SOLID STATE PHYSICS

	1 IA	2 IIA	3 IIIA	4 IVB	5 VB	6 VIB	7 VIIB	8 VIIIB	9 VIIIB	10 VIIIB	11 IB	12 IIB	13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
1	**1** 1.0079 **H** Hydrogen																	**2** 4.0025 **He** Helium
2	**3** 6.941 **Li** Lithium	**4** 9.0122 **Be** Beryllium											**5** 10.811 **B** Boron	**6** 12.011 **C** Carbon	**7** 14.007 **N** Nitrogen	**8** 15.999 **O** Oxygen	**9** 18.998 **F** Flourine	**10** 20.180 **Ne** Neon
3	**11** 22.990 **Na** Sodium	**12** 24.305 **Mg** Magnesium											**13** 26.982 **Al** Aluminium	**14** 28.086 **Si** Silicon	**15** 30.974 **P** Phosphorus	**16** 32.065 **S** Sulphur	**17** 35.453 **Cl** Chlorine	**18** 39.948 **Ar** Argon
4	**19** 39.098 **K** Potassium	**20** 40.078 **Ca** Calcium	**21** 44.956 **Sc** Scandium	**22** 47.867 **Ti** Titanium	**23** 50.942 **V** Vanadium	**24** 51.996 **Cr** Chromium	**25** 54.938 **Mn** Manganese	**26** 55.845 **Fe** Iron	**27** 58.933 **Co** Cobalt	**28** 58.693 **Ni** Nickel	**29** 63.546 **Cu** Copper	**30** 65.39 **Zn** Zinc	**31** 69.723 **Ga** Gallium	**32** 72.64 **Ge** Germanium	**33** 74.922 **As** Arsenic	**34** 78.96 **Se** Selenium	**35** 79.904 **Br** Bromine	**36** 83.8 **Kr** Krypton
5	**37** 85.468 **Rb** Rubidium	**38** 87.62 **Sr** Strontium	**39** 88.906 **Y** Yttrium	**40** 91.224 **Zr** Zirconium	**41** 92.906 **Nb** Niobium	**42** 95.94 **Mo** Molybdenum	**43** 96 **Tc** Technetium	**44** 101.07 **Ru** Ruthenium	**45** 102.91 **Rh** Rhodium	**46** 106.42 **Pd** Palladium	**47** 107.87 **Ag** Silver	**48** 112.41 **Cd** Cadmium	**49** 114.82 **In** Indium	**50** 118.71 **Sn** Tin	**51** 121.76 **Sb** Antimony	**52** 127.6 **Te** Tellurium	**53** 126.9 **I** Iodine	**54** 131.29 **Xe** Xenon
6	**55** 132.91 **Cs** Caesium	**56** 137.33 **Ba** Barium	**57-71** La-Lu Lanthanide	**72** 178.49 **Hf** Halfnium	**73** 180.95 **Ta** Tantalum	**74** 183.84 **W** Tungsten	**75** 186.21 **Re** Rhenium	**76** 190.23 **Os** Osmium	**77** 192.22 **Ir** Iridium	**78** 195.08 **Pt** Platinum	**79** 196.97 **Au** Gold	**80** 200.59 **Hg** Mercury	**81** 204.38 **Tl** Thallium	**82** 207.2 **Pb** Lead	**83** 208.98 **Bi** Bismuth	**84** 209 **Po** Polonium	**85** 210 **At** Astatine	**86** 222 **Rn** Radon
7	**87** 223 **Fr** Francium	**88** 226 **Ra** Radium	**89-103** Ac-Lr Actinide	**104** 261 **Rf** Rutherfordiu	**105** 262 **Db** Dubnium	**106** 266 **Sg** Seaborgium	**107** 264 **Bh** Bohrium	**108** 277 **Hs** Hassium	**109** 268 **Mt** Meitnerium	**110** 281 **Ds** Darmstadtiu	**111** 280 **Rg** Roentgenium	**112** 285 **Uub** Ununbium	**113** 284 **Uut** Ununtrium	**114** 289 **Uuq** Ununquadiu	**115** 288 **Uup** Ununpentiu	**116** 293 **Uuh** Ununhexium	**117** 292 **Uus** Ununseptiu	**118** 294 **Uuo** Ununoctium

Alkali Metal
Alkaline Earth Metal
Metal
Metalloid
Non-metal
Halogen
Noble Gas
Lanthanide/Actinide

Z mass
Symbol
Name

57 138.91 **La** Lanthanum	**58** 140.12 **Ce** Cerium	**59** 140.91 **Pr** Praseodymiu	**60** 144.24 **Nd** Neodymium	**61** 145 **Pm** Promethium	**62** 150.36 **Sm** Samarium	**63** 151.96 **Eu** Europium	**64** 157.25 **Gd** Gadolinium	**65** 158.93 **Tb** Terbium	**66** 162.50 **Dy** Dysprosium	**67** 164.93 **Ho** Holmium	**68** 167.26 **Er** Erbium	**69** 168.93 **Tm** Thulium	**70** 173.04 **Yb** Ytterbium	**71** 174.97 **Lu** Lutetium

