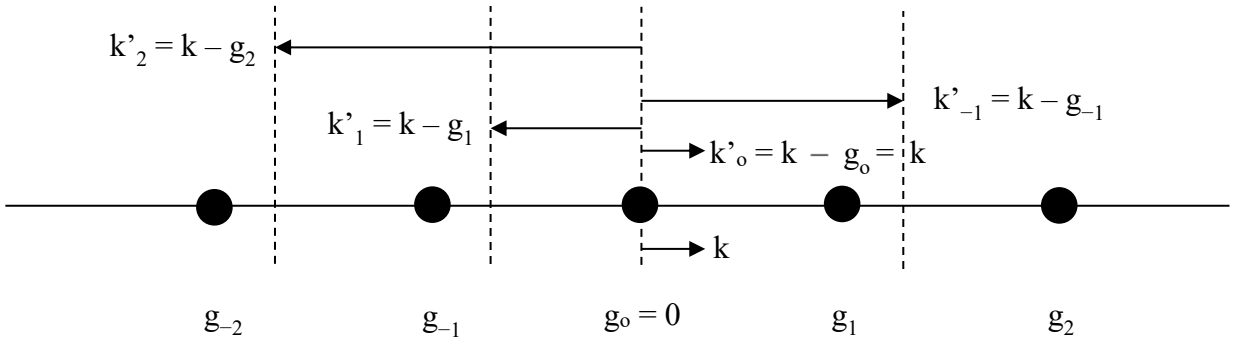
The summation $\sum_{\mathbf{k}'}$ is extended to all wave vectors that enter into the series development of the wave function Ψ(**r**), and that in the summation are indexed using the symbol **k'**. The other summation instead refers to the waves entering the development of the potential in series, indexed with **g**. Overall, therefore, all the N_g*N possible combinations of **k'** and **g** must be considered, each of which gives rise to a wave $e^{i(\mathbf{k}'+\mathbf{g})\cdot\mathbf{r}}$ with wave vector **k'** + **g**.

To proceed, it is better to subdivide all the N_g*N terms of the sum by grouping them according to the value of the wave vector **k'** + **g** (i.e. grouping togheter all those waves where the sum of **k'** and **g** is

the same whereas their individual values are different). We indicate with **k** = **k'** + **g** each of these possible values. Obviously, for a fixed **k**, to each vector **g** of the reciprocal lattice corresponds a vector **k'**, such that the sum of the two returns precisely **k**. This vector **k'** clearly is:

$$\mathbf{k}' = \mathbf{k} - \mathbf{g}$$

To understand how to construct the vectors **k'** that satisfy this condition let us consider the case in one dimension illustrated in the figure:



We can see, indicated with the larger circles, the vectors of the reciprocal lattice g and to distinguish them we use an integer index: $g_o = 0, g_{\pm 1}, g_{\pm 2}...$ The vectors k on the other hand are denser and distributed along the whole continuous line. Let k be one of these vectors and let us suppose that we want to know which are all the other vectors k' which, added to all the possible vectors g, result in k. The thing is simple enough to understand by looking at the figure. What is immediately noticeable is that the vectors k' that satisfy this condition in turn form a replica of the reciprocal lattice, shifted by a quantity k.

Therefore, if we want to group together all the terms of the double summation corresponding to waves with the same wave vector k, we have to take only the terms in which k' belongs to the replica of the reciprocal lattice shifted by k.

In three dimensions this is immediately generalized saying that, to select only the waves with a particular value **k** of the wave vector in the double summation, we must consider all and only the terms in which **k'** = **k** − **g** for all the possible vectors **g** of the reciprocal lattice. Also in this case we are dealing with the set of vectors obtained by shifting of a vector **k** the entire reciprocal lattice.

That being said, the previous summation can be rewritten in this way:

$$\sum_{\mathbf{k}'}\sum_{\mathbf{g}} C_{\mathbf{k}'} V_{\mathbf{g}}\, e^{i(\mathbf{k}'+\mathbf{g})\cdot\mathbf{r}} = \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} V_{\mathbf{g}}$$

The advantage of proceeding in this way is that the previous double summation, on two indices, becomes a summation on a single index for what concerns the waves (**k** and not any more the sum of two terms **k'** + **g**). The second summation on the vectors **g**, on the other hand, is the value of the numerical coefficients of the wave with wave vector **k**:

$$\sum_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} V_{\mathbf{g}}$$

The fact of being able to express the term $V(\mathbf{r})\Psi(\mathbf{r})$ as a summation of many waves $e^{i\mathbf{k}\cdot\mathbf{r}}$ for each of which the coefficient of the series development is known, makes it similar to the other two summations in the equation and therefore allows to collect at common factor the term $e^{i\mathbf{k}\cdot\mathbf{r}}$:

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \left[\left(\frac{\hbar^2 |\mathbf{k}|^2}{2m} - E \right) C_{\mathbf{k}} + \sum_{\mathbf{g}} V_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} \right] = 0$$

This equation is the Schrodinger equation in the reciprocal space. Since it must be true for all the **r** values**,** the term in square brackets must be null for all the values of **k**:

$$\left(\frac{\hbar^2 |\mathbf{k}|^2}{2m} - E \right) C_{\mathbf{k}} + \sum_{\mathbf{g}} V_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} = 0$$

This is only one of the terms of the summation and refers, among all the possible waves, to the one with a particular value **k** of the wave vector, $e^{i\mathbf{k}\cdot\mathbf{r}}$. It is an algebraic equation in which the unknowns are the coefficient $C_{\mathbf{k}}$ together with all its counterparts $C_{\mathbf{k}-\mathbf{g}}$ corresponding to the wave vectors $\mathbf{k}-\mathbf{g}$ of the translated replica of the reciprocal lattice. It is possible to write a similar equation for each of the other coefficients. These equations are different because each one of them is characterized by a different value of the wave vector, whose square modulus explicitly appears in the term in brackets. This system contains N_g equations, i.e. the number of vectors **g** of the reciprocal lattice that we decided to use in the calculations.

6.10 The solutions of the Schrodinger equation with a periodic potential

We are solving the Schrodinger equation with a periodic potential and we have seen that, fixed the value of the vector **k**, we arrive at a system of N_g algebraic equations.This system of equations is the matrix expression of an eigenvalue equation. Each of the possible solutions of the system is a vector whose components are N_g values of the coefficients C. The eigenvalues are the possible values of energy E.

This system of equations admits non trivial solutions when the determinant of its coefficients is equal to zero. Solving the corresponding secular equation, algebraic and of degree N_g in E, it is possible to calculate the N_g eigenvalues of the energy E_n, which will all be referred to a particular value of **k**. We can therefore write:

$$E = E_n(\mathbf{k})$$

The index n is called the <u>band index</u>.

If we consider solutions $\Psi_{n,\mathbf{k}}(\mathbf{r})$ corresponding to the same **k** but with different index n, these are orthogonal to each other, being independent solutions of the system.

If we consider any other wave function characterized by a different value of **k**, i.e. one of the solutions of a different system of equations, the two will be orthogonal because such are the waves that enter into their development in series. Ultimately, the set of all the solutions of all the systems of equations define the eigenstates of the periodic potential, each of which is characterized by two quantum numbers, n and **k**.

A single solution of the system of equations corresponding to a given value of $E_n(\mathbf{k})$ consists of the coefficients $C_{n,\mathbf{k}-\mathbf{g}}$. The corresponding wave function

$$\Psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{i(\mathbf{k}-\mathbf{g})\cdot\mathbf{r}}$$

is an eigenstate of the system with energy eigenvalue $E_n(\mathbf{k})$ that we can rewrite in this way:

$$\Psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{i(\mathbf{k}-\mathbf{g})\cdot\mathbf{r}} = \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{-i\mathbf{g}\cdot\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{-i\mathbf{g}\cdot\mathbf{r}}$$

The summation has the periodicity of the direct lattice, since it contains only waves of the type $e^{-i\mathbf{g}\cdot\mathbf{r}}$. We indicate this periodic function as $u_{n,\mathbf{k}}(\mathbf{r})$ and we write the expression for the wavefunction in this way:

$$\Psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r})$$

This is precisely the form of the wave functions foreseen by the Bloch theorem, which we have therefore demonstrated in an alternative way to that seen above and based on translational symmetry in real space. In addition to this, operating in the reciprocal space has provided us with useful indications on the way to be taken to make real calculations of the electron states.

6.11 The nearly free electron model

We have seen that the eigenfunctions in a periodic potential are in turn periodic in the reciprocal space and therefore also the eigenvalues of the energy are periodic in the reciprocal space:

$$\Psi_{n,\mathbf{k}+\mathbf{g}}(\mathbf{r}) = \Psi_{n,\mathbf{k}}(\mathbf{r})$$
$$E_n(\mathbf{k}+\mathbf{g}) = E_n(\mathbf{k})$$

Let us see what the fact that dispersion relations are periodic in the reciprocal space means in the limiting case of free electrons. In this case the dispersion relation is obtained immediately starting from the classical definition of the kinetic energy:

$$E = \frac{p^2}{2m}$$

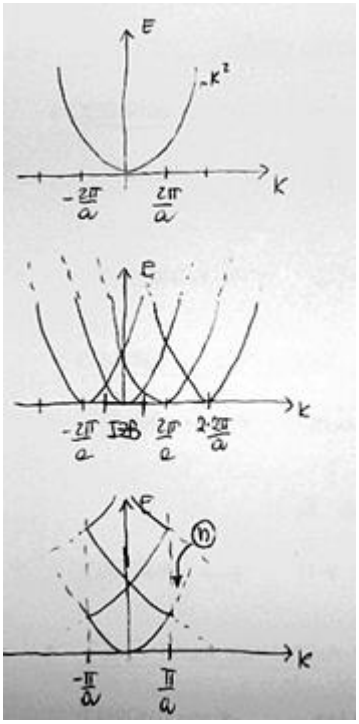
In quantum mechanics the momentum **p** is equal to ħ**k**. We therefore have:

$$E = \frac{\hbar^2 |\mathbf{k}|^2}{2m}$$

In an E-k plot the dispersion relation for free electrons is a parabola defined for all the values of k as shown in the upper part of the figure. Stated in these terms, there is no reciprocal lattice and therefore no periodicity in the reciprocal space. In order to be able to deal with a truly periodic structure, we must introduce a periodicity, for example by assuming that there is a very weak potential periodical, so weak that it does not significantly alter the electron states and the form of the dispersion relation just seen, but that it makes it possible to talk about the periodicity of the structure.

Let us reason in 1D and call a the lattice parameter. What the theory tells us is that, since there is a lattice in real space, the dispersion relation must be replicated indefinitely in the reciprocal space in positions defined of the reciprocal lattice $g_n = (2\pi/a)n$. We will then have many replicas of the parabola of the free electrons, with the axes of all these parabolas separated by a distance $(2\pi/a)$, as in the in intermediate part of the figure.

The presence of replicas of the free electrons parabola means that the dispersion relation is no longer a single valued function but that, for a given value of k, we have several possible values of the energy E. Furthermore, it is easy to see that the complex set of curves expressing the dispersion relations is periodic in k and therefore it is sufficient to consider only what happens within the first Brillouin zone, as in the lower part of the figure.



The model just introduced, in which the electrons move in the presence of a weak periodic potential, is called the Nearly Free Electrons model. For now, the effect of the periodic potential that we have highlighted is the appearance of replicas of the parabolic dispersion relationship. In reality, the periodic potential also produces a distortion of the parabola, particularly at the edge of the Brillouin zone.

To understand the physics of this effect let us consider the one-dimensional case. Electron states are plane waves that propagate along the lattice and for each value of the vector k there are two waves that propagate in the opposite direction (otherwise we would have a net current). As they are plane waves, they will be subject to diffraction as a result of their interaction with the lattice. The Bragg condition is:

$$2a \sin\theta = n\,\lambda$$

a is the lattice parameter, $\lambda = 2\pi/k$ the wavelength and θ the angle between the propagation direction and the lattice planes. Since we are considering a 1D lattice the waves propagate in a direction perpendicular to the planes ($\theta = \pi/2$) separated by a distance a, and therefore the Bragg condition becomes:

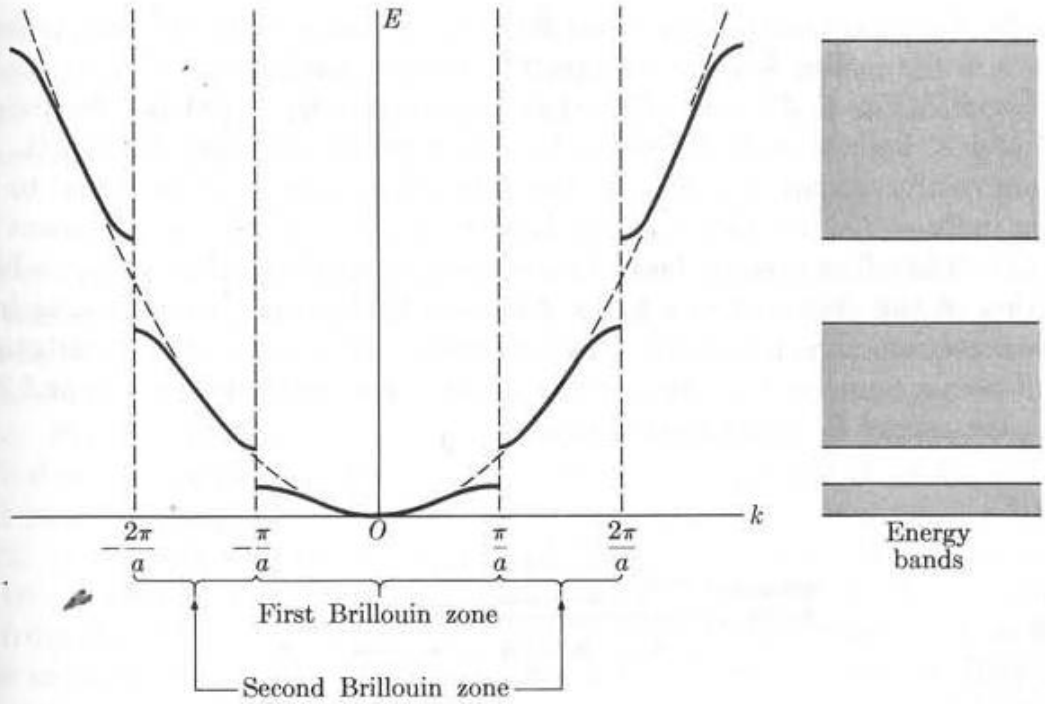


Fig. 6–25. Brillouin zones in a linear lattice.

$$2a = n\,\lambda = n\,\frac{2\pi}{k}$$

The waves that produce diffraction are those that satisfy this condition, which corresponds to a well defined value of k:

$$k = n\frac{\pi}{a}$$

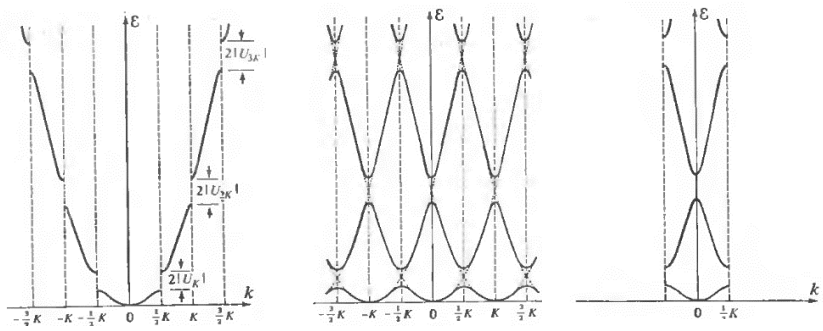
If n = 1 and n = −1 these are waves whose wave vectors are located at the boundary of the Brillouin zone as the vectors of the reciprocal lattice are defined in this way:

$$g = n\frac{2\pi}{a}$$

The diffraction condition then becomes:

$$k = n\frac{g}{2}$$

These are waves identified by the vectors of the reciprocal lattice that are located on the edges of the Brillouin zone. The other waves do not produce diffraction and therefore, since we are assuming that they are nearly free electrons, their characteristics are not modified by the presence of the potential.



A) Distortion of the dispersion relation for free electrons due to periodic potential.
B) The relation must be periodic in reciprocal space, this is the <u>extended-zone</u> scheme.
C) The extended-zone scheme is redundant and can be limited to the first zone, in which case it is referred to as a <u>reduced zone</u> scheme.

On the edge of the zone, instead, each wave produces a diffracted wave with wave vector −k. The state of the system will therefore be a linear combination of waves propagating in the two opposite directions, $e^{\frac{igx}{2}}$ and $e^{-\frac{igx}{2}}$.

Let us consider two possible combinations of these waves:

$$\Psi^{+} = e^{\frac{igx}{2}} + e^{-\frac{igx}{2}} = \cos\left(\frac{\pi}{a}x\right)$$

$$\Psi^{-} = e^{\frac{igx}{2}} - e^{-\frac{igx}{2}} = \sin\left(\frac{\pi}{a}x\right)$$

Each of these two states has a charge density proportional to $|\Psi|^2$. The function $|\Psi^{+}|^2$ has its maxima in correspondence of the points of the direct lattice where the ions are located, the one $|\Psi^{-}|^2$ instead

has its maxima between one ion and the other. Physically it means that in the first case the electrons are attracted by the potential hole of the positive ions, and therefore their energy is lower if compared to the case of the free electrons. In the second case the electrons move away from the positive ions and for this reason the energy of this state is greater than that of the free electrons.

We have therefore come to the conclusion that, for some particular values of the vector k, the same wave vector corresponds to two states with different energy.
By making the calculations it is possible to demonstrate that for values of k lower than these particular values the energy is lower than in the case of free electrons, while just above the values are higher. This, as we can see from the figure, leads to the opening of an energy "gap" in correspondence of these values of the vectors k.

Even in the case just seen, the dispersion relations must be periodic in the reciprocal space, so the distorted curve just seen can be represented both in the extended zone diagram and in the reduced zone diagram, as shown in the figure.

6.12 The tight binding model

Previously we considered an extreme case of a solid, the one constituted by isolated atoms and therefore with an overall wave function given by the incoherent sum of many localized wave functions:

$$\Psi_n^{tot} = \sum_{\mathbf{R}} \Psi_n(\mathbf{r} - \mathbf{R})$$

When the atomic states begin to overlap, an interaction occurs that establishes a phase between adjacent sites. The overall state of the lattice can be expressed in this way:

$$\Psi_{n,\mathbf{k}}^{tot} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \Psi_n(\mathbf{r} - \mathbf{R})$$

Let us now consider the 1D case for simplicity:

$$\Psi = \sum_n e^{ikna} \Phi(x - na)$$

where $(x - na)$ represents the atomic function centered on the n-th site. First we verify that this wave function satisfy the Bloch theorem. We multiply and divide by e^{ikx}:

$$\Psi = e^{ikx} \sum_n e^{-ik(x - na)} \Phi(x - na)$$

If we define:

$$u(x) = \sum_n e^{-ik(x - na)} \Phi(x - na)$$

This is a periodic function that meets the condition $u(x + a) = u(x)$. To understand it, replace x with the value $(x + a)$:

$$u(x+a) = \sum_n e^{-ik[x-(n-1)a]}\Phi[x-(n-1)a]$$

In the case of a lattice made up of many sites, at the infinite limit, the two summations are identical, and this proves that the function u(x) is periodic.

The energy expectation value for a wave function of this type, which we remind cannot be normalized, is given by the expression:

$$E_{ave} = \frac{\int \Psi^* \hat{H} \Psi dx}{\int \Psi^* \Psi dx}$$

$\hat{H}$ is the periodic Hamiltonian operator. Doing the calculations we get:

$$E_{ave} = E_{at} - \alpha - 2\beta \cos ka$$

E_{at} is the binding energy of the atomic state, α and β they are constants that depend on the wave function and potential of the Hamiltonian. This expression tells us that the possible energy values are within this range, or <u>band</u>, of energy:

$$E_{at} - \alpha - 2\beta < E < E_{at} - \alpha + 2\beta$$

The tight binding model allows us to understand the mechanism of formation of bands, as seen in the figure that shows how a single energy level is divided into many levels as the atoms approach forming a band of permitted levels.

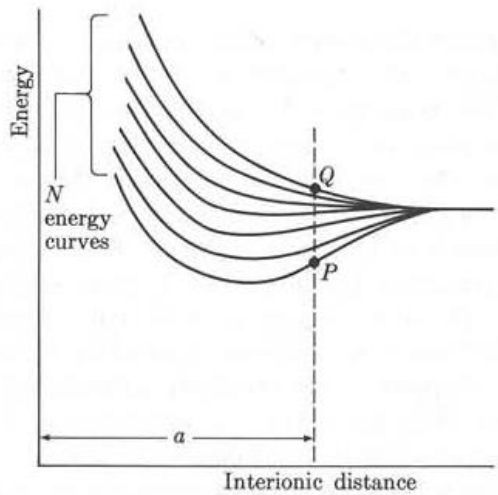


Fig. 6–17. Energy levels in a linear crystal lattice as a function of the interionic distance.

6.13 The occupation of the electron states

So far we have seen what are the possible electron states in the single particle approximation. In a solid these states are occupied by the electrons by respecting the principle of exclusion of Pauli,

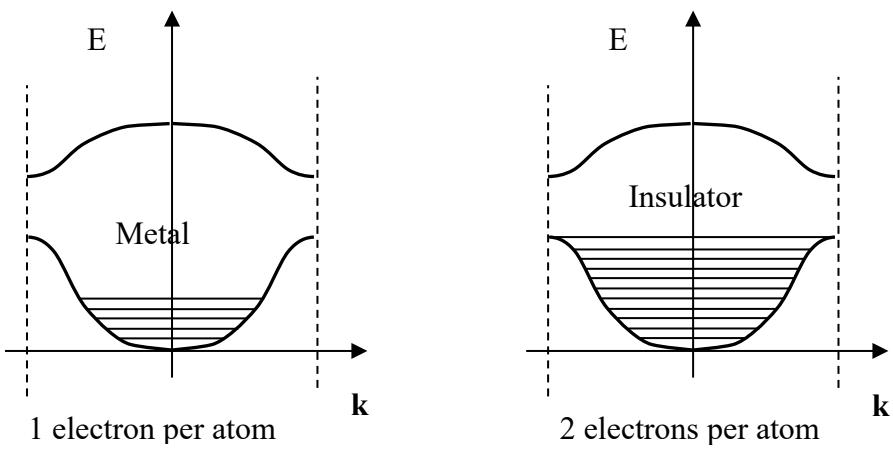
according to which in a given energy level there can be at the most two electrons but with opposite spin. To try to understand what happens, we proceed in order starting from the case in one dimension and then moving on to the more general one.