89 227 **Ac** Actinium	**90** 232.04 **Th** Thorium	**91** 231.04 **Pa** Protactinium	**92** 238.03 **U** Uranium	**93** 237 **Np** Neptunium	**94** 244 **Pu** Plutonium	**95** 243 **Am** Americium	**96** 247 **Cm** Curium	**97** 247 **Bk** Berkelium	**98** 251 **Cf** Californium	**99** 252 **Es** Einsteinium	**100** 257 **Fm** Fermium	**101** 258 **Md** Mendelevium	**102** 259 **No** Nobelium	**103** 262 **Lr** Lawrencium

Mendeleev's periodic table of chemical elements.

3. MECHANICAL PROPERTIES

This chapter is devoted to understanding the macroscopic behavior of a solid that is subject to mechanical forces, such as applied stress or atmospheric and/or hydrostatic pressure. This understanding will be based on the microscopic point of view starting from the atomic structure of solids. For conceptual clarity, the treatment is restricted to isotropic materials.[1] Our interest in these properties arise from their importance in ***materials science and engineering***. Many advances in our technology and material-driven society occur because of development of better materials.

We start by introducing the fundamental definitions for the most relevant quantities of interest. The first of such quantity is the applied ***stress*** on a solid σ, which is defined as the force F per area A perpendicular to the applied force (Fig. 3.1a).[2] If the applied force pushes the solid inwards, we speak of ***compressive stress***, and in the opposite case we speak of ***tensile stress***.

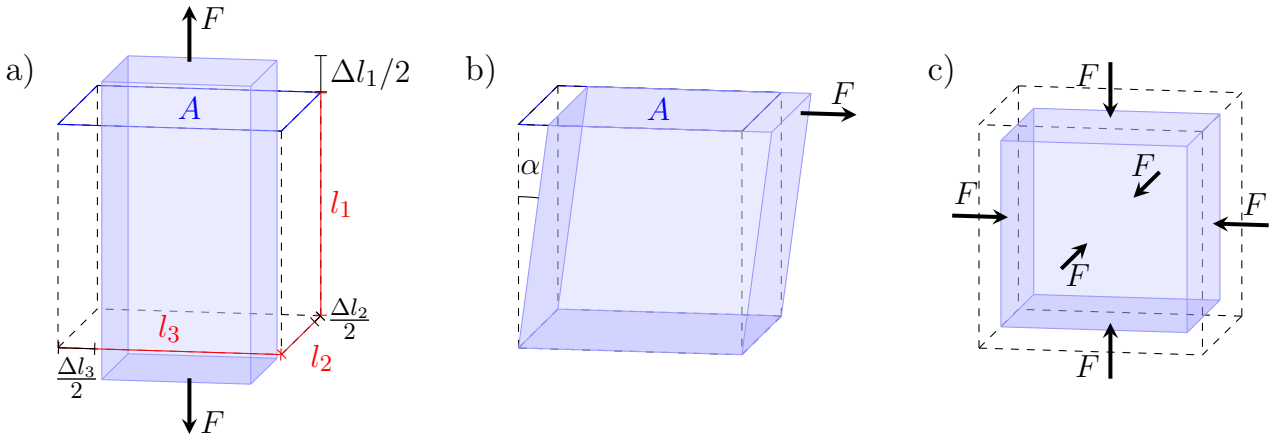

Fig. 3.1: Geometries for material tensile stress, shear and compressive stress, and the respective strains. a) Tensile stress σ is defined as the ratio of the pulling force F and the cross-sectional area A ($\sigma = F/A$). Tensile stress results in material strain ϵ along the directions 1, 2 and 3, where index 1 denotes the direction of the applied stress. b) Shear stress τ results in a material deformation characterized by the shear angle α. c) Hydrostatic pressure results in a compressive uniform stress and consequent strain.

Solids under stress will deform. This deformation is called ***strain*** ϵ and is defined as the relative change of the length of the solid $\epsilon = \Delta l/l$.[3] In most of cases, the solid under stress is strained in many directions. In order to remove the related ambiguity, we use

[1] Generalization of the mechanical properties of solids to include anisotropic materials is mathematically non-trivial. Relevant quantities become ***tensor fields*** extending quickly the treatment beyond the scope of an introductory course.

[2] Note that stress σ is defined similarly to the familiar concept of pressure and has the same units, i.e. Nm^{-2} or Pa.

[3] Strain is in principle unitless, but many engineers find the convention of using units such as μm/m highly illustrative. This convention is similar to the one used with the coefficients for thermal expansion.

notation $\epsilon_i = \Delta l_i / l_i$ where the index i corresponds to the direction with $i = 1$ reserved for the direction parallel to the applied stress. Solids often contract along the direction of the applied compressive stress.