First, however, it is essential to reiterate what has already been said regarding the number of possible states within the first Brillouin zone.

> *In a finite size system consisting of N elementary cells, within each band there are N electron states that, overall, can accommodate 2N electrons, two for each state with opposite spin.*

<u>The one dimensional case</u>

Let us imagine a solid, unidimensional and with only one atom per unitary cell, whose band structure is the one schematically represented in the figure. In one dimension the band structure is simple because there is only one possible direction for the **k** wave vectors. Moreover, in the figure, it has also been assumed that the dispersion relations are symmetrical by inversion of the vector **k** even if this is not always true, it is only true if the system has symmetry of inversion, if it is not so it is necessary to take into account the spin of the electrons, up and down, and it is shown that $E(\mathbf{k}_{up}) = E(-\mathbf{k}_{down})$.



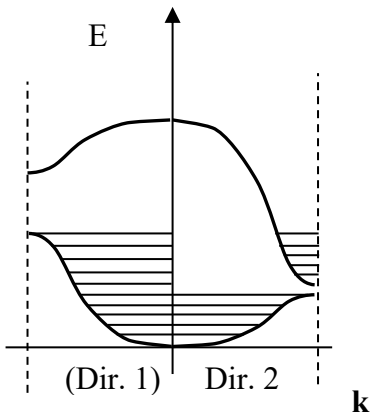
N possible values of **k**, 2N electron states

In the case on the left of the figure we see what happens if each of the atoms contributes to the Bloch states with only one electron, as happens for example in sodium where each atom share with the others only one electron when the metallic bond is formed. The possible states are N, each of these can accommodate two electrons with opposite spin. In this way the lower band is only partially filled, which explains why sodium has a metallic character.

In the case on the right we can see what happens if each cell contributes with two electrons: the band is completely full and the solid is an insulator. In reality this is a situation that occurs quite rarely and to understand the reason we have to take into account the three-dimensional nature of solids.

The three-dimensional case

In two or three dimensions it is possible to move in the reciprocal space along different directions, each of which corresponds in general to a different dispersion relation. In the case of the figure two generic directions (indicated by Dir. 1 and Dir. 2) of a hypothetical solid are indicated, along with their relative dispersion relations. It is noticeable how along the direction indicated as Dir. 2 the upper band goes down, in the energy scale, lower than the top of the lower band along direction Dir. 1.



The electrons, as they are inserted in the available energy levels, are also placed in these states, partially filling the corresponding band. It is for this reason that, even with two electrons per atom, we can still have a metallic character.

In reality things are more complicated and what we see is that:

- All elements with one valence electron are metals. This applies to alkaline elements (Li, Na, ...) and also to noble metals that contain one sp electron in the valence band.

- Divalent elements are metals (Be, Mg, Ca, Sr, Ba).

- Trivalent elements are all metals (Al, Ga, In, Tl).

- Tetravalent elements have complex behavior. Carbon forms diamond that is an insulator but also graphite that is a conductor. Si and Ge are semiconductors, Sn and Pb are metals.

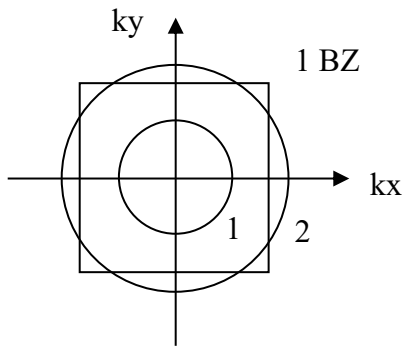
- Pentavalent elements (As, Sb, Bi) crystallize with two atoms per cell so each cell contributes with ten electrons and this gives them a weak metallic character.

6.14 The Fermi surface

In the free electron model the electrons fill states that, in the reciprocal space, fall within a sphere whose radius k_F is called called Fermi wave vector. The surface of this sphere is called Fermi surface.

If there is a periodic potential, the dispersion relations are more complex and this leads to situations that are also very complicated to describe, with Fermi surfaces taking on rather elaborate forms.

Let us consider a solid with a very weak periodic potential, in which electrons can be considered as



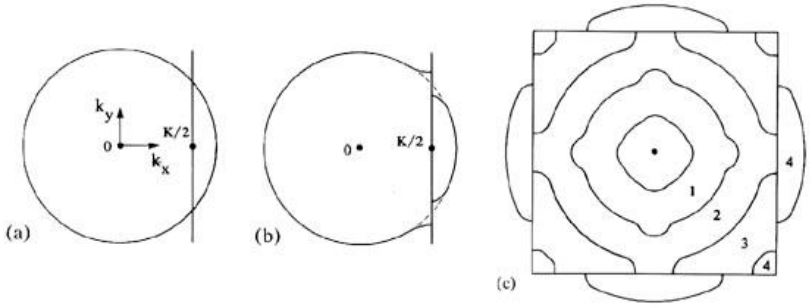
free but in which it is nevertheless possible to define the first Brillouin zone, as for instance in the NFE model.

The figure shows in case in two dimensions. Within the zone there are N states that can accommodate 2N electrons with opposite spin. In the case in which each atom contributes with only one electron the Fermi sphere (in this case the circumference) is contained within the zone, as shown by circle 1. If there are 2N states then the sphere of Fermi intersects the boundaries of the zone.

The figure is actually not correct because when we approach the edge of the zone, the Fermi surface distorts due to an important property of the relations of dispersion $E_n(\mathbf{k})$. The property, which is demonstrated quite easily in the nearly free electrons model and which is almost always generalized even in the case of an arbitrary potentials, is the following:

$$\text{grad } E(\mathbf{k}) = \frac{\hbar^2}{2m}\left(\mathbf{k} - \frac{1}{2}\mathbf{g}\right)$$

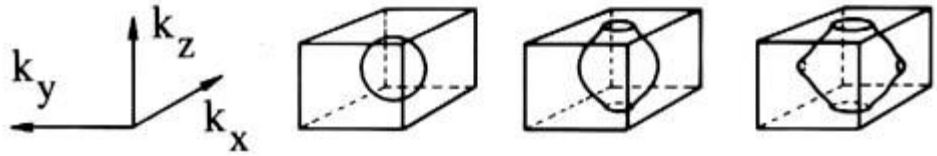
This means that the gradient of E(**k**) must be null when **k** is equal to **g**/2, i.e. when **k** is on a Bragg plane. In the 1D case this means that the dispersion curve is horizontal at the edges of the area, where the gap opens. In more dimensions it means that the Fermi surface must be perpendicular to the Bragg planes. Let us consider, as seen for free electrons, the 2D projection of theFermi surface.



Part (a) of the figure is that of the free electrons. Part (b) shows how the Fermi surface distorts near the area edge to be perpendicular to the Bragg plane. In part (c) curve 1 represents a low electron

concentration, with the radius of the Fermi circle much smaller than the distance from the center to the surface of the Brillouin zone. In the case of curve 2 (more electrons added) there is an increasing perturbation of the Fermi surface for some electron states closer to the edge of the Brillouin zone. Curve 3 represents electrons with a strong perturbation of the highest states, but still the occupied electron states are all within the first zone. The addition of more electrons lead to curve 4, which shows how some electron states have enough energy to cross the boundaries of the Brillouin zone. Note that the lines are constant energy curves and therefore the outermost electrons in curves 2 and 3 have the same energy in all directions and this also applies to those outside the 1BZ of curve 4.

To see what happens in 3D, refer to the latter figure.

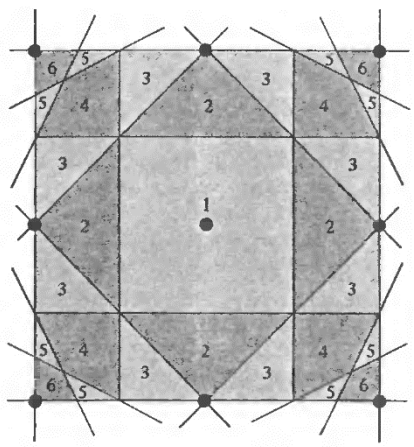


6.15 The successive Brillouin zones

If there are more than two electrons available from each atom, as happens with the elements of the central columns of the periodic table, then the Fermi surface extends well beyond the first Brillouin zone. In this case it is likely that the surface also intersects other Bragg planes, in addition to those that constitute the edges of the first zone. For each of these planes, the condition set out above continues to apply, namely that the Fermi surface is always perpendicular to the planes it intersects. Moreover, at these intersection points, the dispersion relations open up to form energy gaps, which explains the considerable complication of the band structure and of the Fermi surface. To characterize this situation we speak of other zones of Brillouin, subsequent to the first.

The first Brillouin zone is defined as the Wigner-Seitz cell of the reciprocal lattice. An alternative, but equivalent, definition is that the Brillouin zone is the set of points, around a point of the reciprocal lattice, that can be reached without crossing Bragg planes.

Figure 9.7
Illustration of the definition of the Brillouin zones for a two-dimensional square Bravais lattice. The reciprocal lattice is also a square lattice of side *b*. The figure shows all Bragg planes (lines, in two dimensions) that lie within the square of side 2*b* centered on the origin. These Bragg planes divide that square into regions belonging to zones 1 to 6. (Only zones 1, 2, and 3 are entirely contained within the square, however.)



If we cross a Bragg plane, we leave the first Brillouin zone and enter what is called the second zone, which is therefore defined as the set of points of the reciprocal space that can be reached after crossing

a Bragg plane. In general, the n-th Brillouin zone is defined as the set of points that can be reached by crossing (n − 1) Bragg planes.

It is shown that the different zones of Brillouin, whatever the n, have the same volume and therefore are all elementary cells, even if they look quite strange, of the reciprocal lattice. Using this property it is possible to define a procedure to obtain the Fermi surface but the thing is rather complicated.

6.16 The density of states

Also for the electrons in the Bloch states it is possible to define the density of states. To do this, the same methodology adopted for calculating the density of states in the case of the lattice oscillations is adopted. The goal is to find a function g(E), called density of states, such that the product g(E)dE represents the number of electrons that can be accommodated within an energy interval dE centered around the value E of the energy itself. To obtain it, it is advisable to shift attention to the values of the wave vector k, where we have seen that the counting of the states is carried out in a very simple way. To do this we write:

$$g(E)dE = \rho(k)dk$$

The intervals dE and dk are linked together by the dispersion relations, i.e. dE = E(k)dk, while ρ(k) expresses the density of the states in a neighborhood of the k value of the wave vector. It is worth rewriting this relationship in a different way:

$$g(E) = \rho(k)\frac{dk}{dE} = \rho(k)\frac{1}{\frac{dE}{dk}}$$

Let us consider the simple case of free electrons in one dimension. We know that, along the k axis, the points representing the wave vectors are equally spaced at a distance of 2π/L, where L represents the length of the sample. The corresponding density is therefore L/2 π. Since the dispersion relation of the free electrons is symmetrical, the states included within an energy interval dE are those corresponding to the intervals dk centered around the k and −k values of the wave vector. Furthermore, in each of these states two electrons with opposite spin can be found. Therefore, the density ρ(k) is actually equal to four times that of the wave vectors along the k axis and its value is 2L/π.

The dispersion relation is the usual one of free electrons and therefore we can write that:

$$E = \frac{\hbar^2 k^2}{2m} \quad ; \quad \frac{dE}{dk} = \frac{\hbar^2 k}{m}$$

Putting the various pieces together we finally obtain:

$$g(E) = \rho(k)\frac{dk}{dE} = \frac{2L}{\pi}\frac{m}{\hbar^2 k}$$

By inverting the dispersion relation we can express k as a function of E obtaining the desired result:

$$g(E) = \rho(k)\frac{dk}{dE} = \frac{2L}{\pi}\sqrt{\frac{m}{2\hbar^2 E}}$$

With similar considerations it is easy to verify that in two dimensions the density of the electronic states is constant while in three dimensions it is proportional to the square root of the energy. The only thing to remember is that the region of the k space to consider is, in 3D, a sperical shell with radius k and in 2D a circular crown, also with radius k.

At the opposite extreme we find the atomic electrons, each of which continues to maintain the value of the energy it has in the isolated atom and therefore the "band" that follows is a function strongly peaked around the value of the electron binding energy.

Then there are many intermediate situations, such as that of the d-electrons in the transition elements or of the f-electrons in the rare earths, which give rise to quite narrow bands but that somehow interact with those determined by the s and p electrons, more delocalized.

Also in the case of the densities of the electron states we observe the singularities of Van Hove, in analogy with what happens in the lattice oscillations.

6.17 Comparison of electron states and lattice oscillations

Now that the electron states have been discussed, it is appropriate to make a comparison with lattice oscillations.

We will begin with the dynamic equations that describe the two systems. In the case of the lattice oscillations we have operated in a classical context while for the electrons we have adopted the quantum approach:

Lattice oscillations	Electron states
$M_s \ddot{\mathbf{u}}_{s,\mathbf{R}} = -\sum_{s',\mathbf{R'}} G_{s\mathbf{R},s'\mathbf{R'}} \mathbf{u}_{s',\mathbf{R'}}$	$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$

In both cases they are partial differential equations in which the unknowns are, respectively, the atomic displacements $\mathbf{u}_{s',\mathbf{R'}}(t)$ and the wave function $\Psi(\mathbf{r}, t)$. However, these are two different equations in many respects. The first difference consists in the fact that in the case of the atomic displacements the function is defined only in the discrete set of points identified by the indices s and **R** while the wave function is defined in the whole space. If in the equation of the lattice oscillations only the interaction between the first neighbours is considered, as was done in the study of the particular cases of the linear chain of atoms, and the number of different atoms in the cell is stretched to infinity, we obtain the equation that describes the oscillations of the continua. It is the well known D'Alembert wave equation.

The equation for the lattice oscillations is the equivalent of the fundamental law of dynamics and is a second order differential equation with respect both to the time and to the spatial coordinates. There are general solutions and to find a particular solution we have to assign two initial conditions, for example the position and the speed. The equation for the electron states, on the other hand, the Schrodinger equation, is of the second order with respect to the spatial coordinates but of the first with respect to time. Therefore, in the initial conditions only one of them must be imposed, the

position or the speed, in line with the principle of indeterminacy according to which both are not simultaneously known.

In both cases the equations are linear, i.e. for them the sum of two possible solutions is also a valid solution.

A method of solving such equations is the separation of variables, i.e. assuming that the solution is given by the product of one part depending only on the position and one part dependent only on time. As far as the time-dependent part is concerned, it is easy to see that it has to be a complex exponential. In the case of the oscillations of the lattice we write $e^{-i\omega t}$, where it represents the pulsation, linked to the frequency by the relation $\omega = 2\pi f$. In the case of electron states the exponential term is $e^{-i\frac{E}{\hbar}t}$ since $E = \hbar\omega$.

As far as the spatial part is concerned, we know that, by virtue of translational symmetry, the solutions must have the form imposed by the Bloch theorem. Taking into account this consideration, it is possible to hypothesize a form of this type for the solutions of the two equations.

Lattice waves	Bloch waves
$\mathbf{u}_{s,\mathbf{R}}(t) = \boldsymbol{\varepsilon}_s e^{i\mathbf{q}\cdot\mathbf{R}} e^{-i\omega t}$	$\Psi(\mathbf{r},t) = u(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} e^{-i\frac{E}{\hbar}t}$

The exponential terms are the same in the two expressions and the fact that for the lattice oscillations we consider the vector **R** of Bravais lattice while for the electrons we consider the vector **r** depends only on the fact that the first is defined only in discrete points while the second covers the whole space.

The Bloch theorem states that the excitations of a periodic structure are the product of a periodic function for an arbitrary wave. In both cases the periodic function is obviously represented by ε_S for the lattice oscillations and by $u(\mathbf{r})$ for the electrons. Starting from this hypothesis on the form of the solutions it is possible to solve the problem, at least in principle, by replacing them in the respective dynamic equations.

In the case of lattice oscillations, the previous expression for $\mathbf{u}_{s,\mathbf{R}}(t)$ is directly inserted in the dynamic equations.

In the case of electron states, however, we found it more convenient to operate in reciprocal space. We have considered the development in Fourier series of the periodic potential, which concerns only the wave vectors of the reciprocal lattice $V(\mathbf{r}) = \sum_{\mathbf{g}} V_{\mathbf{g}}\, e^{i\mathbf{g}\cdot\mathbf{r}}$ and the development in series of the wave function $\Psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}}$. In this way the unknowns to be determined become the C_k coefficients of the expansion of the wave function which, given the Bloch theorem, must take this form:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{-i\mathbf{g}\cdot\mathbf{r}}$$

Where the periodic function u(**r**) is expressed by summing up all the vectors of the translated reciprocal lattice of **k**.

$$u(\mathbf{r}) = \sum_{\mathbf{g}} C_{n,\mathbf{k}-\mathbf{g}} e^{-i\mathbf{g}\cdot\mathbf{r}}$$

Once inserted the solutions in the equations we have seen that, both for the lattice oscillations and for the electrons, we move from a system of differential equations to a system of algebraic equations. In the case of the lattice oscillations we are dealing with 3n equations, 3 for the different components of the displacement vector of each of the n atoms present in the basis. In the case of electron states, however, the number of equations is infinite because, in principle, all the vectors of the reciprocal lattice must be considered. Of course, in a real calculation the number of equations will necessarily be finished, even if with the current means of calculation it can also assume a very high value.

The two linear algebraic systems that we have obtained in the two cases are the following.

Classical equations of motion:
- Unknowns $\boldsymbol{\varepsilon}_s$;
- 3n equations ;
- Same **q** for all equations

$$M_s\omega^2\boldsymbol{\varepsilon}_s - \sum_{s'} G_{s,s'}(\mathbf{q})\,\boldsymbol{\varepsilon}_{s'} = 0$$

Schrodinger equation in the reciprocal space:
- Unknowns $C_{\mathbf{k}}$;
- ∞ equations;
- Same **k** for all equations.

$$\left(\frac{\hbar^2|\mathbf{k}|^2}{2m} - E\right)C_{\mathbf{k}} + \sum_{\mathbf{g}} V_{\mathbf{g}} C_{\mathbf{k}-\mathbf{g}} = 0$$

In both cases, these are eigenvalues equations that admit non trivial solutions if the secular equation is satisfied. In the case of lattice oscillations this means that each system admits 3n distinct solutions, each characterized by a particular value of ω. In the case of electron states, on the other hand, there are infinite solutions, each with a different energy value E.

As the values of the wave vector change, the solutions, and the respective eigenvalues, change and this leads to the definition of the 3n dispersion relations ω(**q**) in the case of lattice oscillations and to the infinite number of E(**k**) for the electrons. All these dispersion relations are periodic in the reciprocal space with the periodicity of the reciprocal lattice.

7. Transport of electrons

7.1 General considerations

In a solid the electrons are placed in the available electron states which, in the approximation adopted by us, are those of a single particle. Each of these states is uniquely identified by the band index n and the wave vector **k**. The principle of exclusion determines the way in which electrons are arranged in these states: each state $\Psi_{n,\mathbf{k}}$ can host at most two electrons with opposite spin.

In equilibrium conditions, and by assuming that the dispersion relations are symmetrical with respect to **k**, for each state characterized by a wave vector **k** there is another one characterized by a wave vector −**k** with the same value of energy. In this situation of symmetry of the dispersion relations the speed in the state −**k** is equal in modulus, but with opposite direction, with respect to that of the state **k** and therefore the two corresponding currents compensate each other. This applies to all states and therefore, of course, in solid in equilibrium there is no net current flow.

To produce a net flow of charge it is necessary to alter this configuration of equilibrium and to produce an imbalance in the distribution of the electrons in such a way that the average value of their **k** is different from zero. The easiest way is to apply an electric field. This promotes some of the electrons near the Fermi level in more energetic states but, unlike thermal excitation, the presence of the electric field sets a privileged direction with regard to the **k** vectors of the excited states. Overall, therefore, there is an imbalance, in the reciprocal space, which corresponds to a net flow of charge.

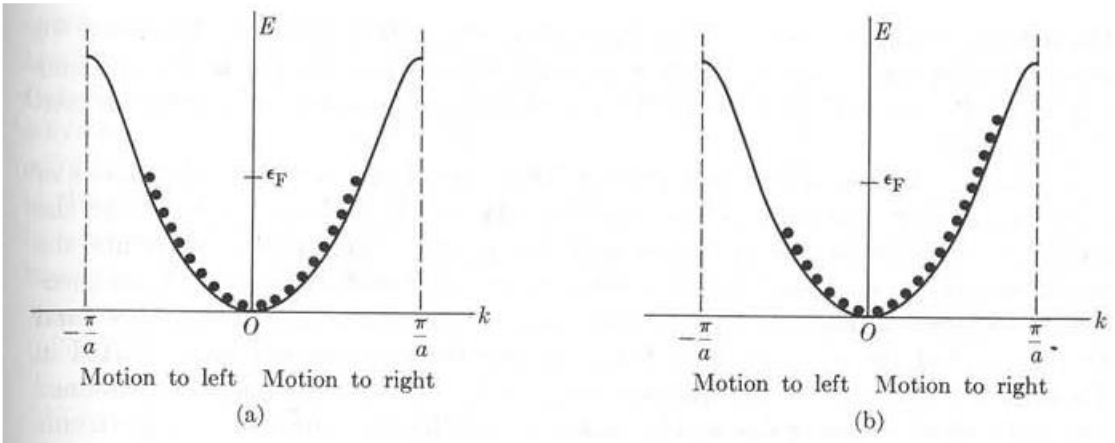


Fig. 6–43. Occupation of energy levels in the first Brillouin zone: (a) no electric field applied, (b) external electric field applied from right to left.

7.2 The semi-classical model

The dynamics of electrons in a solid is a terribly complicated subject but fortunately the essential aspects can be adequately described by a model in which electrons are treated as classical particles that move according to laws formally similar to the classical ones. We talk about a <u>semiclassical model</u> because we do not really have to deal with classical particles but with localized packets built using the initially delocalized states, that is the Bloch waves.

Let us start by summarizing the fundamental characteristics of the dynamics of classical particles, for which it is possible to simultaneously assign position and velocity (or momentum) with arbitrary precision. To these classical particles the law of Newton applies, which can be formulated in two different ways that link force **F**, respectively, to the momentum **p** and to the acceleration **a**:

$$1)\ \mathbf{F} = \frac{d\mathbf{p}}{dt}$$

$$2)\ \mathbf{F} = m\mathbf{a}$$

Electrons have an undulatory nature but it is possible to make them resemble classical particles by building localized packets. The expectation values of the coordinates and the momentum of these packets move according to the classical laws, at least if some conditions are met that we will specify later.

In the case of free electrons or electrons in motion in a region with a slowly variable potential, the verification of all this is quite simple, we will see it in detail below, and the laws of the motion of electron packets coincide exactly with the classical ones. In the case of electrons in motion in the periodic potential of a solid, where instead the potential varies rapidly and therefore the previous condition is lost, it is possible to demonstrate that laws <u>formally</u> equivalent to the classical ones are valid except that instead of the momentum **p** the <u>crystal momentum</u> $\hbar\mathbf{k}$ must be considered and instead of the mass m the <u>effective mass</u> m* must be used, a tensorial quantity linked to the dispersion relations E(**k**).

The semiclassical approach to the motion of electrons in a solid is similar to that used for thermal conduction, which is modeled starting from localized packets of lattice waves that propagate like classical particles.

In the case of lattice oscillations, conduction occurs only in the presence of a temperature gradient, which corresponds to a gradient in the concentration of the phonons. This is also possible with electrons, and the electronic conduction effects linked to the temperature gradient are called thermoelectronic effects (such are the Seebeck effect, the Peltier effect and the Thomson effect). Electrons, however, are charged particles and are therefore also subject to the action of electric and, if they are moving, magnetic fields. In the following we will deal only with the motion of electrons in the presence of an electric field.

7.3 Free electrons or in the presence of a weakly variable potential

Let us consider first the relatively simple case of packets of free electrons or in presence of a slowly variable potential.

<u>Phase and group velocity</u>

A monochromatic wave of wave vector k (wavelength $\lambda = 2\pi/k$) and pulsation ω (frequency $f = \omega/2\pi$) moves with a speed called <u>phase speed</u> which is calculated by multiplying the number of waves passing per unit of time, i.e. the frequency, by the length of each wave:

$$v_{phase} = f\lambda = \frac{2\pi}{k}\frac{\omega}{2\pi} = \frac{\omega}{k}$$

In a wave the frequency ω and the wave vector k are related to each other by the dispersion relation $\omega(k)$. In cases where the relationship between ω and k is linear, the phase velocity is constant for all waves, regardless of the value of k. Otherwise there is an arbitrary dispersion relation $\omega(k)$ and the phase speed of the wave depends on k.

If we construct a packet made up of waves centred on a frequency ω_o, to which corresponds a central wave vector k_o, it is demonstrated that the group velocity with which the packet propagates, defined as the derivative with respect to the time of the value of expectation of the position of the packet, is the derivative of the dispersion relation $\omega(k)$ calculated for $k = k_o$:

$$v_{group} = \left(\frac{d\omega}{dk}\right)_{k = k_o}$$

This formula is valid whatever the dispersion relations is. If it is linear, the group velocity coincides with the phase velocity of all the waves and the packet propagates keeping its shape unchanged, otherwise it widens and disperses.

Free electron waves

Free electrons are plane waves of the type $e^{i(kx-\omega t)}$ and the relations of De Broglie provide the link with the classical quantities:

$$E = \hbar\omega \text{ e } p = \hbar k$$

The dispersion relation $E = E(k)$ for free electrons is obtained starting from the classical definition of the kinetic energy:

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$$

Written in term of the pulsation ω the dispersion relation is:

$$\omega = \frac{E}{\hbar} = \frac{\hbar k^2}{2m}$$

The phase velocity of the electron in this state is therefore:

$$v_{phase} = \frac{\omega}{k} = \frac{\hbar k}{2m}$$

If we consider an electron packet, and taking into account the fact that $\omega = E/\hbar$, the group velocity can be expressed in this way:

$$v_{group} = \left(\frac{d\omega}{dk}\right)_{k = k_o} = \frac{1}{\hbar}\left(\frac{dE}{dk}\right)_{k = k_o}$$

This expression coincides with the one we had previously obtained for the value of the speed of electrons in the Bloch states.

If the dispersion relation is the one of free electrons the group velocity is:

$$v_{group} = \frac{1}{\hbar}\left(\frac{dE}{dk}\right)_{k=k_o} = \frac{\hbar k}{m} = \frac{p}{m} = 2\, v_{phase}$$

We therefore obtain the classical relation between speed and momentum: v = p/m. Furthermore, it must be noted that in this case the phase and group velocity are not the same, as instead happens for electromagnetic waves or acoustic phonons with large wavelength. This happens because the dispersion relation of the free electrons waves is not linear.

Let us now consider a wave packet described by a wave function Ψ(x, t). The average value of the position of the packet, or expectation value $\langle x \rangle$, is calculated in this way:

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \left|\Psi(x,t)\right|^2 dx$$

Since the wavefunction depends on time, the same applies to the expectation value $\langle x \rangle$. The expectation value $\langle p \rangle$ of the momentum of the particle is instead:

$$\langle p \rangle = -i\hbar \int_{-\infty}^{+\infty} \Psi^*(x,t) \frac{\partial}{\partial x} \Psi(x,t) dx$$

If we define the speed of the particle as the temporal derivative of $\langle x \rangle$, with some calculation we obtain the analogous of the classical relation that binds momentum and speed p = mv, the difference is that this time the expectation values are involved:

$$m \frac{d\langle x \rangle}{dt} = \langle p \rangle$$

To achieve this result it is necessary to remember that, in calculating the time derivative of the wave function, according to the time-dependent Schrodinger equation we have $\frac{d\Psi}{dt} = \frac{1}{i\hbar}\hat{H}\Psi$.

<u>Propagation of plane waves packets in the presence of a potential</u>

If the packet is located in a region of space where there is a potential V(x) it is shown that:

$$m \frac{d^2\langle x \rangle}{dt^2} = -\int_{-\infty}^{+\infty} \frac{dV}{dx} \left|\Psi(x,t)\right|^2 dx = \left\langle -\frac{dV(x)}{dx} \right\rangle$$

The two previous formulas constitute the <u>Ehrenfest theorem</u>. Both are always valid, for any wave function and for any potential, they are exact relations derived without making any approximations. The term in parentheses represents the expectation value of the potential derivative calculated using the square modulus of the wave function as a weight function.

The case of a slowly varying potential

If the wave function describes a localized packet, the spatial average of the potential derivative is obviously calculated only in the area where it is different from zero, i.e. where the particle described by the packet is located. If the potential varies slowly with respect to the packet size (i.e. the potential can be considered constant within the packet size) it is no longer necessary to calculate the average of the potential derivative in the whole space, just use the value it assumes at the point $\langle x \rangle$ where the packet is located (let us remember that, in calculating the expectation value, the potential multiply the wave function). But the derivative of the potential, changed of sign, is by definition the force acting in $\langle x \rangle$:

$$m\frac{d^2\langle x \rangle}{dt^2} = -\frac{dV(\langle x \rangle)}{d\langle x \rangle} = F(\langle x \rangle)$$

This is a particular case of the Ehrenfest theorem, but of great importance, because it establishes that the law of propagation of a packet of waves has as a limiting case, for a slowly variable potential, just the classical law of Newton. This is a manifestation of the principle of correspondence according to which quantum and classical mechanics must lead to the same results in cases where the classical description can be considered valid.

7.4 Electrons in a rapidly changing periodic potential

The approach described above can also be used to study the dynamics of electrons in a solid but the situation is much more complicated for at least two good reasons:

- The first is that the system's eigenstates are not plane waves but Bloch waves. This raises the question of how to build localized packets.

- The second is that, as we will see shortly, the periodic potential varies very quickly with respect to the size of the packets, i.e. there is no reason to think that the approximation just seen that leads to write the quantum equivalent of the law of Newton is still valid.

Localized electron packets

First of all, we want to build packets of electrons that can be treated as classic particles, that is to say that have well defined values of both position and momentum. We will proceed in the same way as we did for the phonons.

We know that in the packets the two quantities x and p have uncertainties Δx and Δp linked together by the Heisenberg principle ($\Delta x \Delta p \sim \hbar$) and therefore we must first clarify how we want these packets of localized electrons to be made so that the model we are building can work. Let us start by fixing the orders of magnitude by considering the case of packets built with plane waves, that is considering free electrons.

In the free electron model the electrons that determine the transport properties are those at the Fermi level, which have wave vector k_F. Consider a chain of N atoms separated by a distance a, each of which contributes with one electron to form the metal. We know that the possible values of k are

equally spaced and separated by a distance $\frac{2\pi}{Na}$ and therefore, by assuming to accommodate two electrons for each value of k as foreseen by the exclusion principle, the portion of the Brillouin zone corresponding to these occupied states is $\frac{N}{2}\frac{2\pi}{Na} = \frac{\pi}{a}$. This portion must be centered on the origin and therefore the value of the Fermi wavevector is equal to half of this portion:

$$k_F = \frac{1}{2}\frac{N}{2}\frac{2\pi}{Na} = \frac{\pi}{2a}$$

The condition imposed on the packets is that the uncertainty about k is small compared to k_F:

$$\Delta k << k_F \rightarrow \Delta k << \frac{\pi}{2a}$$

By remembering that $p = \hbar k$, and therefore $\Delta p = \hbar \Delta k$, from the uncertainty relation we obtain:

$$\Delta x = \frac{\hbar}{\Delta p} = \frac{\hbar}{\hbar \Delta k} >> \frac{2a}{\pi}$$

In practice it means that, in order to be able to talk about wave packets with a small k uncertainty compared to k_F, their spatial extension must be much greater compared to the lattice parameter. This is the same condition that we obtained by modelling the thermal conduction with localized phonons. If we are considering electrons in nanometric structures, i.e. with a size comparable to the one of the packets, then this approach is obviously no longer acceptable and we need to formulate an authentically quantum theory of transport.

Bloch wave packets

In a lattice, to construct the packets one must start from the consideration that the eigenstates of the system are not plane waves but Bloch waves. Packets constructed using Bloch waves must have, like those constructed with plane waves, well defined values of position **r** and wave vector **k**.

First of all, it is assumed that electrons always remain in the same band, i.e. transitions between different bands are prohibited. Therefore, the index n indicating the band is constant. To switch from one band to another, tunneling is necessary, as in the case of the Zener effect, and this only happens if the fields present are quite intense.

The construction of packets using Bloch waves involves a rather heavy formalism based on the so called Wannier functions. To understand what it is about we think to express a Bloch state $\Psi_{\mathbf{k},n}$ as a sum of exponentials multiplied by appropriate a_n functions of position:

$$\Psi_{n,\mathbf{k}} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} a_n(\mathbf{r} - \mathbf{R})$$

The a_n functions are the Wannier function. There is one for each site **R** that depends on **r** and n but not on **k**. We can define these functions for each type of band. When the bands are narrow, the functions are similar to the localized atomic states used in the tight binding model. When we are in a situation better described by an nearly free electron model they are obviously something very

different. It should be noted that the function Ψ we are trying to express is a Bloch wave, i.e. is one of the eigenfunctions of the hamiltonian, and not a test function, as is the case of the functions used in the tight binding model, which are not exact eigenfunctions.

If we reverse the previous relationship we can expres these Wannier functions in terms of Bloch states and demonstrate that these functions are orthogonal to each other, in every sense. They are in fact orthogonal to each other those functions with the same index n, but also those with the same band index and centered on different **R** points of the lattice. This orthogonality makes them mathematically useful even if it is difficult to assign them a precise physical meaning.

Since these functions are localized objects, with the property of constituting an orthonormal set, it is possible to write the Schrodinger equation in the representation of the Wannier functions and it is in this way that we proceed to arrive at the simple results of the semiclassical model, which in the following will be enunciated and somehow justified. It is possible to demonstrate them rigorously, but it is a very complex work (see for example chapter 6 of J. Ziman's text, "The principles of the theory of solids").

Propagation of electron packets in a periodic potential

We know that a single electron in a given Bloch state (n, **k**) propagates with a speed, analogous to the phase speed of the plane waves, given by the following expression:

$$v_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}} E_n(\mathbf{k})$$

We build a packet of electrons by superimposing Bloch waves with the same index n but with **k** within a certain interval **k**. Assuming also that everything takes place within a single band, we can assume that for the wave packet we can define the analog of the group speed of the plane waves packets, whose expression will be like the previous one with the difference that the vector **k** must be the central one of the packet.

In the absence of applied external forces, these packets propagate in the crystal without losing energy, as it happens for electrons in Bloch states, even though a rapidly varying periodic potential is present. If these packets were small in size with respect to the lattice parameter, they could be treated as particles subject, instant by instant, to a well-defined force. However, we know that these are packets that extend over many lattice cells and therefore the overall force acting on them due to the periodic potential has an average value of zero. For this reason, we can consider their motion as inertial and neglect the presence of the periodic potential that comes into play indirectly through the previous formula, which makes the speed depend on the band structure.

If we decide to neglect the periodic potential then, in the presence of an external force, we can only consider the potential connected to it, which we can consider as slowly variable with respect to the size of the packet. Since the packets correspond to charged particles we will also be able to evaluate the variation dE of their energy by calculating the work done by the force when the packet undergoes a dx shift using the classic formula:

$$dE = Fdx = Fvdt$$

To take into account the fact that packets actually propagate in a periodic potential, we can express speed using the quantum expression that links the velocity of Bloch waves to the dispersion relations:

$$dE = Fdx = Fvdt = F\left(\frac{1}{\hbar}\frac{dE}{dk}\right)dt$$

If we use the expression $dE = \frac{dE}{dk}dk$ we get:

$$\frac{dE}{dk}dk = F\left(\frac{1}{\hbar}\frac{dE}{dk}\right)dt$$

Hence:

$$F = \hbar\frac{dk}{dt} = \frac{d}{dt}(\hbar k)$$

The important result just obtained is that the temporal derivative of the crystal momentum $\hbar k$ is equal to the acting force. This equation, generalized in 3D and referring to the case in which both an electric field and a magnetic field act, takes this form:

$$\frac{d(\hbar\mathbf{k})}{dt} = \mathbf{F} = -e\left[\mathbf{E} + \mathbf{v}\times\mathbf{B}\right]$$

The previous formula is equivalent to the law of Newton (in the form **F** = d**p**/dt), the difference is that instead of the real momentum p the crystal momentum ħ**k** appears. This difference is due to the fact that the electron packet is propagating in a potential rapidly changing within the packe itself

An important physical consequence of this law is that, by applying a constant force, a linear variation in time of **k** is produced, i.e. the motion of the particle in space **k** is uniform.

7.5 Effective mass

Let us deal now with the law of Newton expressed in the form in which the mass appears, **F** = m**a** = md**v**/dt, and let us see how it can be reformulated in the case of the rapidly variable periodic potential. Let us therefore proceed trying to express the temporal derivative of the speed of the Bloch waves starting, for simplicity, from the unidimensional case:

$$a = \frac{dv}{dt} = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{dE}{dk}\right)$$

Speed v depends on time through k = k(t). It is therefore a question of calculating the derivative of a function of a function:

$$a = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{dE}{dk}\right) = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt}$$

As we have seen previously $\frac{d(\hbar k)}{dt} = F$ and so we can write:

$$a = \frac{1}{\hbar^2}\frac{d^2E}{dk^2}F$$

The classical equation of motion would be:

$$a = \frac{1}{m}F$$

Where m represents the mass of the electron. Formally, also the expression previously obtained can take this form if instead of mass m the <u>effective mass</u> m* thus defined is considered:

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{d^2E(k)}{dk^2}$$

This relation tells us that the details of the motion of a particle in a solid under the action of a force depends on the band structure. The law of motion for the electrons in the periodic potential can be therefore written as:

$$a = \frac{1}{m^*}F = \frac{1}{\hbar^2}\frac{d^2E(k)}{dk^2}F$$

Now let us move on to the case in 3D, obviously more complicated. We know that the electron is moving in the reciprocal space and therefore its vector **k** undergoes, as time goes by, variations that we can express in this way:

$$d\mathbf{k} = d\mathbf{k}_x + d\mathbf{k}_y + d\mathbf{k}_z = \sum_j d\mathbf{k}_j$$

The speed of the electrons depends on **k** and therefore, to variations d**k** are associated variations d**v**. If we consider only one of the components of the speed, the i-th, after a variation d**k** this component will undergo a variation dv_i that we can express by developing $v_i(\mathbf{k} + dk)$ in series:

$$dv_i(\mathbf{k}) = v_i(\mathbf{k} + d\mathbf{k}) - v_i(\mathbf{k}) = v_i(\mathbf{k} + \sum_j d\mathbf{k}_j) - v_i(\mathbf{k}) =$$

$$v_i(\mathbf{k}) + \sum_j \frac{\partial v_i(\mathbf{k})}{\partial k_j} dk_j - v_i(\mathbf{k}) = \sum_j \frac{\partial v_i(\mathbf{k})}{\partial k_j} dk_j$$

The terms dk_j, expressed in scalar notation, represent the modulus of the vcomponents $d\mathbf{k}_i$ of the displacements d**k**. From a mathematical point of view what we have done is to express the variation of the function $v_i(\mathbf{k})$ along a d**k** direction, an operation that mathematically takes the name of <u>directional derivative</u>.

We know that speed can also be expressed in another way, that is, as the gradient, calculated in reciprocal space, of the dispersion relation E(**k**):

$$\mathbf{v}(\mathbf{k}) = \frac{1}{\hbar}\nabla_k E(\mathbf{k}) = \frac{1}{\hbar}\left[\mathbf{u}_x\frac{\partial E(\mathbf{k})}{\partial k_x} + \mathbf{u}_y\frac{\partial E(\mathbf{k})}{\partial k_y} + \mathbf{u}_z\frac{\partial E(\mathbf{k})}{\partial k_z}\right] = \frac{1}{\hbar}\sum_j \mathbf{u}_j\frac{\partial E(\mathbf{k})}{\partial k_j}$$

The three versors $\mathbf{u}_i$ form an orthonormal basis in the reciprocal space and the corresponding components of the velocity vector are $v_i(\mathbf{k}) = \frac{1}{\hbar}\frac{\partial E(\mathbf{k})}{\partial k_i}$. Using this expression we can write the variations of speed v_i in terms of the derivatives of the dispersion relations:

$$dv_i(\mathbf{k}) = \sum_j \frac{\partial v_i(\mathbf{k})}{\partial k_j} dk_j = \frac{1}{\hbar}\sum_j \frac{\partial}{\partial k_j}\left(\frac{\partial E(\mathbf{k})}{\partial k_i}\right) dk_j = \frac{1}{\hbar}\sum_j \left(\frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}\right) dk_j$$

Dividing by dt we obtain the acceleration of the particle:

$$a_i(\mathbf{k}) = \frac{dv_i(\mathbf{k})}{dt} = \frac{1}{\hbar}\sum_j \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}\frac{dk_j}{dt}$$

Remembering that the derivative of the crystal momentum is force, and therefore $\frac{dk_j}{dt} = \frac{1}{\hbar}F_j$, we can write:

$$a_i(\mathbf{k}) = \frac{1}{\hbar^2}\sum_j \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j} F_j$$

The term F_j represents the j-th component of the force acting on the particle. The expression just obtained is formally equivalent to the law of Newton ($a_j = F_j/m$) if we replace the mass with the effective mass m* defined in this way:

$$\left(\frac{1}{m^*}\right)_{i,j} = \frac{1}{\hbar^2}\frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}$$

In this case the effective mass is a tensor that depends on the two indices i and j and that binds together the vectors that represent the force acting on the particle and its acceleration. This tensor can be diagonalized, i.e. three directions can be found in which the acceleration has the same direction as the force. This tensor tells us that the motion of a particle in a solid under the action of a force depends on the band structure.

The dispersion relations $E(\mathbf{k})$ and the effective mass m* are experimentally measurable quantities. As far as the dispersion relations are concerned, the most common experimental technique is ARPES (Angle Resolved Photoelectron Spectroscopy).

For measuring the effective mass there are several experimental methods. The most common is the absorption of microwaves in the presence of a magnetic field B. There is a resonance peak at a frequency:

$$f_c = \frac{eB}{2\pi m^*}$$

7.6 Dynamics of an electron in a band

Let us consider the simple case in which the particle moves along one of the main axes (acceleration and force are parallel). In this case the problem is one dimensional and the effective mass is a scalar:

$$m^* = \frac{\hbar^2}{d^2E/dk^2}$$

For free electrons (E proportional to k^2) the effective mass coincides with the real mass as it is easily verified by considering the dispersion relation of the free electrons. The presence of the lattice produces a distortion of the dispersion relations, as we have seen in the nearly free electron model. See the figure showing this dispersion relationship and the resulting effective mass.

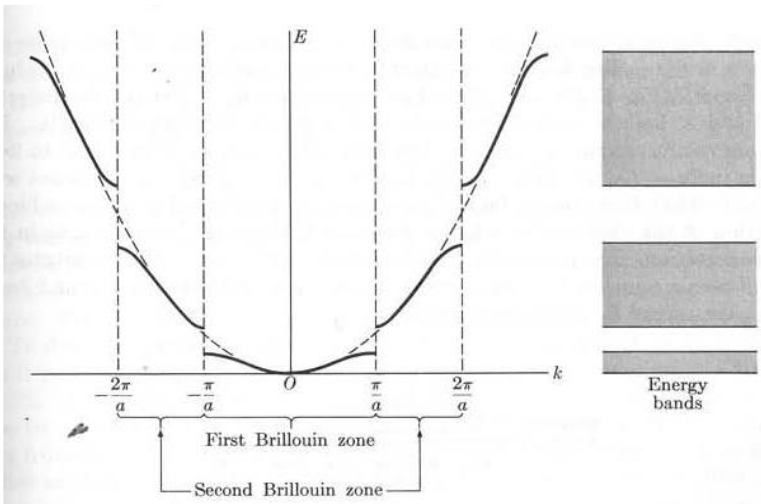


Fig. 6–25. Brillouin zones in a linear lattice.