Normal stress is defined through the force applied perpendicular to the area. Similarly, we can define ***shear stress*** τ as the applied force F per area A, where F is now applied tangentially to A (Fig. 3.1b). Again, the solid is deformed which can be conveniently quantified also as a deformation angle α (Fig. 3.1).

Final situation we consider here is the case when the solid is exposed to uniform compressive stress from all its sides, e.g. due to hydrostatic pressure.[4] We note that other mechanical deformations do exist, but will not be covered during this course.[5]

3.1 Stress–strain Curve

We continue our discussion by considering typical responses of solids under applied stress. Most of materials respond similar to what shown in Fig. 3.2, that shows the material stress as a function of strain. This kind of curve is known as the ***stress–strain curve***. The convention of presenting the strain as the x-axis of the curve might seem odd at first, and arises from the fact how typical experiments are performed in order to extract the information used to plot the stress–strain curve of the material.[6]

Looking at Fig. 3.2, we note that the material's response consists of different phases. For small enough strain, typically much smaller than 1%, the material undergoes ***elastic deformation***. Under this regime, the material returns to its original shape after the applied stress vanishes. In this regime, the relationship between applied stress and the material strain is linear allowing to form simple relationships for the material's elastic constants.

If the applied stress exceeds the so-called yield stress σ_Y, resulting in corresponding yield strain ϵ_Y, material's deformation becomes ***plastic***. This means that the deformation becomes permanent. In this case the material will not anymore return to its original shape once the applied stress is decreased. Finally, at large enough values of applied stress every material will ***fracture***.

---

[4]Relevant scenarios include ***Deep-submergence vehicles***, ***high-altitude balloon*** or maybe even ***satellites***.
[5]An interested reader is referred to the excellent book '*Introduction to Solid State Physics*' by C. Kittel.
[6]Commonly used method to measure the stress–straing curve of a material is to perform ***tensile testing***.

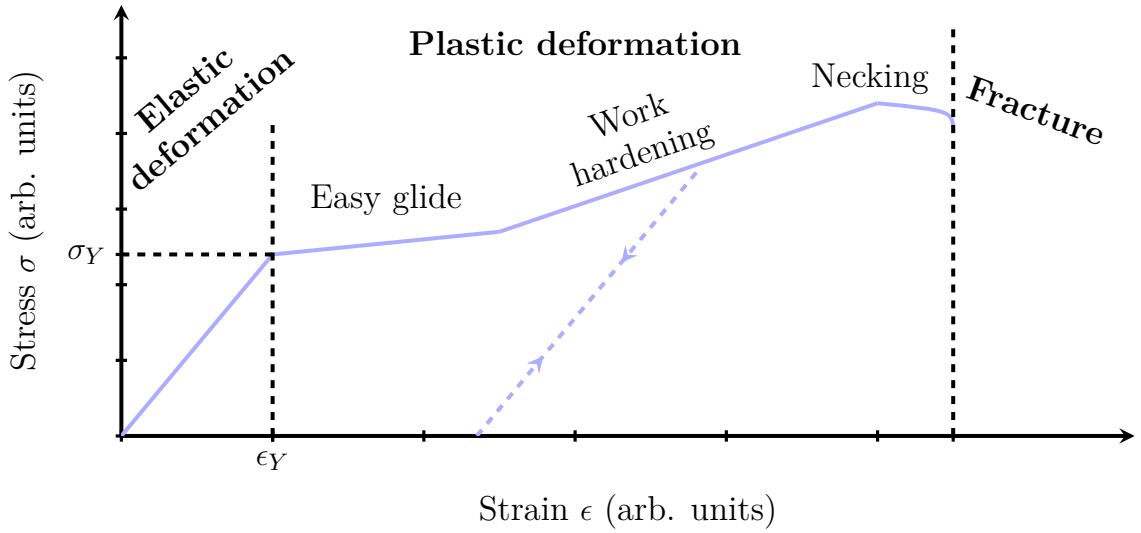

Fig. 3.2: Representative stress–strain curve. Yield strain ϵ_Y and yield stress σ_Y denote the ending position of the region of elastic deformation. The dashed line denotes the material's response to applied stress after the strain has initially reached the work hardening region.

In general, the region of elastic deformation is usually quite small in materials. On the contrary, the region of possible plastic deformation varies widely between materials. At one end, we have materials such as glass or cast iron, that will fracture immediately after the stress exceeds the value of yield stress σ_Y, i.e., at the end of the region for elastic deformation. Such materials are called ***brittle***. Other materials, such as most metals, exhibit a region of plastic deformation, and are called ***ductile***. In the following sections, we extend our treatment further on these properties.

3.2 Elastic Deformation

As mentioned earlier the elastic regime of deformation is usually very narrow. Despite this fact, it is in practice the regime of most interest in terms of materials science applications. Most applications require that the material remains elastic and shows no signs of wear and tear. In the following, we look into the technical details regarding elastic deformations. First we treat the problem from the macroscopic point of view by introducing to readers a few macroscopic ***elastic constants***. After this we will discuss how things look microscopically connecting our treatment to the earlier concepts regarding bonding of solids.

3.2.1 Macroscopic Picture

3.2.1.1 Elastic Constants

Experimental evidence suggests that the relation between the applied stress σ and the resulting strain ϵ is linear. This allows us to define their ratio as ***Young's modulus*** as

$$Y = \frac{\sigma}{\epsilon} = \frac{F}{A}\frac{l}{\Delta l}\,. \tag{3.2.1}$$

Since strain ϵ is unitless, Young's modulus Y has the same unit as stress. Looking at the stress–strain curve of a material (see Fig. 3.2), Young's modulus Y can be interpreted as the slope of the curve in the region of elastic deformation. Similar to Young's modulus, the modulus of rigidity G, also known as the ***shear modulus***, is defined as[7]

$$G = \frac{\tau}{\alpha}\,, \tag{3.2.2}$$

and describes the material's inclination to shear.

The third type of force (or combination of forces) that can deform a solid is a uniform compressive force, arising e.g. from the hydrostatic pressure. The resulting deformation of the solid can be described by ***bulk modulus*** K that is defined through the relation

$$K = -V\frac{\mathrm{d}P}{\mathrm{d}V}\,, \tag{3.2.3}$$

where V is the volume of the solid and dP/dV is the derivative of pressure with respect to the volume. The minus sign in the definition of the bulk modulus K ensures that a solid described by a positive K corresponds to a solid that compresses upon compressive force.[8] Vast majority of solids are compressed when pressed by a force, but some materials do exist that instead expand. In order to comply with this rule of thumb or common sense, most of mechanical properties of solids are defined such as way that positive-valued moduli correspond to compressive type of deformation.

3.2.1.2 Poisson's Ratio

In this section we will generalize the above treatment for the elastic deformations that solids undergo when stressed. We start by making a ***gedanken-experiment*** that there exists a solid material that is totally **incompressible**, meaning that the volume

[7] Engineers may use an alternative definition for the shear modulus: $G = \frac{\tau}{\gamma} = \frac{F/A}{\Delta x/l} = \frac{F}{A\tan\alpha}$.

[8] Materials with negative bulk modulus are rare, but not non-existing. Such properties could be realized e.g. using ***metamaterials***.

of the material would remain constant upon stress. When stress would be applied from two facing sides to such a material, assuming normal (positive) Young's modulus Y the length of the material along that direction would decrease. Because the total volume should remain constant, the logical conclusion is that the other sides of the material, the ones perpendicular to the applied stress, would in this case expand. It turns out, that many materials do behave as outlined, although not all behave as perfectly incompressible.

In order to quantify this effect in a solid, essentially the material's change of dimensions, we define the ***Poisson's ratio*** ν through equations[9]

$$\frac{\Delta l_2}{l_2} = \frac{\Delta l_3}{l_3} = -\upsilon \frac{\Delta l_1}{l_1} = -\nu \epsilon_1 \,. \tag{3.2.4}$$

Again, the minus sign ensures that a positive Poisson's ratio ν complies with the common sense. That is, for positive ν the solid's perpendicular sides expand during compressive stress (vice versa upon tensile stress). In this context, it is useful to introduce terminology that calls the fractional changes in the transverse directions, i.e. terms $\Delta l_2/l_2$ and $\Delta l_3/l_3$ as lateral or ***transverse strain***, while ratio $\Delta l_1/l_1$ is also known as axial or longitudinal strain (see Fig. 3.1a) . Using this terminology, and by re-organizing the terms in Eq. (3.2.4), we see that the Poisson's ratio can be understood as the (negative of the) ratio of transverse strain to axial strain.

The Poisson's ratio ν can be easily shown to be limited to values between -1 and 0.5. Because for most materials the Poisson's ratio ranges from 0 to 0.4, we will only calculate the origin of the upper limit. Let us consider a cubic piece of a solid with volume $V = l_1 l_2 l_3$ where l_i are the side lengths of the volume. We place the solid under axial stress turning the original volume V into

$$\begin{aligned} V' &= (l_1 + \Delta l_1)\,(l_2 + \Delta l_2)\,(l_3 + \Delta l_3) \\ &= l_1 l_2 l_3 + l_1 l_2 \Delta l_3 + l_1 \Delta l_2 l_3 + l_1 \Delta l_2 \Delta l_3 \\ &\quad + \Delta l_1 l_2 l_3 + \Delta l_1 l_2 \Delta l_3 + \Delta l_1 \Delta l_2 l_3 + \Delta l_1 \Delta l_2 \Delta l_3 \,. \end{aligned} \tag{3.2.5}$$

The strain terms Δl_i are far smaller than the actual dimensions l_i of the volume. In consequence, terms in the sum containing two or more strain terms can be neglected

[9]Note here that we restricted our treatment to isotropic materials, which in this case implies the equality between the fractional changes $\Delta l_2/l_2 = \Delta l_3/l_3$. Generalization to even anisotropic, not to mention ***orthotopic materials*** is beyond our scope.