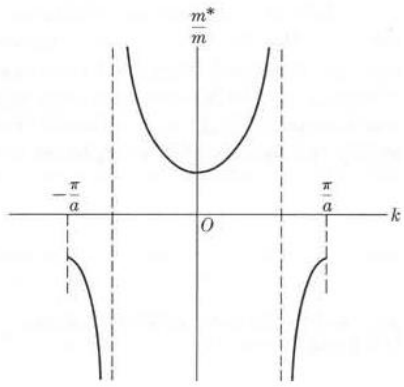


Fig. 6–31. Effective mass as a function of k.

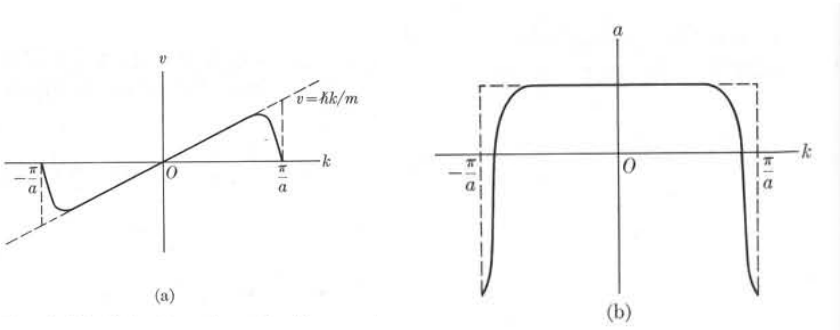
Let us now imagine that in the band there is only one electron, initially in the state k = 0 and that we apply a constant field, which generates a constant force on the electron. The action of the force causes the wave vector **k** to vary, as the semi-classical model tells us:

$$\frac{d(\hbar\mathbf{k})}{dt} = \mathbf{F}$$

This relation shows that with a constant force the variation of **k** is linear in time. As **k** varies, so does the speed since it is proportional to the derivative of the dispersion relation. Initially, the speed increases with the growth of k as for free electrons according to the linear relation $v = \hbar k/m$. However, when the bands begin to be no longer simply parabolic, then the speed begins to increase less quickly and at some point even begins to decrease until it resets to zero at the zone boundary.

This strange behaviour of speed obviously depends on the fact that the electron interacts with the lattice. We can understand all this also considering the trend of the effective mass that increases with the growth of k, and this means that the acceleration produced by the constant force decreases until it completely cancels out when the effective mass assumes an infinite value. After this point the effective mass changes sign becoming negative, and this explains the fact that the speed begins to decrease until it cancels out when we get to the edge of the area.

At the zone edge the Bragg diffraction occurs and the presence of the two waves, the original one and the diffracted one, gives rise to a stationary wave with zero speed. A further increase of k involves leaving the first Brillouin zone and, since the electron remains in the same band, this is equivalent to re-entering the zone from point $k = -\pi/a$. Under the action of the applied force k continues to grow linearly over time, retracing a path similar to the previous one, until it reaches again point $k = 0$, and from here on the cycle repeats itself. The speed and the acceleration, during all this process, have the trend shown in the figure.



All this reasoning can also be done for the other bands and the important result that is reached is that, as a result of a relatively weak applied field, an electron can not leave the band in which it is located. For this to happen, the field must be sufficiently intense and other phenomena must occur, such as the transition to another band due to the tunneling (Zener effect) or after diffusion with other particles from which the electron can absorb sufficient energy (photoabsorption, collisions with electron beams, ...).

The alternating motion of electrons in a band was predicted by Bloch and Zener and is called Bloch oscillations. In practice it is difficult to realize a cycle like the one just described because, in reality, the electrons do not succeed in increasing the value of **k** sufficiently since other scattering occur before reaching the edge of the area (scattering with other electrons, with the lattice oscillations, with defects and impurities, ...).

A numerical estimate allows us to fix the orders of magnitude. If we consider a charged electron $-e$ under the action of an electric field E the acting force is $F = -eE$ and the rate of variation of k is:

$$\frac{dk}{dt} = \frac{F}{\hbar} = -\frac{eE}{\hbar}$$

Assuming that it is $E = 10^5$ V/m, we obtain $dk/dt = 10^{20}\ m^{-1}/s$. k at the edge of the zone has a value $2\pi/a = 10^{11}\ m^{-1}$ and therefore, starting from $k = 0$, 10^{-9} s are needed to reach the edge of the zone. The electrons, however, suffer collisions, which decrease the value of k, and the average time between these collisions is very small, in the order of ps or even less. For this reason, the electrons that start from the bottom of the band cannot reach the zone edge and the oscillations of Bloch are not observed in normal crystals.
Instead, it was possible to observe them in superlattices where the lattice parameter is greater than in crystals and therefore the Brillouin zone is smaller and consequently the time to reach its boundary may be shorter than the time of collision with impurities. The same thing was observed by measuring the motion of ions of cold atoms trapped in a lattice consisting of stationary electromagnetic waves.

7.7 Conduction from a band

What we have been working on so far is the motion of a single electron in a periodic structure. In reality there are many electrons and to understand the properties of solids we have to take into account what they do collectively. Since they do very different things depending on their vector **k**, things are not that simple.

Let us start by expressing the current $d\mathbf{j}_n$ due to the electrons of the n band contained in a volume element d**k** of reciprocal space. Each electron has charge −e and moves with speed **v**(**k**). The density of electron states in the reciprocal space is $1/8\pi^3$, which however must be multiplied by two to take into account the degeneration of spin. We get this expression:

$$d\mathbf{j}_n = -e\mathbf{v}(\mathbf{k})\frac{d\mathbf{k}}{4\pi^3} = \frac{-e}{4\pi^3\hbar}\nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k}$$

First let us consider the case of a fully occupied band. The total current is the integral of the previous expression extended to all the occupied states:

$$\mathbf{j}_n = -\frac{-e}{4\pi^3\hbar}\int_{\text{Occupied}} \nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k}$$

Assuming that the dispersion relations are symmetrical, for each electron state with vector **k** that moves with speed **v**(**k**) there must be one with vector −**k**, and speed −**v**(**k**). Since we are assuming that all these states are occupied, a full band cannot give rise to a net current and therefore $\mathbf{j}_n = 0$ even if a force is acting on the electrons. This is the reason why solids that only have full bands are insulators. (This is true only at T = 0, otherwise there is the possibility that some electrons are promoted in a higher energy band as it happens in semiconductors).

In order to have a net flow of charge the band therefore has to be filled only partially. In the case of a partially occupied band, the expression for the net current takes this form:

$$\mathbf{j}_n = \frac{-e}{4\pi^3\hbar}\int_{\text{Occupied}} \nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k} = \frac{-e}{4\pi^3\hbar}\int_{1ZB} \nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k} - \frac{-e}{4\pi^3\hbar}\int_{\text{Empty}} \nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k} = \frac{+e}{4\pi^3\hbar}\int_{\text{Empty}} \nabla_{\mathbf{k}}E(\mathbf{k})d\mathbf{k}$$

Where the fact that the integral on the 1BZ is null has been exploited.

The obtained result shows how the current in a partially filled band can, at least formally, be considered either as the current of electrons in filled states or as the current of fictitious, positively charged particles in the empty states, the so called holes. It is demonstrated that for these particles is possible to formulate the same dynamical equations that are valid for electrons extending also to the holes the concept of effective mass.

If the band is almost completely full we know that the electrons are placed in the lowest levels and therefore the empty states are the highest. These states are those with large **k** values, which are located on the edge of the area. If the dispersion relation is of the type of the nearly free electrons, it is possible to express it, near the minimum (k = 0) and the maximum (k on the edge of the zone), using a simple parabolic form. The dispersion relation at the bottom of the band is that of the free electrons but, instead of the real mass of the electron, the effective mass is used (the downward arrow means that we are at the bottom of the band):

$$E(k) = \frac{\hbar^2 k^2}{2\left|m^*_{\downarrow}\right|}$$

For the holes in the upper part of the band and close to the edge k_B of the Brillouin zone, on the other hand, the dispersion relation takes this form:

$$E(k) = E_o - \frac{\hbar^2 \left(k - k_B\right)^2}{2\left|m^*_{\uparrow}\right|}$$

$m^*_{\uparrow}$ is the effective mass in the upper part of the dispersion relation, which as we have seen has a negative value (but in the formula above we use the modulus). The effective mass obviously depends on the value of k and therefore is generally different at the bottom and at the top of the band, so it is different for electrons and holes. E_o represents the energy of the highest states in the band.

7.8 The mechanisms of electron diffusion

All we said so far concerns electrons in an ideal crystal that move without encountering any resistance since their state is one of the eigenstates of the Hamiltonian. A situation that is at the antipodes of the Drude model, which foresees collisions with the ions and therefore a mean free path in the order of the lattice parameter. The reality is that in metals electrons suffer collisions with an average path in the order of a of hundreds of Angstroms. To be able to explain these scattering events, something has to be done to break the conditions on which the theory of bands is based.

The first is that the periodicity should cease, and this is the case of:

1. Scattering with static defects (impurities, lattice defects, ...).

2. Scattering with the lattice oscillations.

Then the independence of the particles can be lost and therefore it become possible to have:

3. Electron collisions.

Let us briefly consider these types of scattering. First, however, let us recall the conservation principles relating to collisions between particles in a solid. First of all, must hold the principle of energy conservation, according to which the energy of all particles before the collision must equal that of the particles after the collision.

In empty space we know that the principle of conservation of momentum holds for an isolated system. This principle cannot be valid in a solid because the particles that move and collide in it cannot be considered isolated. However, another rule applies, the conservation of the crystalline moment. According to this rule, the crystalline moment of the particles before and after the collision can differ by an amount $\hbar\mathbf{g}$, where $\mathbf{g}$ is an arbitrary vector of the reciprocal lattice, including the case in which $\mathbf{g} = 0$. The proof of this property is quite complex and will not be reported.

1) Static defects

They can be treated as a perturbation of the Hamiltonian, which depends only on spatial coordinates and not on time. The probability of transition between a state **k** and a state **k'** must therefore be calculated. A static defect, however, is not able to vary the energy of the electrons. It can be demonstrated in a rigorous way but physically it is possible to guess why: a static defect in an infinite lattice has no way of supplying (or absorbing) energy from the electron because it is not clear where it could go to get it. The energy of the electron must therefore remain the same while the crystal momentum can change because the crystal can recoil all together providing the additional momentum $\hbar\mathbf{g}$. In other words, static defects can only produce elastic electron diffusion, as in the collision between two classical particles with very different masses. It is a situation similar to that of diffraction by a periodic potential and also in this case the matrix element is different from zero provided that the moduli of **k** and **k'** are equal and that the energies of the corresponding states are the same.

2) Lattice vibrations

In this case, we can think of ion displacements as time-dependent defects. For this type of scattering the laws of conservation of energy and of the crystal momentum apply (which we limit ourselves to enunciate without demonstrating them, to do so we must use the theory of time-dependent perturbations). The conservation of energy following a collision in which the electron changes its vector **k** into **k'** and produces a phonon with wavevector **q** has this form:

$$E(\mathbf{k}') - E(\mathbf{k}) = \hbar\omega(\mathbf{q})$$

$\hbar\omega(\mathbf{q})$ is obviously the energy of the phonon generated by the collision. The conservation of the crystal momentum, on the other hand, is formulated in this way:

$$\mathbf{k}' - \mathbf{k} = \mathbf{q} + \mathbf{g}$$

The crystal momentum before and after the collision is thus conserved up to an arbitrary vector **g** of the reciprocal lattice.

(3) Electron-electron collisions

The interaction between electrons and the medium in which they propagate is very strong, for the simple reason that the electrostatic interaction of electrons with the other charged particles present in condensed matter is very intense. This is why electrons in material media travel very short distances.

7.9 The electron escape depth and the independent electrons approximation

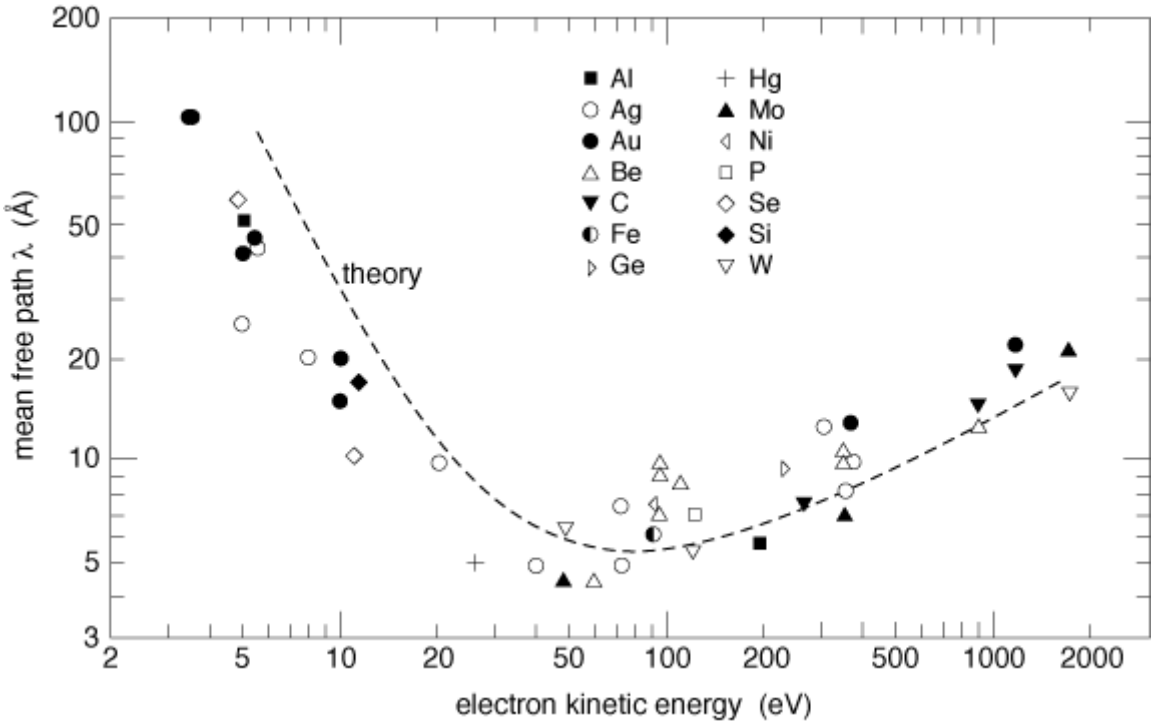
To characterize phenomenologically the motion of electrons in matter we use the Lambert-Beer law in this form:

$$N = N_o\, e^{-x/\lambda}$$

λ is called escape depth or, sometimes, mean free path even though this last term is more frequently used in connection with a single scattering mechanism. In a real situation many different mechanisms act with mean free paths λ_1, λ_2, … The relationship between the escape depth and the individual mean free paths is the following:

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \ldots$$

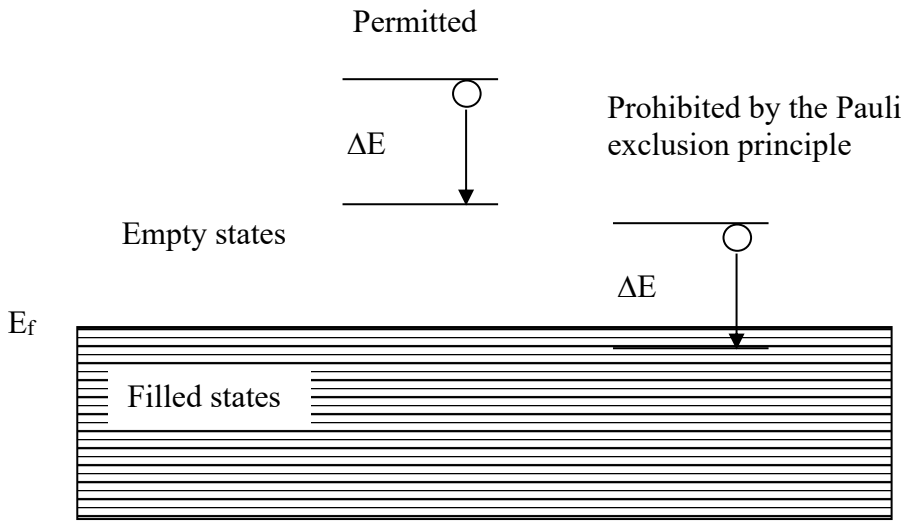
An important physical fact is that λ depends on the energy E of the electrons and the plot that binds the two quantities together is quite independent of the material considered. That is, there is a sort of "universal" curve that binds the energy of the electron to its depth of penetration into the material, as shown in the figure.



The curve has a minimum for electron energies in the order of 100 eV. To understand the shape of this curve, it is better to start with what happens for fairly high energies. In this case the electrons scatter with those of the material in a way very similar to what the alpha particles do with the atomic nuclei, that is to say that the diffusion mechanism is the one of Rutherford. A theoretical analysis of this type of diffusion, whether classical or quantum, leads to the conclusion that the probability of scattering is inversely proportional to the square of the energy of the incident electron. It follows that the corresponding mean free path is larger if the energy of the electron is greater. This allows us to explain the increasing trend of the curve shown in the figure for high energies.

In order to understand the ascent that one observes at low energies, one must resort to quantum considerations. In extreme synthesis, let us consider an electron that moves in a medium suffering energy losses. As known in a solid there are bands of electron states which, in the case of a metal, are full up to the Fermi level. Above the Fermi level there are the empty states in which the electrons injected into the material can stay, as shown in the figure. If an electron is injected into the material, it is placed in one of the empty energy levels and when, following a collision, it loses an amount of energy ΔE, it goes into a lower energy state, as shown in the figure. Obviously, if the lower energy state is in the band of the occupied states, the collision cannot take place because on the basis of the principle of exclusion of Pauli it is not possible to put other electrons in the already occupied states. This impossibility to jump in lower energy filled states is obviously more significant for electrons that are in empty states close to the level of Fermi. This makes it possible to explain, in a qualitative way, why the probability of diffusion decreases as the energy of the electron decreases. Consequently,

due to these purely quantum effects, the depth of penetration increases as the energy of the incident electrons decreases.



All this reasoning is correct if we consider only collisions between electrons, as it is right since this is the major interaction mechanism involved. At very low energies, however, the scattering between the electrons and the lattice vibrations (the phonons) begin to become significant and this is the reason why the depth of penetration does not increase indefinitely as the energy of the electron decreases.

The fact that in a metal low energy collisions between electrons are unlikely is equivalent to saying that electrons, because of the principle of exclusion, interact weakly with each other. This means that, despite the high electron density and the very strong intensity of the Coulomb interaction, the quantum effects lead the electrons to behave as independent particles. This, very roughly, is an argument in favour of the independent particle approximation of the electron states.

7.10 The transport equation

Electrons in a solid can be described in statistical terms by a distribution function f(**r**, **k**, t) defined in this way:

$$f(\mathbf{r}, \mathbf{k}, t)d\mathbf{r}d\mathbf{k}$$

This function represents the number of electrons that, at time t, are in the volume element d**r** and have crystal momentum within d**k**. Obviously, in conditions of equilibrium the f coincides with the Fermi distribution. We call f_o this equilibrium distribution.

If we consider the instant t + dt we know that the coordinates and the crystal momenta of the electrons will vary. The coordinate **r** varies because the electrons are moving at a speed **v**(**k**):

$$\mathbf{r} \rightarrow \mathbf{r} + \mathbf{v}(\mathbf{k})dt$$

while the crystal momentum changes if a force **F** acts:

$$\mathbf{k} \rightarrow \mathbf{k} + \frac{1}{\hbar}\mathbf{F}dt$$

Let us consider the distribution function in the instant t + dt:

$$f\left[\mathbf{r} + \mathbf{v}(\mathbf{k})dt,\ \mathbf{k} + \frac{1}{\hbar}\mathbf{F}dt,\ t + dt\right]$$

The volume element d**r**d**k** remains the same as time passes (Liouville's theorem). If the particles are not created and not destroyed it must be:

$$f\left[\mathbf{r} + \mathbf{v}(\mathbf{k})dt,\ \mathbf{k} + \frac{1}{\hbar}\mathbf{F}dt,\ t + dt\right]d\mathbf{r}d\mathbf{k} - f(\mathbf{r},\mathbf{k},t)d\mathbf{r}d\mathbf{k} = 0$$

In reality, scattering occur and therefore a term must be introduced that describes the variation of f caused by the different scattering mechanisms:

$$f\left[\mathbf{r} + \mathbf{v}(\mathbf{k})dt,\ \mathbf{k} + \frac{1}{\hbar}\mathbf{F}dt,\ t + dt\right]d\mathbf{r}d\mathbf{k} - f(\mathbf{r},\mathbf{k},t)d\mathbf{r}d\mathbf{k} = \left(\frac{\partial f(\mathbf{r},\mathbf{k},t)}{\partial t}\right)_S d\mathbf{r}d\mathbf{k}$$

For small dt the first term can be expanded in series and we arrive at the famous <u>Boltzmann equation</u>:

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}} f + \frac{1}{\hbar}\mathbf{F}\cdot\nabla_{\mathbf{k}} f = \left(\frac{\partial f}{\partial t}\right)_S$$

The first term is the <u>Eulerian derivative</u> of the distribution, i.e. it expresses the variations of f in a fixed point of space. The second term is the one due to diffusion (in this case not in the sense of the collisions but in the statistical sense of the term) caused by the differences of concentration, that is by the gradient of f evaluated in the real space. The third is the drift term linked to the presence of a force.

In order to simplify the discussion let us make the following hypotesys:

- The distribution is stationary, i.e. the Eulerian derivative is null: $\partial f / \partial t = 0$.

- f does not depend on the position ($\nabla_{\mathbf{r}} f = 0$), so there is no diffusion.

Depending on how the scattering term is formulated, the equation may be more or less easy to solve. In general, the term diffusion contains matrix elements that are integral, so the equation is of an integro-differential type. The simplest case is to assume that it is proportional to the difference between the distribution f and its equilibrium value f_o:

$$\left(\frac{\partial f}{\partial t}\right)_S = -\frac{f(\mathbf{k}) - f_o(\mathbf{k})}{\tau(\mathbf{k})}$$

The term τ is the relaxation time. Here we assume that it depends only on **k** and not on **r**, and this is true if the system is spatially homogeneous.