resulting in

$$V' \approx l_1 l_2 l_3 + \Delta l_1 l_2 l_3 + l_1 \Delta l_2 l_3 + l_1 l_2 \Delta l_3 \,. \tag{3.2.6}$$

Because $V' = V + \Delta V$, and using the Eq. (3.2.4), the change in the volume V becomes

$$\begin{aligned} \Delta V &\approx \Delta l_1 l_2 l_3 + l_1 \Delta l_2 l_3 + l_1 l_2 \Delta l_3 \\ &= \Delta l_1 l_2 l_3 + l_1 \left(-v \frac{\Delta l_1}{l_1} l_2 \right) l_3 + l_1 l_2 \left(-v \frac{\Delta l_1}{l_1} l_3 \right) \\ &= (1 - 2v) \Delta l_1 l_2 l_3 \,. \end{aligned} \tag{3.2.7}$$

The upper bound for the Poisson's ratio $\nu = 0.5$ is now seen to correspond to ideal incompressible solids ($\Delta V = 0$). An exemplary material that is almost incompressible is rubber. On the other side, cork is a material that has ν close to zero.

3.2.1.3 Relations between Elastic Constants

As we've seen above, the macroscopic elastic constants can be derived from a few fundamental elastic properties. Subsequently, it is possible to find relations that link these properties together. For example, the Young's modulus Y, the shear modulus G and the Poisson's ratio ν are related by

$$G = \frac{Y}{2(1+v)} \,. \tag{3.2.8}$$

Similarly, the Young's modulus Y, the bulk modulus K and the Poisson's ratio ν are approximately inter-related by

$$\nu \approx \frac{1}{2} - \frac{Y}{6K} \,, \tag{3.2.9}$$

which is known as the ***Lame's relation***. Looking at these relations, we see that the elastic constants for a given material should all have the same order of magnitude. This rule of thumb is often useful when the elastic properties of materials are used in everyday applications.

3.2.2 Microscopic Description

Next, we consider the elastic deformation of a solid from the microscopic point of view, and see that such a deformation can be understood as changing inter-atomic distances. Recalling the inter-atomic potential ϕ between two atoms (Fig. **??**), we see that the equilibrium distance between two atoms corresponds to the minimum in the inter-atomic

potential ϕ. As we consider the solid to be in equilibrium,[10] the electrostatic approximation holds and the resulting conservative force field between the two atoms at this distance r is zero.[11]

When a compressive stress is applied on the solid, the forces affecting the atoms increase and a new equilibrium distance is found. In essence, the distance between the atoms r decreases. Because the atoms return to their original equilibrium distance once the applied stress is released, the deformation is elastic.

In order to understand why the stress–strain relationship is linear, we consider the details of the potential ϕ more carefully. In order to do so, we assume that the potential ϕ is well behaving and can be expanded as a Taylor series close to the equilibrium distance $r \approx a$ resulting in

$$\phi(r) \approx \phi(a) + \frac{\phi'(a)}{1!}(r-a) + \frac{\phi''(a)}{2!}(r-a)^2 + \frac{\phi'''(a)}{3!}(r-a)^3 + \dots \tag{3.2.10}$$

Next, we estimate the roles of different terms for the resulting force that is proportional to the gradient of the potential ($\mathbf{F} = -\nabla\phi$). The first term is a constant and vanished upon derivation over r. The second term vanishes, because the slope of the potential is zero, implying a vanishing $\phi'(a)$ term. It is therefore mainly the third term that is responsible for the elastic behavior of the solid.[12] Furthermore, that term results in a linear inter-atomic force.

More insights into elastic deformations can be found by looking at the functional form of the third term. Because the second derivative of the potential $\phi''(a)$ is related to the curvature of the potential, we see that the elastic behavior of solids should be affected by changes in the curvature of the potential well, motivating us to look more closely into the potentials.

We continue by looking at the values of Young's modulus Y for chemical elements of the periodic table (Fig. 3.3). It becomes clear, that the type of bonding strongly affects the values of Y. Particularly, the transition metals (highlighted in red) in general have a higher Y than simple metals. This trend agrees with the fact that their bonding

[10] Right now We are interested in solids that are mechanically stable.

[11] Conservative force fields are equal to the negative of the gradient of the potential, i.e. $\mathbf{F} = -\nabla\phi$.

[12] Higher-order terms are in general non-zero, but their strengths are often quite weak and can be neglected. Their role becomes non-negligible only at relatively high values of applied stress and only for certain materials, prompting to use ***finite strain theory***.