To understand the physical meaning of τ one can imagine to create an out of balance distribution f by applying, for example, an electric field. Let us consider the case of a static electric field E and make some simplifying hypotheses:

In this way, among the other terms of the Boltzmann equation, the only left is the drift term due to the external field. We will assume that the force is due to the presesnce of an alectric field **E** and therefore $\mathbf{F} = -e\mathbf{E}$:

$$-\frac{e}{\hbar}\mathbf{E}\cdot\nabla_{\mathbf{k}} f = -\frac{f(\mathbf{k}) - f_o(\mathbf{k})}{\tau(\mathbf{k})}$$

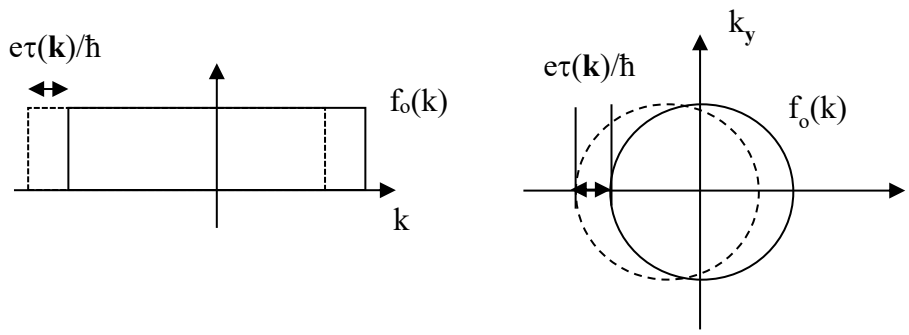
and then, by extracting f(**k**):

$$f(\mathbf{k}) = f_o(\mathbf{k}) + \frac{e}{\hbar}\tau(\mathbf{k})\mathbf{E}\cdot\nabla_{\mathbf{k}} f$$

In the case of small values of field E (Ohm's law) we can assume that $\nabla_{\mathbf{k}} f \cong \nabla_{\mathbf{k}} f_o$ and therefore the two terms on the right represent the beginning of the expansion in series of f_o calculated in the reciprocal space at point $\mathbf{k} + \frac{e}{\hbar}\tau(\mathbf{k})\mathbf{E}$, and therefore:

$$f(\mathbf{k}) \cong f_o\left[\mathbf{k} + \frac{e}{\hbar}\tau(\mathbf{k})\mathbf{E}\right]$$

The meaning of this expression is that the distribution that is determined under the action of the field E is equal to the equilibrium distribution translated however, in the reciprocal space, of a quantity $\delta\mathbf{k} = e\tau(\mathbf{k})\mathbf{E}/\hbar$. The figure shows the case of an nearly free electron gas.



On the left in the figure with the solid line it is shown the Fermi distribution of a free electron gas. Instead the dashed line represents the out of equilibrium distribution created by the applied field. The out of equilibrium distribution is the one at equilibrium but shifted in the reciprocal space,

On the right is shown the case of a Fermi gas in three dimensions where the Fermi sphere undergoes a translation in the reciprocal space.

If the electric field is turned off, the distribution will decay exponentially to the equilibrium one, with a time constant τ.

7.11 Conduction in a metal

If we want to calculate the link between the electric field **E** and the current density **j** in a metal we introduce the concept of conductivity σ or its inverse, the resistivity ρ (in general these are tensors, here we suppose they are scalar):

$$\mathbf{j} = \sigma\mathbf{E} \quad \text{or} \quad \mathbf{E} = \rho\mathbf{j}$$

If we evaluate the current density we can express it as an integral over the 1BZ. In the case of a metal, however, under E_F the terms $v(\mathbf{k})$ and $v(-\mathbf{k})$ cancel each other and only an integral extended to a shell around the Fermi surface is left. The final result is the same as with Drude model:

$$\rho = \frac{m^*}{e^2\tau(E_F)n}$$

m* is the effective mass, n the electron concentration in the solid, e their charge and τ the relaxation time of the electrons at the Fermi level.

Drude's model however is wrong because all electrons are assumed to contribute to conduction. In reality, we know that the only ones to do so are those near the Fermi surface. So the analogy is only formal and in any case in the Drude model the relaxation time τ was not precisely defined while now we know that it is the relaxation time of the electrons at Fermi level. In calculating the conductivity in semiconductors it must be kept in mind that the concentration of carriers n is a function of temperature and therefore it is necessary to carry out appropriate averages on the values of the relaxation time τ which is a function of **k**.

If multiple mechanisms operate within the same sample, the inverse of the time between two successive scattering events is equal to the sum of the inverses of those connected with each of these mechanisms. The most relevant mechanisms in metals are those connected with diffusion by defects and by phonons:

$$\frac{1}{\tau} = \frac{1}{\tau_{def}} + \frac{1}{\tau_{ph}}$$

In a metal the dependence of ρ on temperature is the same as τ, since both n, the concentration of conduction electrons in a metal, and m* do not depend on T.

For defects, the probability of diffusion depends on the speed of the electrons at the Fermi level and on the cross section, both independent of T.

Two cases must be distinguished for phonons. One case is that of the high values of T for which the approximation already made for the thermal conductivity is valid, where the probability of diffusion is assumed to be proportional to the number of phonons present. In this case τ_{ph} results to be inversely proportional to T. In this way, the resistivity ρ, which is proportional to the inverse of $\tau_{ph,}$ is proportional to T.

At low T a theory due to Gruneisen shows that ρ is proportional to T^5. In some alloys containing ferromagnetic impurities, a weak minimum is observed in the resistivity at low temperatures. This is called the Kondo effect.

Ultimately, since ρ is inversely proportional to τ, the total resistivity is given by the sum of two terms linked to the defects and to the phonons:

$$\rho = \rho_{def} + \rho_{ph}$$

This is the so called Matthiesen rule (1864).

The law of Wiedemann-Franz (1853) states that in metals the ratio between the thermal conductivity due to electrons, which in a metal is the dominant mechanism, and the electric conductivity is proportional to T.

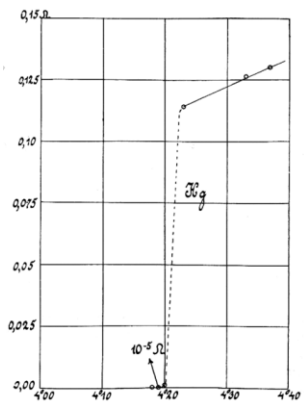
8. Superconductivity

Superconductors are materials in which very strange things occur and the most famous, the cancellation of resistivity, is only one of them. Equally relevant and strange are those related to their magnetic behavior. So strange that it leads us to say that a superconductor is a quantum physical system on a macroscopic scale. Similar and equally strange behaviors are also observed in other phenomena, in many ways conceptually similar, such as superfluidity or Bose-Einstein condensation.

8.1 The cancellation of resistivity

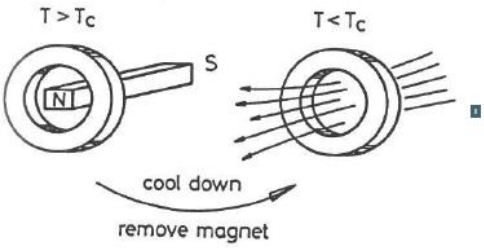
The resistivity in a normal metal shows, at low T values, a trend proportional to T^5 (the Bloch-Gruneisen formula) in addition to a residual value that can also be observed at T = 0 due to diffusion with impurities and defects. In some materials containing magnetic impurities there is also a minimum known as the Kondo effect.

In superconductors the resistivity, instead of continuing to progressively decrease as the material cools down, is abruptly cancelled out below a well-defined temperature, characteristic of the material, called critical temperature (T_c). The figure, taken from the original work, shows the first experimental observation of the phenomenon (Kammerling Onnes, 1911) made by measuring the resistivity of mercury as a function of temperature. As we can see, below a temperature of about 4.2°K the value of resistivity suddenly collapses.



It is important to underline the fact that, below T_c, the resistivity does not limit itself to diminish, even if by a lot, but cancels out completely, as revealed by experiments carried out on rings made of superconducting material. If we induce currents in these rings by varying the magnetic flux inside them, for example by removing a magnet inserted in the central hole of the ring thus giving rise to an electromotive force, these currents never decay. Within the experimental errors, it can be said that the time constant of their possible decay cannot be less than 100.000 years. Superconducting magnets are based on the principle that a continuous current circulates without ever stopping.

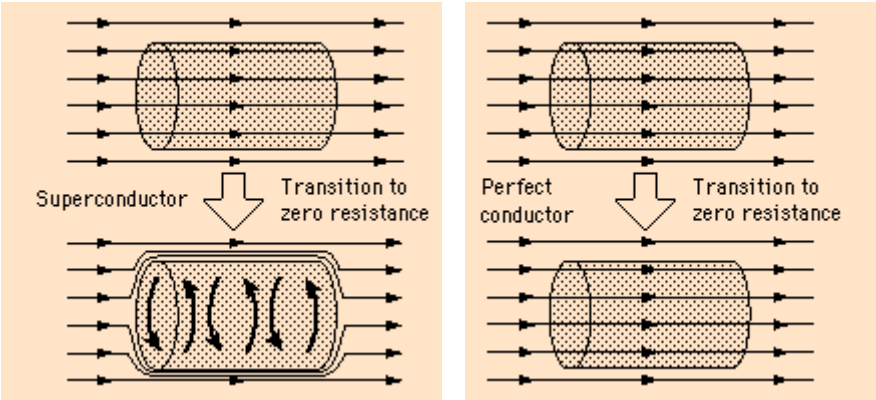
The superconducting chemical elements are about thirty. Some of them (semiconductors like Si) become superconductors at high pressure. So are many compounds, in crystalline, polycrystalline and amorphous form, which means that in the superconducting state it is of no importance the presence



of a perfect lattice. Therefore, superconductivity has nothing to do with what we have seen while dealing with the waves of Bloch, namely that an electron in a perfect lattice has an average speed different from zero and therefore can continue to move indefinitely.

8.2 The Meissner-Ochsenfeld effect and perfect diamagnetism

The Meissner-Ochsenfeld effect (1933) occurs in superconductors. This effect, shown in the left part of the figure, can be observed by cooling a superconductor in the presence of a magnetic field. What we observe is that, when we descend below T_c, the magnetic flux outside the material suddenly increases and this is explained by the fact that the flow lines are expelled from the superconductor, thus cancelling the field inside it and making it increase outside.



To explain this effect, the first temptation is to connect it directly to the cancellation of resistivity below the critical temperature. Unfortunately, however, this simple explanation is not adequate and now we will see that the cancellation of resistivity can only partially explain the shielding of an external magnetic field, but not the Meissner-Ochsenfeld effect which instead consists in a real expulsion of the field already present in the material.

To understand how an external magnetic field can be shielded, let us imagine to apply a magnetic field to a normal conductor and that this field, initially null, grows linearly over time. If we consider an arbitrary closed line inside the material, through it we have a variable magnetic flux that, according to the Farady-Lenz law, generates an electric field that in turn make currents circulate. These currents in turn generate a magnetic field whose flux tends to cancel out the variations of the flux due to the one applied externally.

This phenomenon is well known and occurs in all materials. Induced currents are called eddy currents and are associated with energy dissipation. The magnetic field that they create tends to oppose

variations in the applied one, however, its intensity is generally very small compared to it and this is because of the dissipative phenomena that occur in the material.

Let us suppose instead that our material is a perfect conductor, that is, that the resistivity is null. In this case, since the currents can flow freely regardless of their intensity, the field generated by them is able to completely shield the external field.

Made optimistic by the fact that in a perfect conductor it is possible to completely shield an applied magnetic field, let us try to use these arguments to explain also the Meissner-Ochsenfeld effect. The experimental situation is illustrated in the right part of the previous figure. We take a cylinder of conductive material and start from room temperature, where the resistivity is different from zero. The cylinder is immersed in a constant magnetic field that completely penetrates the material (which therefore must not be ferromagnetic). Let us suppose that, once a certain temperature is reached, the material passes in a state where $\rho = 0$. Since the magnetic flux linked with any closed line inside the material does not change, it follows that nothing must happen, in particular no currents of any kind can arise. Therefore, under these experimental conditions the magnetic field remains within the perfect conductor, unlike in superconductors, where the field is completely ejected once it has dropped below T_c.

Moreover if, once the state in which $\rho = 0$ is reached, the magnetic field applied to the perfect conductor is suddenly shut off, according to the Faraday law currents should be generated to preserve the magnetic flux inside the material. These currents would then continue to flow indefinitely and the material would become a permanent magnet. In superconductors, on the other hand, the currents simply disappear when the external field is removed.

In conclusion, we can affirm that the magnetic behaviour of superconductors is a further aspect of their very peculiar nature that cannot be traced back simply to the cancellation of resistivity. To summarize we can say that a superconductor is both a perfect conductor and a perfect diamagnet.

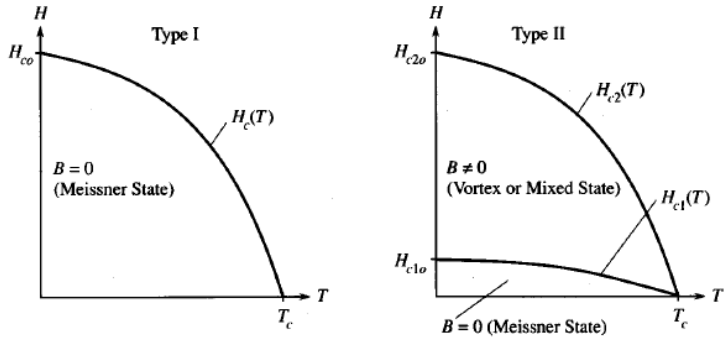
To understand the type of physics to be used to explain superconductivity one can do the following reasoning. A theorem of statistical physics, the Bohr-van Leeuween theorm, states that a statistical system of classical particles in equilibrium cannot possess a net magnetization, even in the presence of an external field. The essential ingredient in the demonstration of the theorem, quite simple, is the fact that a magnetic field does not perform work and therefore can not change the energy of the system. The most important consequence of the theorem is that ferro, para and diamagnetic phenomena can only be explained using quantum mechanics. This consideration also extends to the magnetic behaviour of superconducting materials, which can therefore be considered as the macroscopic manifestation of a quantum behaviour.

8.3 The effects of magnetic fields

Another aspect of the magnetic behavior of superconductors is that a field strong enough destroys the superconducting state and this is a serious problem for applications, particularly for the manufacture of superconducting magnets. Even a current greater than a certain value, called critical current, destroys the superconducting state, and this too is a major technological problem.

The behaviour with respect to the applied magnetic fields leads to a distinction between two types of superconductor. In the normal superconductors called type I, we observe the existence of an H_{co} field, called critical field, whose value depends on T, as shown in the figure, below which the material remains superconducting and there is no flux penetration. For this reason in the figure this condition

is indicated by the term Meissner state. For higher values of the field, the material return in the normal state and the applied field penetrates freely.



There are, however, materials in which the magnetic behaviour is more complicated. There is also in these cases a lower critical field, H_{c1o}, below which the magnetic flux does not penetrate. Above this, however, there is no transition to the normal state, to get to that we must reach a value H_{c2o} said higher critical field. For applied fields between these two values there is a partial penetration of the flow into the material and we speak of mixed state or even state of Shubnikov. These materials are called superconductors of type II. Most superconductors are of this type. Superconductors of the second type are more interesting for applications because they have higher critical temperatures and currents. Among the type II superconductors there are also the so called "high T_c superconductors" discovered in the 80s of the last century.

8.4 The superconducting gap

In a normal metal the specific heat c_n varies with temperature as $T + T^3$, the linear term is that due to electrons, the cubic one to the lattice. In superconductors, on the other hand, the specific heat c_s shows an abrupt increase at Tc and then it drops exponentially. If the material is released from the superconducting state, for example by applying a fairly intense magnetic field, the trend of the normal

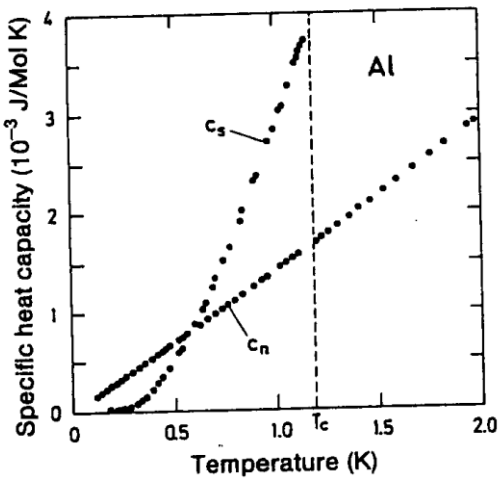
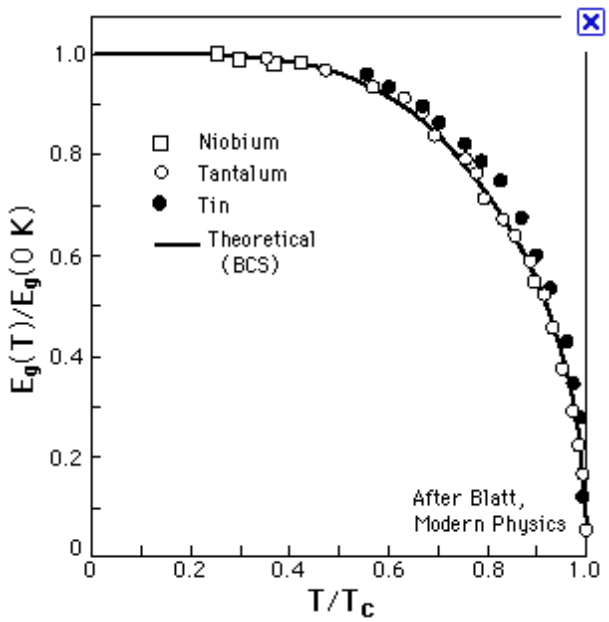


Fig. 10.3. Specific heat of normally conducting (c_n) and superconducting (c_s) aluminium. Below the transition temperature T_c, the normally conducting phase is created by applying a weak magnetic field of 300 G. (After [10.3])

state c_n is recovered. The presence of a discontinuity in the specific heat suggests that the transition to the superconducting state can be considered as a phase transition.

An exponential trend suggests that there is a gap of empty states that separates the fundamental state of the superconducting electrons from the first excited state. The electrons in the fundamental state, in order to reach the first excited state, must overcome this step of energy and it is therefore understandable that this process of thermal activation, regulated by the Boltzmann factor, has an exponential character.



The hypothesis of the existence of a gap is also supported by other experiments, for example by measurements of the I-V characteristics of superconductor-normal metal junctions, in which it is clearly seen that there is a minimum value of V for having a currents flow. Another class of experiments which reveal the existence of a gap is microwave absorption. These measurements also allow to verify that the gap depends on temperature and is null at the critical temperature T_c as shown in the figure.

8.5 Applications of superconductivity

Despite its more than one hundred years of history, superconductivity is still struggling to find large-scale applications. This is both for the costs of cryogenic apparatuses and for the limits imposed by the critical field and by the critical current, still rather low. By far, the most widespread application are superconducting magnets, such as those used in nuclear magnetic resonance imaging or in some large particle accelerators (the Large Hadron Collider is probably the largest application of superconductivity with its 1800 superconducting magnets distributed over a circumference of 27 km).

The most useful application of superconductors could be the transmission without loss of electricity but for now there are only a few pilot plants, which are prohibitively expensive. The situation is similar with regard to another great potential application, magnetic levitation applied to rail transportation.

In addition to these power applications, there are also superconducting electronic devices, the best known of which is the SQUID, an extremely sensitive magnetic field sensor. Another very specialized application is represented by the so called superconducting cavities that are used in some particle accelerators to obtain very high currents.

8.6 Theories of superconductivity

Superconductivity was discovered in 1911 but it took almost 50 years to understand its origin, and this only in some cases. For the so called high T_c superconductors there is no explanation yet. These are the most important moments that have characterized the difficult process of understanding the phenomenon:

- London equations, formulated by the brothers Fritz and Heinz London in 1935. These two equations are obtained by playing with the Maxwell equations applied to a medium with no resistivity and by introducing, without being able to justify it, an ad hoc hypothesis to explain the Meissner-Ochsenfeld effect.

- Theory of second-order phase transitions, formulated by Landau in 1943 and used for the first time to explain the superfluidity of He-4, for which Landau was awarded the Nobel prize. Later adapted to superconductors (Landau-Ginzburg 1950) and used to explain type II superconductivity in a thermodynamic way (Abrikosov 1954). Both Ginzburg and Abrikosov won the Nobel Prize.

- BCS Theory, published in 1957. It is a very complex microscopic theory that earned Bardeen, Cooper and Schrieffer the Nobel Prize. It was developed independently of N. Bogoliubov.

- L. Gor'kov (1959) has shown that the Landau-Ginzburg theory can be derived from the BCS theory. Neither he nor Bogoliubov got the Nobel Prize, a great injustice.

8.7 London equations

If an electric field E is applied to a conductor, in a purely classical model an electron inside it moves according to Newton equation:

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E} - \frac{m}{\tau}\mathbf{v}_D$$

The term on the left is the product of mass and acceleration, the two terms on the right represent the forces acting, one due to the applied field and the other to the viscous friction proportional to the drift velocity $\mathbf{v}_D$. The friction coefficient is m/τ where m is the mass of the electron and τ a characteristic time (relaxation time). This equation leads to the derivation of the Ohm's law, which associates a constant current to a constant electric field. If instead we suppose that $\rho = 0$ the term due to friction vanishes because τ becomes infinite and therefore the particle accelerates indefinitely, as in the case of free electrons. The equation becomes:

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E}$$

Let us suppose that in the superconducting material there is a fraction of carriers able to move without encountering resistance, as supposed in a model due to Gorter and Casimir in 1934. In this case the

superconducting current density is $\mathbf{j}_s = -en_s\mathbf{v}_D$, where n_s is the density of superconducting electrons. Combining it with the previous expression for speed we can write:

$$\frac{d\mathbf{j}_s}{dt} = n_s e\frac{d\mathbf{v}}{dt} = \frac{n_s e^2}{m}\mathbf{E}$$

This is the <u>first London equation</u>. This law relates the electric field and the currents in a material where $\rho = 0$. In the presence of a constant field the current increases indefinitely. In addition, even in the absence of an electric field it is possible to have a stationary current.