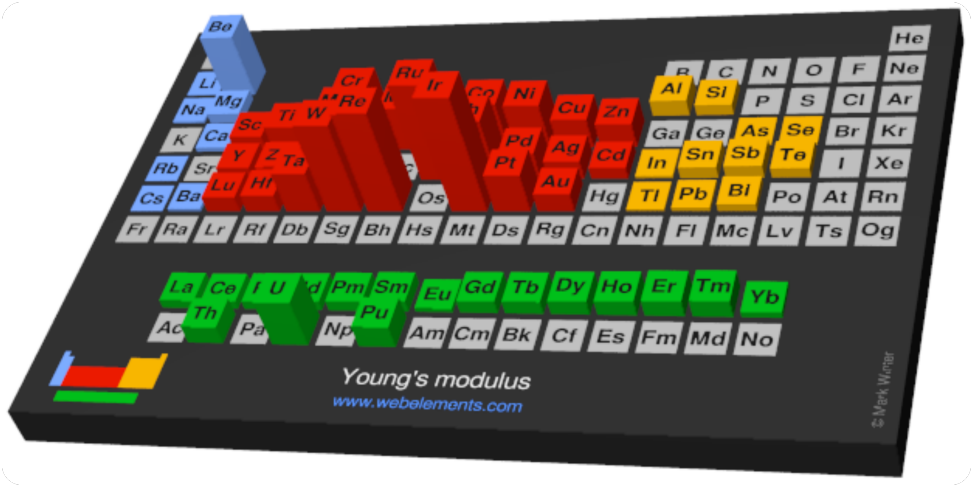

Fig. 3.3: Young's moduli for chemical elements shown in the periodic table. From webelements.com.

strengths are often quite high due to localized d electrons. A useful rule of thumb is also found by comparing the values of Y with the melting points of elements. Often these parameters are strongly correlated, which fact agrees with our intuition based on the shape of the potentials.

Covalently bonded solids show more dispersion of Y values. This general behavior can be understood by realizing that many different kinds of covalent bonding play a role in formation of such solids. A prime example material is graphite, that is bonded by two kinds of bonds. Consequently, graphite is also a good example of an anisotropic solid, whose elastic properties depend on the direction.[13] The stronger bonds are the sp^2 bonds linking the carbon atoms together to form the hexagonal plane. Considerably weaker molecular bonding occurs in the plane that is perpendicular to the hexagonal plane. As a consequence, the Young's modulus of graphite is highly anisotropic (Fig. 3.4).

In addition to crystalline solids, many solids are amorphous. It is worth to notice that the above treatment does not always adequately describe the elastic properties of such materials, because other mechanisms need to be taken into account. For example, ***polymers*** form a large class of materials, that in general have quite low values of Young's moduli Y. This general trend can be understood by noting that the large molecules forming the polymers are in general bent in the solid. Once the polymer is

[13]Note that in our earlier treatment, we specifically restricted our scope to isotropic solids in order to keep treatment conceptually simple.

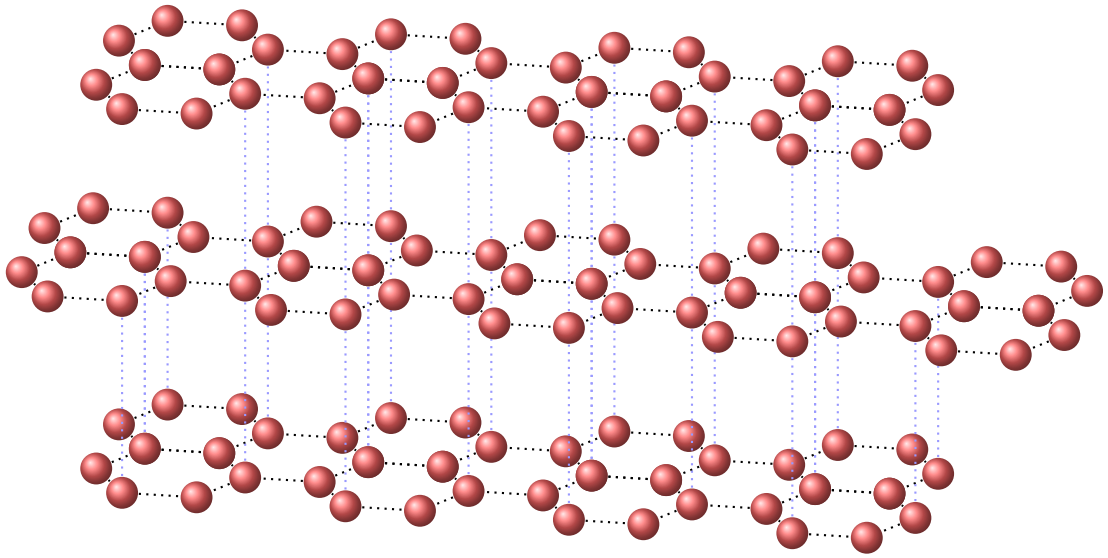

Fig. 3.4: Graphite is an extremely anisotropic solid. The bonds (black dotted lines) associated with the planar honeycomb structure are strong, while the inter-planar bonds (blue dotted lines) are weak.

stressed, these bent molecules tend to first straighten out. Only after this unfolding of molecules has occurred, the applied stress would start to noticeably affect also the inter-atomic distances causing an elastic deformation as described earlier.