Another essential aspect of superconductors, as we have seen, is their magnetic behavior, so we have to write an equation for field **B** as well. Normally the link between the electric field and the magnetic field is given by Faraday-Lenz law:

$$\mathrm{rot}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

Using the expression previously obtained for the electric field in a material where $\rho = 0$, i.e. the first London equation, we obtain:

$$\frac{\partial}{\partial t}\left(\frac{m}{n_s e^2}\mathrm{rot}\,\mathbf{j}_s + \mathbf{B}\right) = 0$$

This equation binds magnetic field and current in the case of an ideal conductor with zero resistivity. The equation is certainly satisfied if the term in brackets is constant over time and this happens if fields and currents are constant. This is not strange because it is well known that constant currents produce constant magnetic fields.

Inside superconductors, however, there are no magnetic fields, i.e. $\mathbf{j}_s = 0$ and $\mathbf{B} = 0$, so the term in parentheses is not simply constant, it is equal to zero. The idea of the London brothers was to suppose that the term in parentheses is always equal to zero, a hypothesis from which comes <u>the second London equation</u>:

$$\mathrm{rot}\,\mathbf{j}_s = -\frac{n_s e^2}{m}\mathbf{B}$$

The second London equation resembles the well known one that says that magnetic field flux lines are closed lines surrounding currents, $\mathrm{rot}\,\mathbf{B} = \mu_o\mathbf{j}_s$. In this case, however, the roles are reversed and the equation means that if there is a magnetic field in the superconductor there are also currents flowing along closed lines wound around the field lines.

It should be pointed out that there are no arguments based on classical electromagnetism capable of justifying this ansatz whose legitimacy occurs a posteriori because by means of it it is possible to describe correctly the Meissner effect (describe, be cautious, not explain). To do so we start from these equations:

$$\mathbf{E} = \frac{m}{n_s e^2}\frac{d\mathbf{j}_s}{dt} \qquad \text{First London equation}$$

$$\mathbf{B} = -\frac{m}{n_s e^2}\,\text{rot}\,\mathbf{j}_s \qquad \text{Second London equation}$$

$$\text{rot}\,\mathbf{B} = \mu_o \mathbf{j}_s \qquad \text{Maxwell equation (with no displacement term)}$$

With some calculation we get:

$$\text{rot rot}\,\mathbf{B} = \mu_o \text{rot}\,\mathbf{j}_s = -\frac{\mu_o}{\lambda_L}\mathbf{B}$$

$$\text{rot rot}\,\mathbf{j}_s = -\frac{1}{\lambda_L}\text{rot}\,\mathbf{B} = -\frac{\mu_o}{\lambda_L}\mathbf{j}_s$$

Hence:

$$\nabla^2\mathbf{B} = \frac{\mu_o}{\lambda_L}\mathbf{B}$$

$$\nabla^2\mathbf{j}_s = \frac{\mu_o}{\lambda_L}\mathbf{j}_s$$

Where we placed:

$$\lambda_L = \frac{m}{n_s e^2}$$

∇^2 is, mathematically speaking, the vector laplacian that acts on a vector and returns another vector (unlike the scalar laplacian that acts on a scalar field and returns a scalar). In Cartesian coordinates the components of $\nabla^2\mathbf{B}$ are simply the second derivatives of **B** along the three spatial directions.

The solutions of these equations, i.e. the spatial dependence of currents and fields, are damped exponentials with a carachteristic attenuation lenght

$$\Lambda_L = \sqrt{\frac{m}{\mu_o n_s e^2}}$$

called the London penetration depth. It represents the depth at which the field penetrates into a material, which is the same as the region where the supercurrents flow. An estimate of the orders of magnitude provides a few tens of nm. Λ_L also depends on T, and the BCS theory shows that the trend is of the type shown in the figure, with a divergence at T_c, as it should be since in a normal conductor (i.e. just above T_c) the magnetic field can penetrate indefinitely (obviously the material must not be a ferromagnet).

One consequence of all this is that no current can flow in the interior of a conducting material as it is the case of normal conductors. The reason is that a superconductor does not tolerate the presence of a magnetic field, including that generated by the current itself. As a result, currents can only flow near the surface of the superconductor. This would be a huge limit to applications if only

superconductors of the first type existed. Fortunately, there are also those of the second type where, as we have seen, the magnetic field can partially penetrate.

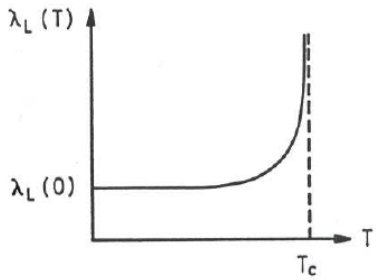
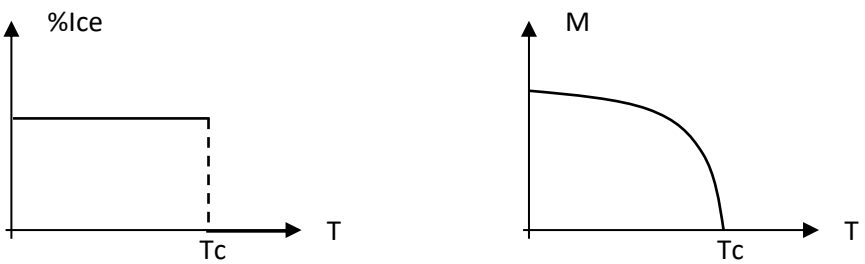


Fig. 7: Temperature dependence of the London penetration depth.

8.8 Phase transitions and their order

Another type of approach to the phenomenon of superconductivity is that based on the theory of phase transitions of Landau. It is a thermodynamic approach that does not explain the microscopic mechanisms but that still has an enormous predictive power. The idea behind this approach is to consider the transition from the normal state to the superconducting state as a phase transition. Whether it is really a phase transition is suggested by the trend of the specific heat that, as we have seen, at T_c changes discontinuously. It is, however, a phase transition different from those more known as fusion or boiling. These are called first order, whereas the superconducting one is called a second order transition



In order to understand what is meant by order of a transition let us think of cooling water by measuring the fraction of ice above and below the transition temperature. For positive temperatures (if measured in degrees Celsius) the fraction of ice is zero, below it there is only ice, so at the transition temperature we can see a discontinuity as shown in the figure. Transitions of this type are of the first order.

If, on the other hand, we consider how the magnetization M varies during the transition of a material from the para to the ferromagnetic state, we observe a continuous trend, as shown in the figure. In this case the transition is considered of the second order. This trend trend is very similar to that of the superconducting gap and this is a further argument, in addition to what has been seen in relation to the specific heat, to support the idea that the superconducting one can be considered a phase transition of the second order.

A more precise definition of the order of a phase transition is that based on the Ehrenfest criterion, which start from the expression of the free energy and calculates the successive derivatives with respect to T in order to obtain the other thermodynamic quantities. By free energy here we mean both the Helmoltz and the Gibbs free energy, which of the two is better to use depends on the type of problem treated, in particular on the thermodynamic variables with whom we work. In the case of the free energy of Helmoltz we know that it assumes its minimum value operating at constant temperature and volume while in the case of Gibbs the minimum is reached by operating at constant temperature and pressure. These variational principles are quite similar to what concerns entropy which, as we know, assumes its maximum value when equilibrium is reached in an isolated system.

The order of a phase transition, according to this criterion, is that of the first discontinuous derivative that is found by subsequently deriving the free energy. Referring to the Helmoltz free energy, we recall that it is defined in this way:

$$F = E - TS$$

If we calculate the differential dF, and remember that $dE = TdS - PdV$, we get:

$$dF = -SdT - PdV$$

If we express F as a function of T and V, $F = F(T, V)$, we obtain these expressions:

$$S = -\left(\frac{\partial F}{\partial T}\right)_V \quad \text{e} \quad P = -\left(\frac{\partial F}{\partial V}\right)_T$$

Entropy is therefore the first derivative of free energy at constant volume.

The specific heat is defined as:

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V$$

We know that $E = F + TS$. If we calculate the derivative by remembering the previous expression for the entropy S, which is the derivative of F with respect to T at constant volume changed in sign, we obtain that the specific heat at constant volume is connected to the second derivative of the free energy:

$$C_V = \left(\frac{\partial F}{\partial T}\right)_V + S + T\left(\frac{\partial S}{\partial T}\right)_V = T\left(\frac{\partial S}{\partial T}\right)_V = -T\left(\frac{\partial^2 F}{\partial T^2}\right)_V$$

Based on this classification, transitions involving latent heat and taking place at a fixed transition temperature (melting of a solid, evaporation of a liquid) are first order transitions because there is a discontinuity in the entropy as a function of temperature. In the case of a mole of substance, and assuming that it is the latent molar heat, once the melting of the substance is over the variation of entropy is $\Delta S = \Delta Q/T = \lambda/T$. The energy of the system in turn increases by an amount equal to λ. The free energy instead remains the same because its variation is:

$$\Delta F = \Delta E - T\Delta S = \lambda - T\frac{\lambda}{T} = 0$$

During the first order transitions, the two phases coexist during the transition that takes place at a constant temperature. In the second order transitions, on the other hand, the transition from one phase to another takes place continuously by changing T and what is observed is a leap in the specific heat but not in entropy. This is for example what happens in the transition between the para and the ferromagnetic state, which for this reason must be considered a second order transition. The temperature below which the phenomenon begins to manifest itself is the critical temperature T_c of the transition.

8.9 The Landau theory of phase transitions

The classification of Ehrenfest is purely thermodynamic and therefore does not provide any information on what is happening in the material at the microscopic level. This is provided by an idea of Landau who suggested to link the second order phase transitions to the variation of the symmetry of the system.

To understand the role of symmetry in phase transitions we think of a magnetic material above critical temperature, better known in this case as Curie temperature, T_{Cu}. In these conditions the material is paramagnetic, i.e. spins are randomly oriented in all directions and the system as a whole is isotropic. In terms of symmetry it can be said that any rotation leads to the same physical situation. We are in a state of high symmetry.

When we go below T_{Cu} the magnetic moments begin to align. In this situation the symmetry is lower because the complete rotational symmetry of the paramagnetic state is lost. The transition to the magnetized state is therefore accompanied by a decrease in symmetry, therefore the term symmetry breaking is used. It should be noted that the change of symmetry does not depend on the actual value of the magnetization. It is sufficient to have a minimal value of the latter to be able to say that the symmetry of the system has changed.

The idea of linking symmetry variations with the appearance of new phases of the physical system is a brilliant intuition of enormous scope. It has been applied to physical systems as diverse as magnetic materials, superfluids, superconductors, and is the basis of some of the most recent theories of elementary particles. The famous Higgs boson prediction was based on Landau's theory of phase transitions. Some people think that the appearance of the presently known fundamental interactions is the consequence of a series of broken symmetry phase transitions that occurred as the temperature of the universe decreased after the big bang.

When a transition occurs in which the symmetry decreases and the "order" is born, it is better to define a parameter η to quantify this order. For instance, in the case of the magnetic transition the order parameter is **M**, the material magnetization. Above the critical temperature there is no order and the order parameter is null, $\eta = 0$.

When the temperature is equal to the critical temperature, the order parameter is null and below it increases gradually and without presenting discontinuities, as instead happens in first order transitions.

The idea behind Landau's theory is to express the free energy F (or G), which is supposed to be continuous, in series of powers close to the critical temperature.

$$F(\eta) = F_o + F_1\eta + F_2\eta^2 + F_3\eta^3 + F_4\eta^4 + \ldots - \eta B$$

The value assumed in equilibrium is the one that minimizes the free energy. Let us specify some aspects of the previous formula:

- The order parameter in general may depend both on position and time, $\eta = \eta(\mathbf{r}, t)$. In this way, spatial and temporal fluctuations can be included in the model.

- F is to be understood as the energy density.

- F_o is the energy of the high symmetry state, which we assume to be constant near the transition temperature. Being constant, and being its derivatives the physically interesting quantities, it can be neglected.

- Temperature dependence is taken into account by means of the development coefficients: $F_i = F_i(T)$.

- The function $F(\eta)$ must be an even function of η for symmetry reasons and therefore odd order terms must be cancelled. Think for example of the magnetic transition where the energy of the system must remain the same after inverting all the spin.

- The the third order term may be different from zero, in which case it can be shown that the transition is of the first order.

- B is a possible externally applied field that couples with the order parameter η that for now we will neglect and therefore we can conclude by assuming that the expression of free energy is:

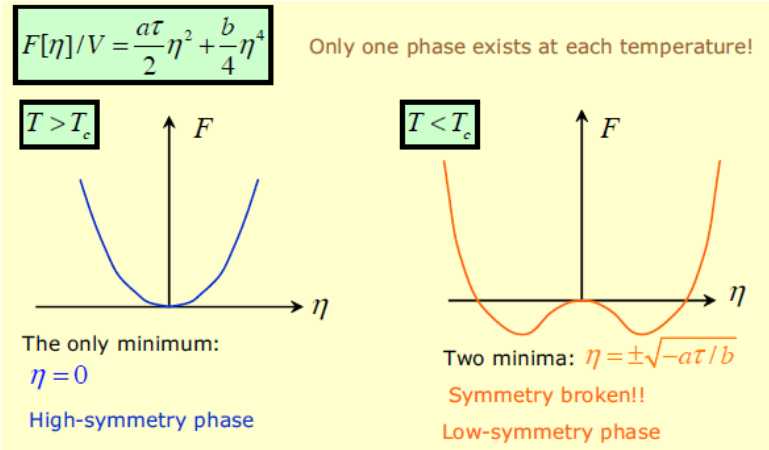
$$F(\eta) = F_2(t)\eta^2 + F_4(t)\eta^4$$

- The two coefficients F_2 and F_4 depend on the temperature and can, at least in principle, assume any value, greater or less than zero. Now let's see what the different possibilities are.

- First let's assume that both are positive. In this case the curve representing the free energy as a function of the order parameter is always positive with the concavity facing upwards and with a minimum at $\eta = 0$. This is what happens above of the critical temperature where the equilibrium state of the system is that in which the order parameter is null.

- F_4 must be always positive, at any temperature. In fact the state of equilibrium of the system is that in which $F(\eta)$ presents a minimum; if F_4 were negative the quartic terms would become prevalent with the growth of η and one would not have a minimum. It is usual to write $F_4 = b/4$, with b always positive.

- F_2 can be both positive and negative. We have already seen how a positive value corresponds to temperatures higher than T_c.For $T < T_c$, on the other hand, the coefficient F_2 can take a negative value. If F_2 is negative and F_4 remains positive, the curve is no longer a parabola but has a relative maximum for $\eta = 0$, i.e. an unstable point, and two minima in symmetrical positions where η is different from zero. The equilibrium states of the system are therefore represented by the two minima. The surprising thing is that in these states of equilibrium the order parameter is no longer null. In other words a phase transition took place because something suddenly appears that was not there before, above T_c. Which of the two minima the



system chooses depends on statistical fluctuations or on the presence of some external field (think, for example, of which direction the magnetization takes falling below the Curie temperature). As the system move in one of these minima, of course, its free energy decreases.

- If F_2 is positive above T_c and negative below, it must be $F_2 = 0$ when $T = T_c$. This is because we are assuming that F_2 is a continuous function of T, which is reasonable since it is used to express the free energy which, in a transition of the second order, must vary with continuity with respect to T. For small temperature variations we can therefore assume that F_2 linearly depends on $\tau = (T - T_c)/T_c$ and we usually write $F_2 = a\tau/2$, with $a > 0$.

After these considerations the expression for $F(\eta)$ takes the following form:

$$F(\eta) = \frac{a\tau}{2}\eta^2 + \frac{b}{4}\eta^4$$

By imposing that the derivative of F with respect to the order parameter is null we obtain the values of η corresponding to the minima of $F(\eta)$, i.e. to the equilibrium values of the order parameter:

$$\eta = \pm\sqrt{-\frac{a\tau}{b}} \propto -\sqrt{T - T_c}$$

There is also a solution for $\eta = 0$ but it is a maximum and therefore corresponds to an unstable equilibrium. The free energy corresponding to the values of stable equilibrium is:

$$F = -\frac{(a\tau^2)}{4b} \propto -(T - T_c)^2$$

From this we calculate the entropy $S = -\partial F / \partial T \propto (T - T_c)$ and the specific heat that for $T = T_c$ results to be:

$$C = T_c \left(\frac{\partial S}{\partial T}\right)_{T=T_c} = -T_c \left(\frac{\partial^2 F}{\partial T^2}\right)_{T=T_c} = \frac{a}{2bT_c}$$

Above T_c the free energy does not depend on temperature because we are assuming that it is constant, and therefore C = 0, while below it assumes the value just seen. This means that at T_c there is a

discontinuity of the specific heat, discontinuity that is not present in the case of entropy. Obviously we are talking only about the part of the free energy and of the specific heat of the new phase characterized by the order parameter. There are also those of the not ordered normal state but, like all the other thermodynamic quantities of the normal state, we can consider them as constant close to T_c.

To sum up, Landau's theory predicts that, below a certain critical temperature T_c, equilibrium states exist for non zero values of the order parameter. These states are equivalent to each other because the energy of the system is symmetrical with respect to the transformation $\eta \rightarrow -\eta$. However, the system is forced to choose one of the two and in doing so two things occur:

- the system goes into a state whose symmetry is less than that of the physical laws that describe it, in the case just seen, the symmetry of mirroring of the function $F(\eta)$ is lost in the actuale state chosen by the system. In these cases we talk of <u>spontaneous symmetry breaking</u> (SSB).

- In the reduced symmetry state, the order parameter takes on a value different from zero.

The breaking of symmetry and the appearance of a value different from zero of the order parameter are therefore to be considered as two different aspects of the phase transition, inextricably linked together.

8.10 The Ginzburg-Landau theory of superconductivity

In the case of the magnetic transition, the reduction of symmetry and the order parameter are easy to spot. But in the case of superconductivity? Let us start with the order parameter, later we will think about symmetry. Landau's stroke of genius, one of many, was to assume that it was a complex number, a kind of wave function, dependent on position and time:

$$\Psi(r,t) = |\Psi(r,t)| e^{i\varphi(\mathbf{r},t)}$$

Today we are able to justify this hypothesis because it has been demonstrated that the square modulus of Ψ represents the density of supercarriers n_s. At that time instead nothing of the microscopic mechanisms of superconductivity was known. However, there was the precedent of the theory of the formation of superfluid He-4 in which a complex parameter is also used whose square modulus represents the density of the superfluid component. The fact that superconductivity is also an exquisitely quantum phenomenon able to manifest itself on a macroscopic scale probably suggested to use also in this case an order parameter similar to a normal wave function.

Once the hypothesis has been made that the order parameter is a complex function of position and time, it is a matter of writing the expression of the free energy as foreseen by Landau's theory and of calculating the minimum to which the state of equilibrium corresponds. Since the order parameter is a function of position and time, the functional expressing the energy of the system must be minimized. Doing all this is obviously not easy and in the following we will only summarize the most important results that are obtained, by making some comment on the salient aspects.

The main advantage of the Ginzburg-Landau theory is that, although it is a thermodynamic theory, it allows to take into account the spatial variations of the superconducting state because the order parameter is a function of position and time. These variations, which it would be better to call <u>fluctuations</u>, play an important role as we saw when, dealing with the shielding of the magnetic fields, we introduced a characteristic length, the London penetration depth. We will see later that, by

exploiting this prerogative, the theory was generalized by A. Abrikosov who correctly interpreted the phenomenon of partial penetration of the magnetic field in superconductors of type II. Its main disadvantage is that it only works near the critical point.

8.11 The Ginzburg-Landau equations

In the Ginzburg-Landau model, based on the Landau theory of phase transitions, the free energy of the system is expressed through this functional:

$$F = F_n + \alpha|\Psi|^2 + \frac{\beta}{2}|\Psi|^4 + \frac{1}{2m^*}\left|\left(-i\hbar\nabla - 2e\mathbf{A}\right)\Psi\right|^2 + \frac{|\nabla\times\mathbf{A}|^2}{2\mu_o}$$

Let us see the meaning of the different terms of the expression that in this case refers to a particular point of space, therefore it represents the spatial density of energy. The total energy of the system will therefore be an integral over the entire space of the energy density.

- $\Psi(\mathbf{r},t)$ is the complex order parameter depending, in general, on position and time.

- F_n is the energy density of the normal state, which in a small range around the critical point we will assume to be constant and therefore will be neglected.

- The two quadratic and quartic terms in Ψ are those provided by the Landau theory of phase transitions.

- The fourth term is the quantum expression of the kinetic energy of the electric supercurrent that is flowing. To understand where this term comes from, let us remember that the kinetic energy is equal to $\mathbf{p}^2/2m$ and that, for an electron whose charge is $-e$ in presence of a magnetic field whose vector potential is $\mathbf{A}$, instead of the normal momentum we must consider the canonical momentum $(\mathbf{p} - e\mathbf{A})$. The quantum operator of the momentum is obtained by substituting to $\mathbf{p}$ the expression $-i\hbar\nabla$. This term refers to the <u>spatial fluctuations</u> of the order parameter and assumes its minimum value when this is constant. Fluctuations therefore have an energy cost that must be taken into account. The term $-e$ represents the charge of the electron, multiplied by two for taking into account the fact that the carriers are pairs of electrons, while m* is the effective mass, in turn multiplied by two.

- If a magnetic field is present close to the surface of the superconductor (not inside due to the Meissner – Ochsenfeld effect), then the fifth term must also be added, remembering that $\mathbf{B}$ is the curl of $\mathbf{A}$.

In order to solve the problem it is first necessary to express the total energy of the system by integrating its density over the whole space. Once this expression has been written, it must be minimized with respect to the functions that express the spatial variations of the order parameter $\Psi(\mathbf{r})$ and the vector potential $\mathbf{A}(\mathbf{r})$. In this way we get to the equations of Ginzburg - Landau:

$$\alpha\Psi + \beta|\Psi|^2\Psi + \frac{1}{2m^*}\left(-i\hbar\nabla - 2e\mathbf{A}\right)^2\Psi = 0$$

$$\mathbf{j}_s = \frac{e}{2m^*}\mathrm{Re}\left\{\Psi^*\left(-i\hbar\nabla - 2e\mathbf{A}\right)\Psi\right\}$$

The first of the two is obtained by minimizing F with respect to $\Psi(\mathbf{r})$, the second with respect to $\mathbf{A}(\mathbf{r})$ and provides the superconducting current $\mathbf{j}_s$. These are two rather complex, non linear equations.

8.12 The characteristic lengths of superconductivity

The first of the Ginzburg-Landau equations allows us to deal with spatial variations of the order parameter. An easy case to deal with is that of the interface between two regions where the material is in a normal state in one side and in a superconducting state in the other. By solving the equation with these boundary conditions one discovers that the transition from the superconducting state to the normal state, described by the complex parameter $\Psi(\mathbf{r})$, takes place within a transition region whose thickness is of the order of the so called <u>coherence length</u>:

$$\xi = \sqrt{\frac{\hbar^2}{4m^*|\alpha|}}$$

The coherence length gives an idea of the extension that the spatial fluctuations of the superconducting state can assume.

The second equation instead relates the supercurrents to the order parameter. If we consider the case of a uniform system, that is where Ψ it is constant and therefore $\nabla\Psi = 0$, the second equation is simplified and becomes:

$$\mathbf{j}_s = -\frac{e^2|\Psi|^2}{m^*}\mathbf{A}$$

If we calculate the curl of the two terms and remember that $\mathbf{B} = \mathrm{rot}A$ we obtain:

$$\mathrm{rot}\,\mathbf{j}_s = -\frac{e^2|\Psi|^2}{m^*}\mathbf{B}$$

This expression coincides with the second London equation already obtained previously on the condition that the square modulus of Ψ represents the density of superconducting carriers that had previously been called n_s. We had already seen that from the London equation it is possible to obtain the length of attenuation of the magnetic field and the supercurrent:

$$\Lambda_L = \sqrt{\frac{m}{\mu_o n_s e^2}}$$

The comparison between these two characteristic quantities in a given material, the coherence length and the London penetration lenght, allows us to make very important predictions about the properties of superconducting materials, in particular it allows us to understand why there are two types of superconducting materials, those in which the magnetic field does not penetrate completely and those in which there is partial penetration.

8.13 The lower critical magnetic field

The transition to the superconducting state occurs for the usual reason that all physical phenomena occur, namely because in doing so the system reduces its energy. Once this state is reached, let us try to disturb it by applying a magnetic field. As we know, when a magnetic field is applied to a superconducting material, supercurrents are created which, in turn, generate a field within the material that perfectly screen the external one.

Circulating the supercurrents and generating a magnetic field that compensates for the applied one, however, costs energy. The amount of energy we need to supply to the system obviously depends on the intensity of the field we want to screen. The larger the field, the more energy must be expended to shield it.

There is a threshold beyond which the energy spent for screening the external field becomes higher compared to the one gained with the transition to the superconducting state, which is essentially constant. Once this threshold is reached, the system no longer finds it convenient to remain in the superconducting state. This explains in a qualitative way the fact that there is a critical magnetic field and also the fact that this field increases as the temperature decreases This under the reasonable assumption that the energy gain of the system following the superconducting transition increases as T decreases.

What happens if the value of this critical field is exceeded serves to distinguish the two known types of superconductors, those of type I and those of type II. To understand the reasons for the different behavior in these two cases, it is necessary to deal in more detail with what happens at the interface between a material in the superconducting state and a normal one.

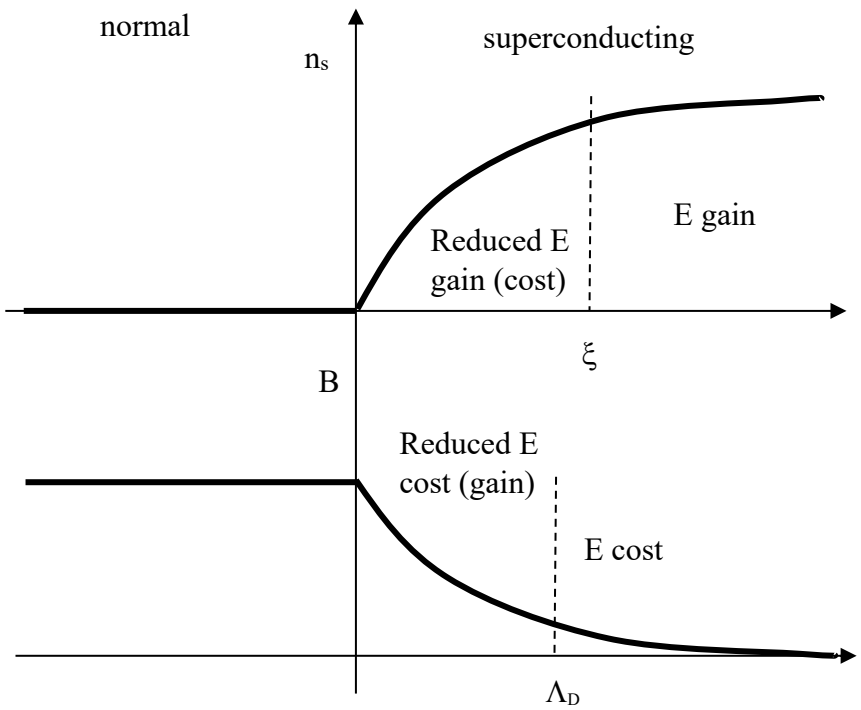
8.14 Energy considerations and type II superconductors

The Ginzburg-Landau theory allows us to work with spatially inhomogeneous systems such as an interface. Abrikosov, developing this theory, was able to correctly interpret what happens in type II superconductors.

Let's look at the figure that shows what happens at the interface between a normal material and a superconductor. The two graphs show, respectively, the trend of the density of the superconducting carriers n_s and that of the magnetic field B versus position. In the first case the transition zone to the superconducting state has a thickness of the order of the correlation length ξ. In the case of the magnetic field, the transition zone has a thickness of the order of the London penetration length, Λ_D.

If we consider samples of much larger dimensions than ξ and Λ_D then the role of the interfaces is negligible. The relevant contributions to the total energy, in this case, are those of the bulk, i.e., the energy gain due to the superconducting transition and the energy cost associated with shielding the magnetic field. However, when the critical field is reached, these two contributions cancel each other out and therefore in calculationg the energy balance the role of the interfaces becomes relevant.

As regards the density of superconducting carriers, in the transition zone of thickness ξ the energy gain associated with the transition is reduced because the concentration of carriers is lower. We can consider this reduced gain as an energy cost that we indicate with symbol E_ξ. On the other hand, if the magnetic field is considered, in the transition zone within Λ_D the energy cost is lower since the magnetic field partially penetrates. We can therefore consider this as a decrease of the energy of the system that we indicate as $-E_\Lambda$.



We are interested in what happens at the interface to understand if the possible formation of alternating layers of material in the normal state and in the superconducting state can be energetically favored. In making this estimate we will assume that the amount of energy lost or gained is proportional, respectively, to the correlation length ξ and to the London penetration length Λ_D. We therefore distinguish two cases.

$\xi > L_D$: in this case $E_\xi - E_\Lambda > 0$ and this means that, once the critical field has been exceeded, the system has no way of remaining in the superconducting state because the creation of layers in which the normal state alternates with the superconducting state has a higher energy cost than returning all the material to its normal state. This is what happens in type I superconductors.

$\xi < L_D$: in this case instead $E_\xi - E_\Lambda < 0$ and this means that the formation of an interface leads to a decrease in the energy of the system. For this reason the system finds it more convenient to separate into zones in the normal state alternating with zones in the superconducting state. This is roughly what happens in superconductors of type II.

The reasoning just made allows us to qualitatively understand how an energy balance allows us to understand the different behavior of superconducting materials. A more detailed theory, formulated by Abrikosov, leads to predicting a more complex situation as represented in the figure. For this theoretical prediction, confirmed by experiments, Abrikosov was awarded the Nobel Prize.

According to this theory, in type II supeconductors the magnetic flux flow lines penetrate in some points of the material that, locally, is not in the superconducting state. Around these flow tubes there are areas of the material that are in the superconducting state and in which supercurrents circulate that prevent the field from penetrating the superconducting area.

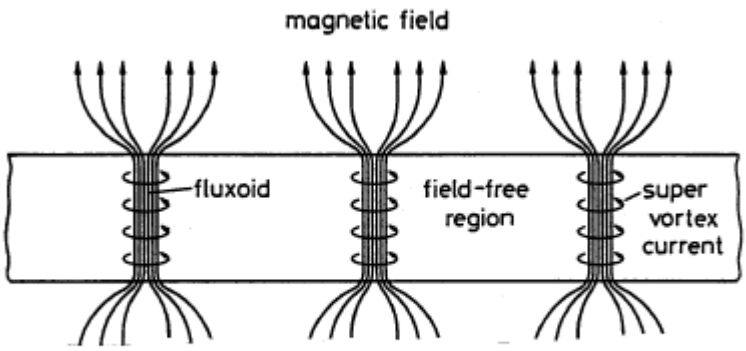


Fig. 11: Flux tubes in a type II superconductor.

These supercurrent vortices, called Abrikosov vortices, are arranged in an ordered way forming real lattices, the lattices of Abrikosov. The theory correctly foresees that the concatenated flux at each vortex is equal to a quantum of the magnetic field Φ_o.

To understand whether a material is type I or II, the so-called Ginzburg-Landau parameter is introduced:

$$\kappa = \frac{\Lambda_L}{\xi}$$

If $\kappa < \frac{1}{\sqrt{2}}$ then the superconductor is type I, otherwise type II.

8.15 The vector potential

Any model of superconductivity, even if of an exclusively thermodynamic type, must necessarily take into account the essential aspects of the phenomenon. The first is the presence of supercurrents, which, as we know, can flow indefinitely as a consequence of the cancellation of resistivity. The second is the effect of the magnetic fields, whether generated by the same currents or applied from outside.

To take into account the magnetic fields it is convenient to use the vector potential **A** which is defined by the relationship that binds it to the magnetic field **B**:

$$\mathbf{B} = \nabla \times \mathbf{A}$$

It is possible to demonstrate that the equation of motion of an electron with charge –e in the presence of a vector potential can be written by replacing in the law of Newton (or in the Schrodinger equation) the ordinary momentum **p** = m**v** with the so called canonical momentum also called conjugated momentum:

$$\mathbf{p} - e\mathbf{A}$$

In quantum mechanics the wave function of a free particle with momentum **p** is described by a wave function of the type $e^{i\mathbf{k}\cdot\mathbf{r}}$ or, remembering that $\mathbf{k} = \mathbf{p}/\hbar$, as $e^{i\frac{\mathbf{p}}{\hbar}\cdot\mathbf{r}}$. A displacement $\Delta\mathbf{r}$ therefore corresponds to a variation $\Delta\varphi = \frac{1}{\hbar}\mathbf{p}\cdot\Delta\mathbf{r}$ in the phase of the wave function.

If instead of the simple momentum **p**, the conjugated momentum $\mathbf{p} - e\mathbf{A}$ is considered, the phase shift becomes:

$$\Delta\varphi = \frac{1}{\hbar}(\mathbf{p}\cdot\Delta\mathbf{r} - e\mathbf{A}\cdot\Delta\mathbf{r})$$

In addition to the term related to the ordinary momentum, which has to do with the motion of the particle and therefore, if it is charged, with the corresponding current, there is also an additional phase shift due to the presence of the vector potential:

$$\Delta\varphi = -\frac{e}{\hbar}\mathbf{A}\cdot\Delta\mathbf{r}$$

The fact that the vector potential produces a clearly identifiable effect directly on the wave function makes it the most appropriate physical quantity to describe the quantum interaction of charged particles with the electromagnetic field.

To understand in physical terms the effect of the vector potential on the state of a charged particle, let us consider the case of a solenoid of infinite length in which a current flows. As we know, inside the solenoid there is a direct constant magnetic field $\mathbf{B}_o$ directed along the axis of the solenoid while outside the field is zero. Vector potential **A**, on the other hand, does not cancel itself outside the solenoid, let us see why.

According to Stokes' theorem, the flow of **B** through a surface S is equal to the line integral of the vector potential **A** along a line Γ that surrounds the surface (this is because **B** is the curl of **A**):

$$\Phi_{\mathbf{B}} = \iint_S \mathbf{B}\cdot dS = \int_\Gamma \mathbf{A}\cdot d\mathbf{l}$$

In the case of the solenoid we consider a circumference perpendicular to the solenoid and whose centre is on the axis of the solenoid itself. Suppose the circumference is larger than the solenoid. The flux of the magnetic field through this circumference is easy to calculate and is equal to $\pi R^2 B_o$; the circulation of the vector potential along the same circumference is equal to $2\pi rA$. Equalizing these two expressions we obtain the value of the vector potential A outside the solenoid:

$$A = \frac{1}{2}\frac{R^2 B}{r}$$

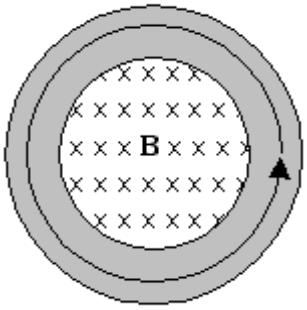
The fact that outside the solenoid the vector potential is different from zero, even if the field is null, is at the basis of the Ehrenberg-Siday-Aharonov-Bohm effect, illustrated in the figure. The electrons emitted by a source are passed through two slits in such a way as to produce two coherent beams, which as is known produce a diffraction figure on a screen placed downstream of the two slits. Before interfering with the screen, however, the two beams pass on both sides of a solenoid. If a current

flows in the solenoid, a magnetic field is generated which is null on the outside, unlike the vector potential which is also present on the outside.

The presence of the vector potential introduces a phase shift of opposite sign in the two beams since the vector potential is oriented differently in the two cases. The different phase variation on the two beams produces a variation observable in the interference figure. This effect, first predicted in 1949, was observed experimentally only much later, in 1998.

8.16 Quantization of the magnetic flux

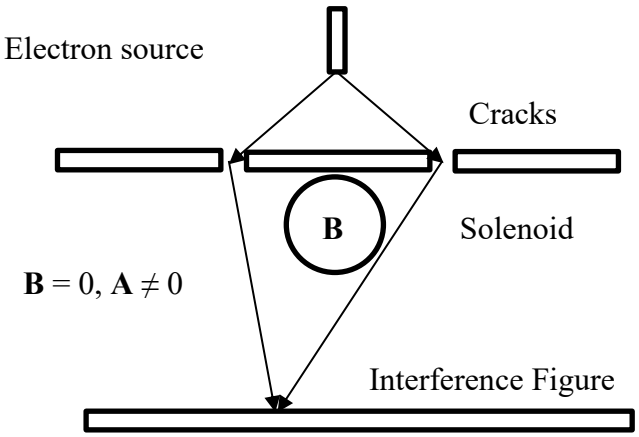
Let us take a ring made of a superconducting material of type I, that is able to completely expel the magnetic field from its interior. Let us suppose that there is a magnetic field **B** in the hole of the ring, and that this field is associated with a vector potential **A**.



Let us consider a circular path inside the ring like the one indicated by the continuous line shown in the figure. The integral of the phase of the order parameter calculated along this closed path must of course be null, or a multiple of 2π.

$$\oint d\varphi = n \cdot 2\pi$$

Let us suppose that the order parameter behaves exactly like the wave function of a charged particle, whose phase varies spatially due to two factors. The first is the movement of the charges connected with the presence of a supercurrent. In the case we are considering, however, we neglect this term because we know that the supercurrent, if present, flows only within a surface layer whose thickness is of the order of a few tens of nm. If the ring we are considering is of macroscopic dimensions, we can assume that the current inside it is null.



The phase change is therefore only due to the vector potential .

$$\oint d\varphi = \oint -\frac{e}{\hbar}\mathbf{A}\cdot d\mathbf{l} = -\frac{e}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} = n\cdot 2\pi$$

The line integral of **A**, however, according to Stokes' theorem, is equal to the magnetic flux chained, and therefore, by neglecting the minus sign:

$$-\frac{e}{\hbar}\oint \mathbf{A}\cdot d\mathbf{l} = -\frac{e}{\hbar}\Phi_{\mathbf{B}} = n\cdot 2\pi$$

From which we get:

$$\left|\Phi_{\mathbf{B}}\right| = n\cdot\frac{\hbar}{e}2\pi = n\cdot\Phi_o \quad \Rightarrow \quad \Phi_o = \frac{2\pi\hbar}{e} = \frac{h}{e}$$

In other words, the flow through the closed line is quantized, being Φ_o the flow quantum! This effect was predicted by Fritz London in 1948 and was then verified experimentally in the 1960s. However, the measurement have led to the conclusion that the flux is not h/e but h/2e:

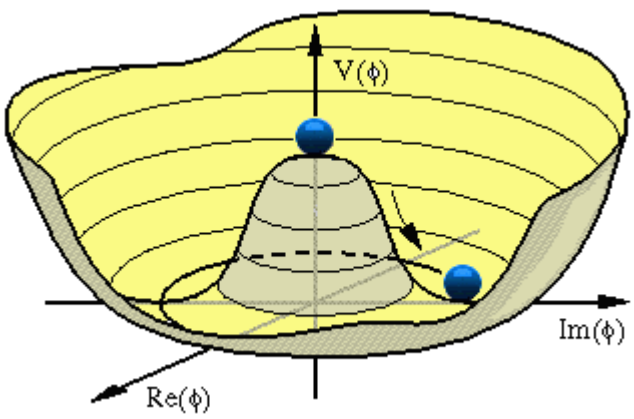
$$\Phi_o = h/2e = 2.07 \text{ x } 10^{-15} \text{ Wb}$$

This result confirm that charge carriers are not just electrons but pairs of them. The BCS theory in fact explains the superconductivity in terms of pairs of electrons weakly linked to each other, the so called Cooper pairs.

8.17 Symmetry breaking with a complex order parameter

If the order parameter is a scalar quantity, as in the case of magnetization in the para-ferromagnetic transition, the potential energy below T_c can be described by a curve with two minima in the plane (F, η).

In the case of a complex parameter such as in the superconducting transition, both the real and the imaginary part (or the modulus and the phase) must be considered, and therefore the potential energy of the system becomes a surface in a three dimensional diagram. Below T_c this surface takes on a shape, shown in the figure, similar to the bottom of a bottle of wine or a sombrero. It is because of these similarities that physicists speak of "mexican hat potential".

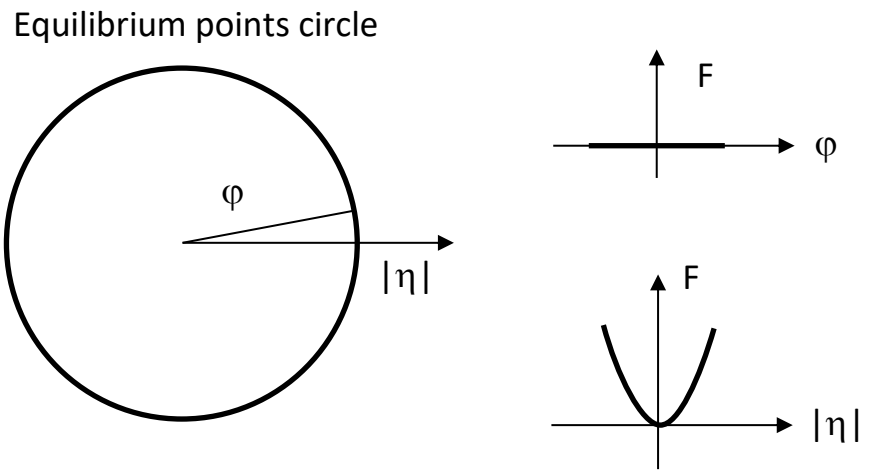


Also in such potential the position corresponding to a null value of the order parameter, $\Psi = 0$, is an unstable equilibrium position. Positions of stable equilibrium, on the other hand, are those that are at the bottom of the surface. When the material passes to the superconducting state, its position is a point on the bottom of the potential and in doing so it "chooses" a particular value of the phase, which is instead undefined when $\Psi = 0$. Therefore a symmetry breaking occurs, similar to the one we have already mentioned in the case of the scalar potential and that leads to a non symmetric state even though the potential is symmetric. This breaking of phase symmetry is precisely that connected, within the Landau-Ginzburg theory, with the second order phase transition to the superconducting state.

8.18 Excitations with a complex order parameter

The complex order parameter η that we are considering is a function of the spatial position and time $\eta(\mathbf{r}, t)$, what in physics is called a "field". As is known, in a field there may be excitations in the form of waves, such as in the case of the electromagnetic field or the field of the displacements of the atoms in a solid where the lattice oscillations take place. These are excitations characterized by an oscillation frequency ω and by a wave vector $\mathbf{k}$ linked together by appropriate dispersion relations $\omega(\mathbf{k})$.

Suppose we position ourselves in one of the possible equilibrium points of the chosen system by opting for a particular phase value among those on the bottom of the sombrero potential, which all together constitute a circumference. In the state of minimum energy, in which there are no excitations, the order parameter assumes the same complex value in every point of space and therefore also the same phase. However, it is possible that there may be excitations of this field just as in the case of the field of atomic positions in a lattice.



We distinguish two distinct types of excitations: those where the modulus of the order parameter varies over time and those in which the phase varies. The left part of the figure shows the minimum energy circumference in the sombrero hat potential. The two graphs on the right describe the energy trend moving along the radial direction and along the circle. What mainly distinguishes these two excitations is the trend of the dispersion relations when $\mathbf{k} = 0$. In this limiting case, the law that describes the variation over time of the field is the same in all points in space, that is, the field moves as a single rigid entity.

Let us first consider the case of oscillations in which only the phase varies, that is, those in which the order parameter moves along the circumference of minimum energy with constant modulus. It is therefore a situation of an indifferent equilibrium and therefore by varying the value of the field no recall forces are manifested. The fact that there are no recall forces means that oscillatory motions cannot arise.

On the other hand, in the case of radial excitations, where the value of the modulus of the order parameter varies, a recall force arises which tends to bring it back to the equilibrium value. This allows us to hypothesize excitations in which the trend over time of the order parameter has the same phase in all points of space, that is, it moves in unison throughout the system. In this case there are global oscillations with a well-defined frequency value even though the value of the wave vector is equal to zero.

The analogy with phonons helps to better understand the nature of these two types of excitation.

In the well-known case of acoustic phonons in a one-dimensional chain of equal atoms we have seen that a force acts on each of them and that this force depends on the phase variation between one atom and another. To be exact, this force is proportional to the sine squared of (½ ka), where a is the lattice parameter and k the wave vector. From this it follows that, as k decreases, the force acting on each atom of the chain decreases and with it the frequency of oscillation.