We also note that it is important to make the distinction between the curvature of the potential ϕ and the actual depth of the potential well, which gives rise to the cohesive energy of the solid. It is specifically the curvature of the potential ϕ that is expected to affect the elastic properties of (crystalline) solids. While in most cases a high curvature of ϕ implies a deep potential well, this does not always need to be the case. A good

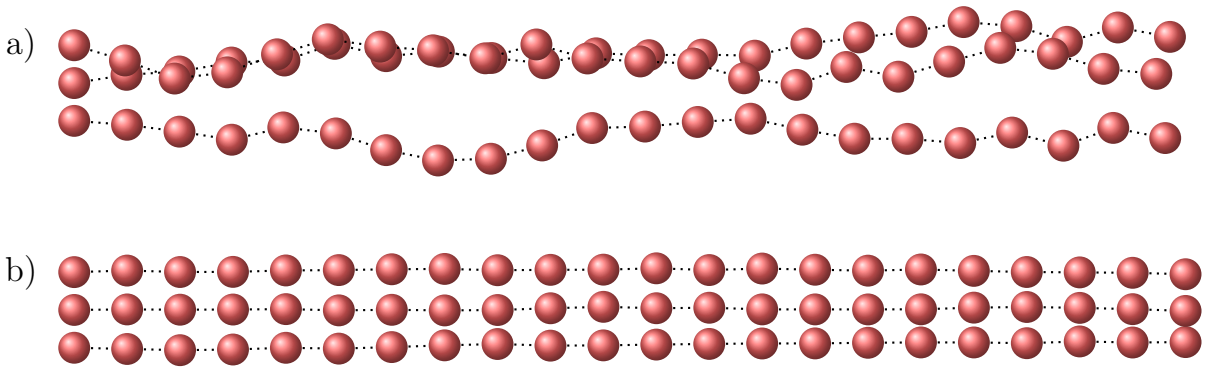

Fig. 3.5: Mechanism behind elastic deformation of polymers. Polymers tend to naturally form a) bent configurations, that can be b) straightened relatively easily by applying tensile stress. This phenomena gives rise to the general trend that the Young's moduli Y of polymers tend to be moderately low.

example solid is diamond, which is associated with both a very high cohesive energy and a high value of Y.[14]

Finally, we take a moment to outline how our treatment could be generalized. The obvious extension would be to extend our scope to anisotropic solids using an approach known as ***infinitesimal strain theory***. This generalization is conceptually quite straightforward. Unfortunately, the mathematical treatment becomes somewhat cumbersome because the relevant quantities stress σ and strain ϵ become tensors.[15] Complicating things further, the subsequent derivation of the material parameters, such as the Young's modulus Y, requires use of tensor calculus. As a result, the material parameters become ***tensors of even higher order***.

For our scope, it is far more interesting to consider what kind of physics results if we extend our treatment to time-varying elastic deformations. Particularly, we are interested in fast and oscillating deformations, that could be driven by time-varying external forces caused, for example, by incident electromagnetic fields. Such lattice vibrations are also known as ***phonons***, and will be introduced and treated more carefully later on in Chapters. We only note here, that phonons often play an important role in the thermal and electrical conductivities of solids.

3.3 Plastic Deformation

Next we consider in more depth the regime of plastic deformation, that is visible in the stress–strain curves of some solids.[16] We will try our best to link the region of plastic deformation and respective phenomena to the microscopic description of solids, but notice that such phenomena cannot be completely understood by only considering perfectly crystalline solids. Therefore, this section motivates the introduction of ***lattice defects***, which are later on found to be an important concept in more detailed understanding of solid-state physics.

3.3.1 Estimate of the Yield Stress

As we learned earlier, the yield stress σ_Y is a point of no return for a solid. Once the applied stress becomes stronger than the yield stress σ_Y, the solid no longer returns to

[14] Diamond is in many ways a special material. It has a quite high refractive index n and it is highly thermally conductive, while it still is a dielectric.

[15] In essence, stress σ_{ij} and strain ϵ_{ij} would be expressed as quantities closely resembling 3×3 matrices.

[16] Note here that not all materials exhibit plastic deformations. Such brittle materials fracture after the yield stress σ_Y is exceeded.

its original shape after the stress is decreased.

It can be intuitively understood that the yield stress values σ_Y of different materials are of considerable practical interest. For conceptual clarity, we provide next an order-of-magnitude estimate for the yield shear stress τ_Y, that would correspond to a case of perfectly crystalline solid. We start our calculation by considering an hcp lattice, as shown in Fig. 3.6a. By applying a shear stress τ on the lattice, the lattice is deformed (Fig. 3.6b). For small enough shear stress value the shear angle α is close to the ratio of layer distance a the displacement of layers x as given by

$$\alpha = \tan^{-1}\left(\frac{x}{a}\right) \approx \frac{x}{a}\,. \tag{3.3.11}$$

By inserting the above relation to the earlier definition of shear modulus given by Eq. (3.2.2), we can relate the shear modulus τ with the microscopic displacements:

$$\tau = G\alpha \approx \frac{Gx}{a}\,. \tag{3.3.12}$$

For small values of shear stress τ and subsequent small changes in the deformation angle α, the atoms will return to their equilibrium positions once τ vanishes. In this case, the response of the solid remains elastic.