In the case of lattice oscillations, in addition to massless excitations such as acoustic phonons, there are optical phonons, whose dispersion relations do not pass through the origin. In this case the physical meaning is easily understood by considering the oscillations of a lattice consisting of two different atoms in the case in which **k** = 0, a situation where the two sub-lattices defined by the two atomic species oscillate in phase opposition with respect to each other. other. These oscillations, that we could define as "internal" of the system, correspond to a non-zero frquency, which manifests itself even when there are no spatial fluctuations of the system.

The excitations whose dispersion relationship passes through the origin in physics are called "massless" to distinguish them from the other called "massive" ones. To understand the origin of this terminology, let us examine the relativistic relationship that links momentum and energy (in turn linked to the wave vector and frequency through De Broglie relations):

$$E^2 = p^2c^2 + m^2c^4$$

In this relationship, when the momentum tends to zero, the energy does not tend to zero but to a finite value, $E = mc^2$. As is well known, this is the most famous formula in physics that expresses the so-called zero-point energy, that is, the energy that a body possesses when it is stationary, and establishes the equivalence between mass and energy. If, on the other hand, the particles are massless then the dispersion relationship passes through the origin: E = pc, as in the case of photons.

In the case of optical phonons it can be said that even in the absence of spatial modulation the system of the two sub-lattices possesses an internal energy, which is responsible for the fact that the dispersion relationship has a non-zero value when **k** = 0, which is equivalent to the zero point energy in the case of relativistic particles.

The fact that a massless excitation is associated with a spontaneous symmetry breaking of a system with a sombrero-like complex order parameter is the subject of a very important theorem, the Goldstone theorem, formulated in the early sixties of the last century. This theorem states that any spontaneous breaking of a continuous symmetry gives rise to a massless excitation with zero spin

whose quantum what is called the Goldstone boson. In the physics of condensed matter there is the already mentioned case of phonons, whose existence is to be connected with the solidification of a liquid in which, passing from the liquid to the crystalline state, the density loses its complete translational symmetry. Another case is represented by the magnons, the quanta of the spatial fluctuations of the magnetization density, consequent to the occurrence of the ferromagnetic transition.

Although it may seem bizarre, the Goldstone theorem is important, more than for the observed particles whose existence it predicts, for those it predicts but which have never been observed. The first case of this kind was that of superconductivity, where it has never been possible to observe excitations of the density of the superconducting component whose dispersion relationship passes through the origin (at least in three dimensions, in systems of lower dimensionality things change).

We can understand that massless excitations cannot exist in a superconductor by using a simple argument. The superconducting state admits excited states which, however, as we have seen, are separated from the ground state by an energy gap. This means that there is a minimum energy value for these excitations which therefore cannot be without mass.

How is it possible to explain the absence, in some cases, of the Goldstone boson? Answering this question is by no means simple and we must first introduce the concept of gauge invariance. As we will see, continuing along this line we will arrive very, very far.

8.19 Gauge invariance and superconductivity

The vector potential is not univocally defined because it is always possible add to it the gradients of an arbitrary scalar function without affectng the physical laws. Let us suppose we have a vector potential **A** and we add to it the gradient of an arbitrary scalar function $\chi(\mathbf{r})$:

$$\mathbf{A}(\mathbf{r}) \rightarrow \mathbf{A}(\mathbf{r}) + \text{grad}\,\chi(\mathbf{r})$$

The magnetic field **B** remains the same because the curl of a gradient is always equal to zero:

$$\mathbf{B}(\mathbf{r}) = \text{rot}\left[\mathbf{A}(\mathbf{r}) + \text{grad}\,\chi(\mathbf{r})\right] = \text{rot}\mathbf{A}(\mathbf{r})$$

This property is called gauge invariance. When we have gauge invariance the degree of symmetry is high because, in fact, we can add to the vector potential the gradient of any scalar function and the magnetic field **B** remains unchanged. As known, the Noether theorem states that continuous symmetries are associated with conservation laws. In this case, the gauge symmetry of the electromagnetic field is associated with the conservation of the electric charge.

In superconductors the gauge invariance is lost because the vector potential is uniquely defined and not up to the gradient of an arbitrary scalar function. The system is therefore less symmetrical and for this reason the term symmetry breaking is used. To understand why in the superconducting state the gauge invariance is no longer valid, let us take the second London equation written in terms of the vector potential **A**:

$$\mathbf{j}_s = -\frac{e^2 n_s{}^2}{m}\mathbf{A}$$

Since the current is uniquely defined, so must be **A**, that the equation says is proportional to $\mathbf{j}_s$. To be precise, in order to satisfy the conservation of charge (div$\mathbf{j}_s$ = 0) it must also be div**A** = 0. This condition is called the London gauge.

One could think of interpreting the failure of gauge invariance as a breaking of symmetry and linking it, in the spirit of the Landau theory, to the superconducting transition. This is because we go from a situation in which it is possible to add an arbitrary function to the vector potential to one in which the vector potential is uniquely fixed, which corresponds to a decrease in symmetry. In reality this is not the breaking of symmetry associated with the manifestation of superconductivity because it is not a true symmetry of the physical system. The gauge invariance of the electromagnetic field is in fact a manifestation of the uncertainty with which the vector potential is defined, in a sense it is a manifestation of the inadequacy of our description of the electromagnetic phenomena. The breaking of the gauge symmetry has a role in the superconductivity phenomenon but is not the one directly responsible for its establishment. As we have seen before, the broken symmetry is that relative to the superconducting state phase, a state with a defined phase even though all the possible values of the phase are energetically equivalent.

In the expression of the free energy of a superconductor the vector potential enters with two terms, one in which it appears alone and the other where it appears coupled to the superconducting order parameter. As a consequence of this it is possible to demonstrate, and this Philip Anderson did, that in a superconductor the massless exitation provided by the Goldstone theorem is no longer present. A colorful way used by many physicists is to say that the gauge field, in this case the vector potential, "eats" the Goldstone boson.

To justify this situation in a simple way, we have already emphasized the fact that there is an energy gap in superconductors and therefore any excitation inside them must have a minimum energy and excitations with lower energy are not admissible. This is true both for the "plasma" excitations of the electrons and for the electromagnetic field inside the material, whose dispersion relation therefore does not pass through the origin as it happens in vacuum.

This mechanism that leads, as they say, photons to acquire "mass" is known as Anderson mechanism and is a direct consequence of the interaction between electrons in the superconducting state and the electromagnetic field. Incidentally, the fact that it is not possible to have low energy, i.e. low frequency, electromagnetic excitations inside a superconductor is nothing more than the Meissner-Ochsenfeld effect where static or slowly changing fields are perfectly shielded.

These ideas, developed by Philip Anderson in the early 1960s for the case of superconductivity, were extended by other authors to elementary particle physics. Also in the physics of the elementary particles in fact there are examples of violation of the theorem of Goldstone and the most relevant is that of the weak interaction to which are not associated particles without mass and null spin. To try to explain this anomaly the Anderson mechanism was taken up by many authors and extended to the case of the relativistic fields of elementary particles and this allowed to solve the problem of the absence of Goldstone bosons in the weak interaction. In this version, the mechanism foresees that the fields describing the particles sensitive to this type of interaction (the leptons) interact with an additional field, which does not represent a particle (such as the electron field) or a force (such as the electromagnetic field). This additional field, added ad hoc, is the Higgs field. From the interaction of the so called electroweak field, that is, the one that includes the weak and electromagnetic interactions, with the Higgs field it descends that its possible excitations are four. One is a particle of zero mass, the well known photon. The other three, called intermediate vector bosons, were hypothesized by Salam, Glashow and Weinberg (who for this reason took the Nobel prize) and

subsequently observed at CERN in Geneva by a group led by Carlo Rubbia and Simon van der Meer (who in turn were awarded the Nobel prize).

The Higgs field is a field whose energy levels are quantized. The fundamental quantum of the Higgs camp is the famous Higgs boson, observed for the first time in 2013, fifty years after its prediction. These are very high energy quanta, about 125 GeV.

8.20 BCS Theory: Cooper Pairs

To understand the mechanism that gives rise to the superconducting state and explain its extraordinary properties (i.e. the cancellation of resistance and the expulsion of the magnetic fields) it is necessary to develop a microscopic model of the phenomenon. This was done by Bardeen, Cooper and Schrieffer and the theory they developed (BCS theory) is able to adequately explain the phenomenon of superconductivity in many cases even if it is not yet known whether it can be the basis for the understanding, still missing, of the same phenomenon in the so called high critical temperature superconductors.

The starting hypothesis of the model is that between two electrons of opposite spin there is a weak interaction of attractive type that gives rise to a bound state of the two called Cooper pair. By assuming that the two electrons involved have opposite spin, it follows that the total spin of the pair is null, and therefore the system of the two must be considered as a boson. These bosons, as known, can condense in their fundamental state and what derives from it is a coherent sum of the wave functions that describe the single pairs. This is the condensation of bosons also known as Bose-Einstein Condensation (BEC). This, in a nutshell, is the nature of the superconducting state: a condensate of Cooper pairs. There are other examples of boson condensation such as superfluidity or the condensation of cold atoms. Even photons can give rise to condensation, obviously not in the void where their mutual interaction is null, but inside some material medium. These are very topical subjects and a lot of research is presently done on them.

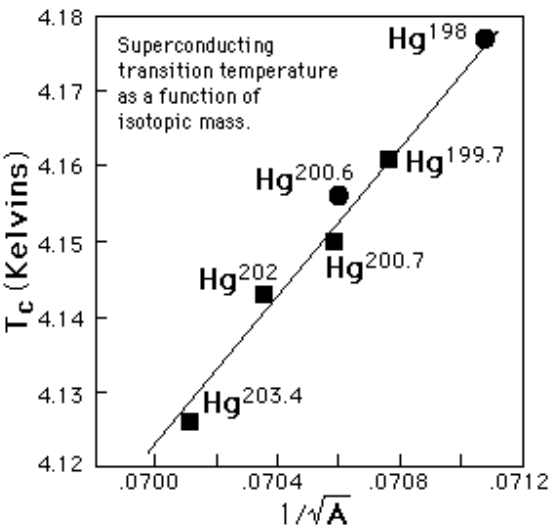
Bose-Einstein Condensation of Trapped Electron Pairs. Phase Separation and Superconductivity of Metal-Ammonia Solutions

RICHARD A. OGG, JR.
Department of Chemistry, Stanford University, California
March 2, 1946

The probable explanation of the above phenomena is to be found in the behavior of trapped electron pairs, recently demonstrated[3] to be a stable constituent of fairly dilute metal-ammonia solutions. In the concentration range characterized by liquid-liquid phase separation, experimental studies[4] show the solute to be diamagnetic at temperatures just above the consolute point. This suggests the electron constituent to be almost exclusively in the trapped electron pair configuration. Because of their zero angular momentum, such pairs must obey Bose-Einstein statistics.

The idea that electron pairs with opposite spin can be formed, thus giving rise to pairs of bosonic nature, was first proposed in 1946 by Richard Ogg who was working on superconductivity in solutions. However, this did not have a particular following despite the fact that it was published in the most important journal of physics, Physical Review.

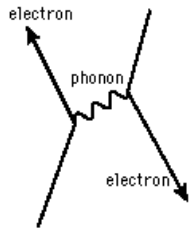
In order to formulate a microscopic model based on the idea of Cooper pairs, it is necessary to formulate ad hypothesis on the physical origin of the attractive interaction between the electrons. The possible role of phonons in establishing the superconducting state was highlighted for the first time by the discovery of the so called isotopic effect: the critical temperature depends on the atomic mass of the atoms of the material, which in turn determines the vibrational properties. The effect was first observed in 1950 by E. Maxwell and by C.A. Reynolds et al. who, in the case of mercury, discovered that T_c is inversely proportional to the square root of the mass of the isotope of which the material is made. Since the frequency of oscillation is inversely proportional to the root of the oscillating mass, the isotopic effect tells us that the critical temperature is proportional to the energy of the phonons $\hbar\omega$.



At first glance, this is quite strange, since phonons normally contribute to slowing down the motion of electrons, and this reinforces the belief that superconductivity has nothing to do with normal phenomena of electronic transport.

A mechanism to explain the existence of an attractive interaction between electrons based precisely on the intervention of the lattice oscillations was proposed by Froelich in 1950, even before the observation of the isotopic effect. This mechanism involves an electron passing through and deforming the lattice causing an increase in the positive charge density of the ions, which in turn attract a second electron. To visualize the thing we think of rolling a ball on an elastic membrane, its weight creates a hole in which a second ball can fall.

The process just described can be interpreted, in a quantum vision, as the creation of a phonon by the first electron and by the subsequent reabsorption of this by the second electron. This way of describing the interaction of two particles through the exchange of the quanta of a mediating field, in this case the lattice, is the way in which all interactions are described within the framework of quantum field theory. The particle responsible for the interaction, in this case the phonon, lives a short life and for this reason it is called "virtual particle".



Describing the formation of Cooper pairs as the exchange of virtual phonons between two electrons allows to fix the order of magnitude of the energies involved in their formation. Since the phonons have a maximum energy of the order $\hbar\omega_D$, the binding energy of the pairs will be lesser or comparable with this one.

This representation of the phenomenon also tells us that only those electrons that are in a narrow shell around the Fermi sphere, whose thickness is of the order of $\hbar\omega_D$, are involved in this interaction. The inner ones cannot participate because the principle of exclusion forbids them to move from the level in which they are, there are no other free ones in the immediate vicinity.

Starting from this idea Cooper considered the case of a gas of free electrons to which are added two electrons at the Fermi level and with equal and opposite values of vector **k** and spin. By igniting a weak attraction potential, as it happens according to the mechanism proposed by Froelich, we create a bound state of the two electrons whose energy is less than the Fermi energy. This has the consequence that the minimum energy state of the system is no longer the traditional one of a Fermi gas, with all the energy levels filled by electrons. In fact, there is the possibility that the system reduces its energy by forming pairs, and this leads to talk of instability of Fermi gas.

Cooper's original work referred to a single pair, the generalization to the collective behavior of the Fermi gas of electrons was done later and led to the formulation of the BCS theory. The fact that this is a very complex issue can be understood by estimating the spatial dimensions of the Cooper pairs.

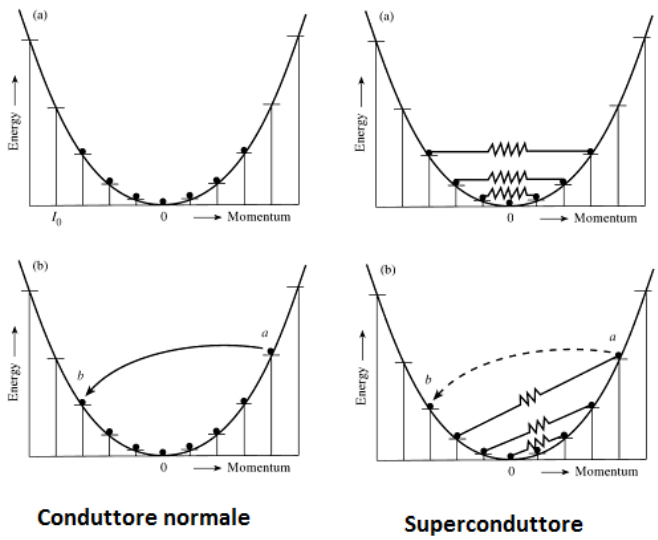
Let us consider an electron whose motion triggers a deformation of the lattice, and therefore an oscillation of the ions at a frequency $\omega_D/2\pi \sim 10^{13}$ Hz. The amplitude of the oscillation (and therefore the increase of positive charge) is maximum after a period whose order of magnitude is the inverse of the frequency of oscillation (10^{-13} s). In this time interval the electron, which moves at a speed $v_F \sim 10^6$ m/s, travels a distance of about 10^{-7} m = 1000 Å. At this point a second electron that is in the deformed zone (the first is now far away) is attracted by the higher density of positive charge that has been created. This so called "correlation distance" gives an idea of the spatial extension of the Cooper pairs and the fact that we can actually speak of an attractive type of interaction. At such great distances, in fact, the coulomb repulsion between the electrons is completely shielded by the other electrons and therefore only the attractive interaction mediated by the lattice survives.

This representation of the mechanism of formation of couples also allows other aspects to be highlighted. The first is that it is a "delayed" interaction and not an instantaneous one. Moreover, couples are not stable systems, they born and die continuously. Finally, within the spatial dimensions of a single pair there are many other electrons and therefore the pairs that these can eventually form overlap spatially. In doing so, the phases of the wave functions that describe the couples also engage with each other and it is precisely in this that the formation of the superconducting state consists.

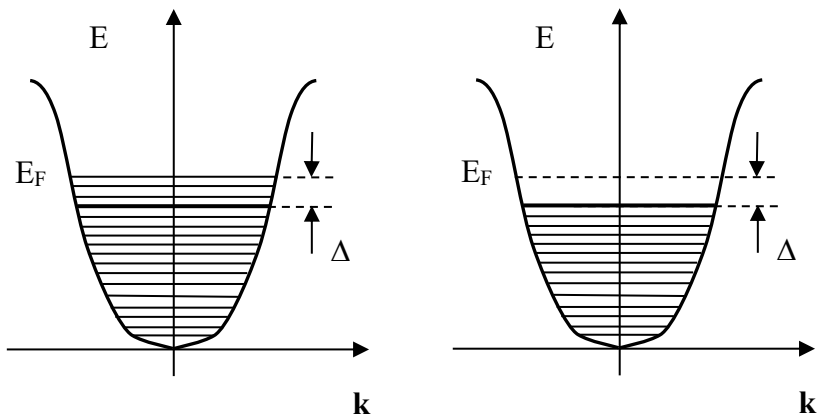
8.21 The BCS state

Once accepted the idea that an attractive interaction between electron pairs can be born, it remains to be understood what it can give rise to by considering all the electrons potentially able to form pairs, i.e. those within an interval $\hbar\omega_D$ around the Fermi level. This is very complicated because if it is true that overall we obtain a reduction in the energy of the system, it is also true that the effect due to the formation of each pair depends on how many pairs are already present and therefore we cannot limit ourselves to summing up the individual effects. In order to obtain the configuration of the many body system , it is necessary to minimize the energy by hypothesizing all the possible configurations of the pairs. Starting from a description of this kind of the superconducting state, the BCS theory allows to deduce many of the properties observed experimentally.

Let us start by considering the cancellation of resistivity. In a normal conductor, applying an electric field unbalances the distribution of electrons in the reciprocal space, and when the field is removed the electrons return, through scattering processes, in their initial states. In the superconductor the electrons attract each other and, in the figure, we represent this interaction with "springs". The springs connect electrons with opposite **k** vector, as required by BCS theory. After creating this asymmetric distribution by applying an electric field, how does the system return to the state of equilibrium once the field is removed? The simplest thing would be to blow up the electron from state a to state b of the figure. This electron however is chained to another electron, and therefore things cannot proceed in this way. The only possible mechanisms should be the simultaneous passage of several electrons in the right states, but this is quite unlikely. For this reason, once a superconducting current has been established, it tends to remain because of the difficulty for the system to relax in the state of equilibrium.



To understand the fact that there is a gap we can reason in this way. We start from a normal gas of free electrons and we turn on the attractive interaction between the electrons that leads to the formation of the bound states that correspond to the Cooper pairs. Initially let us suppose that only one pair is formed, whose state is at an energy $\Delta \simeq \hbar\omega_D$ below the level of Fermi, as we can see in the figure where the energy level of the pair is indicated by the thickest line. In this state the two electrons involved maintain the same momentum they had in the gas of free electrons, **k** and −**k**.



Just as a couple is formed, many others can be formed and, for simplicity's sake, let us suppose that the energy level in which the couples are placed always remains the same. The fact that all the pairs bring themselves in the same state is possible because, unlike the electrons, their spin is null and therefore they behave like bosons. The fact that the overall momentum of the pair, i.e. the sum of **k** and −**k**, is null means that the wavelength of de Broglie associated with the pairs is very large and this allows their consistent overlap. It is for this reason that the transition to the superconducting phase is comparable to the Bose-Einstein condensation of Cooper pairs.

The process of forming pairs proceeds until there are electrons in the occupied states above the energy level of the pairs. Once the maximum value of the energy of the electrons not yet coupled becomes equal to that of the pairs, that is $E_F - \Delta$, we reach a situation of equilibrium that constitutes the superconducting state. This state can be seen as the superimposition of a normal gas of free electrons with the condensate of Cooper pairs. The electrons involved in the formation of pairs are very few, they are a fraction of the order of $\hbar\omega_D/E_F \simeq 10^{-3} - 10^{-4}$. Although it is a small part of the electrons, it is still enough to give rise to the phenomenon of superconductivity and it should be noted that even in the normal conduction of metals the fraction of electrons involved is the same.

Let us just take care of the superconducting fraction and its excited states. It is evident that above the energy state $E_F - \Delta$ the coupled electrons can only move into the state at the level of Fermi and this corresponds to breaking the pair (for the other normal electrons the continuous states of the of Fermi gas are obviously accessible). Therefore, the first excited state for superconducting electrons is separated from the state in which the Cooper pairs are located by an energy gap Δ.

Finally, without going into detail, it should be noted that even the magnetic properties of superconductors can be obtained within the BCS theory by regaining the results of the thermodynamic theory of Ginzburg-Landau. To do this we proceed starting from the fundamental superconducting state and, by introducing a vector potential **A**, we obtain the London equations by means of which it is possible to explain the Meissner-Ochsenfeld effect.

Suggested texts

Ashcroft - Mermin, Solid State Physics (Saunders College)

The text is from 1976 but at that time the foundations of physics of solids had been laid for some time and therefore the book has lost nothing of its formative value. However, many of the most recent developments in solid physics are not present (high-temperature superconductors, quasicrystals, more current investigation techniques such as inelastic X-ray diffusion or tunnel effect microscopy). These are subjects that are not dealt with in the course of the course.

M. Marder, Condensed Matter Physics (John Wiley and Sons)

It is a recent and updated text even if the formalism used often goes far beyond the level of knowledge required (for example, the second quantization method is widely used).

H. Ibach, H. Luth, Solid State Physics (Springer)

This is also a recent and updated text, and the level of treatment accessible. Its limitation is that topics are often dealt with very concisely, which makes it difficult to understand.

J. Ziman, Principles of the Theory of Solids (Tamburini)

Text from the 1960s, very well written, presenting the concepts of solids theory in a refined and elegant way. At a more advanced level, by the same author, Electrons and Phonons.