For larger values of shear stress τ, the adjacent atomic layers start gliding over each other giving rise to the easy-glide region often seen in the stress–strain curves (Fig. 3.2). This case is illustrated in Fig. 3.6c. Another stable configuration for the lattice is found once the displacement of layers x reaches equals the lattice constant b of the solid (Fig. 3.6d). Therefore, a simple estimate for the yield shear stress τ_Y is reached by inserting $x = b$ into Eq. (3.3.12)[17]

$$\tau_Y \approx \frac{Gb}{a}\,. \tag{3.3.13}$$

Admittedly this derivation is simplistic. However, it should provide us a reasonable order-of-magnitude estimate. Surprisingly, several orders of magnitude discrepancies are found when this estimate is compared against experiments. For example, the measured Young's modulus Y for aluminum is close to 70 GPa while its yield stress τ_Y is only roughly 30 MPa. The main reason for this discrepancy is the fact that the calculation

[17]Note that this derivation differs slightly from the one shown in P. Hofmann's book. However, since we're only looking for a rough order-of-estimate, our logically sound treatment is adequate.

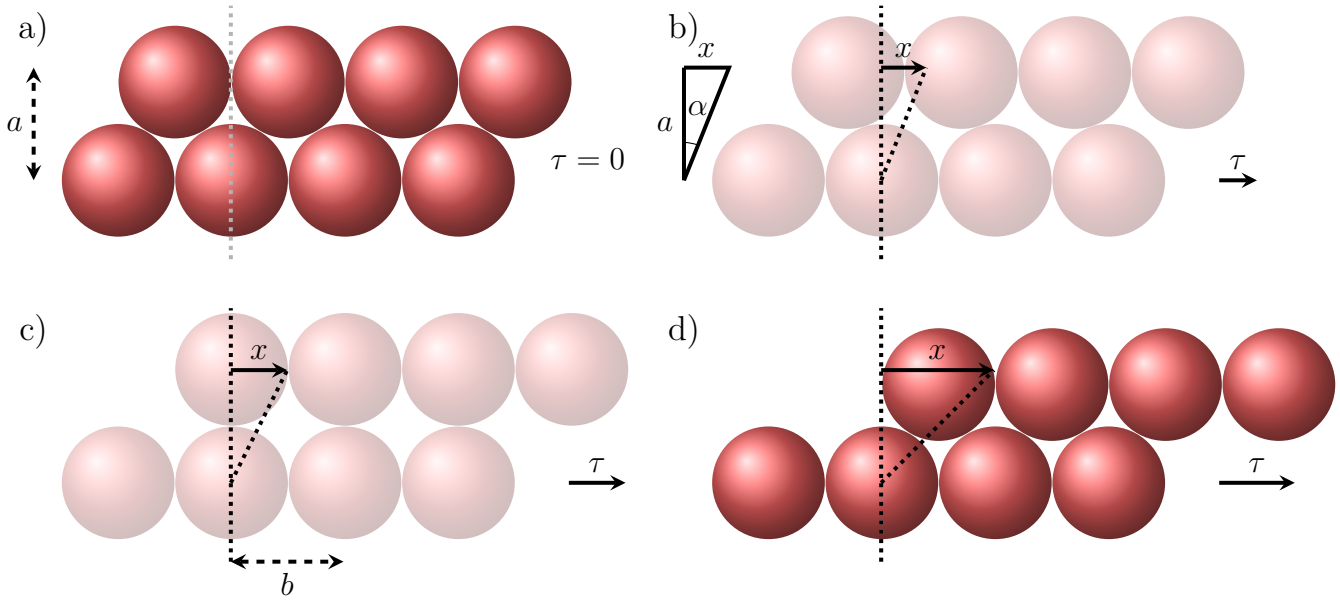

Fig. 3.6: Explain the variables used in the figure. Interatomic distance b. 'Estimate of the yield stress for shearing a solid. (a) Atoms in equilibrium position. (b) Distortion for a small shear stress. (c) Meta-stable equilibrium. (d) New stable equilibrium for the sheared solid.'

assumed a perfect crystalline solid, motivating our next discussion of lattice defects. In other words, in order to really understand the microscopic origins of the plastic deformation region, we need to consider lattice defects.

3.3.2 Point Defects and Dislocations

First we note that in the broader context of solid-state physics and materials science, the lattice defects, also called crystal imperfections, are an important research topic of their own. For example, defects are paramount in the proper behavior of photonic materials/devices such as photonic crystal waveguides and cavities, quantum dots and wells, and other nanoscale emitters. In essence, defects are in many cases more than just minor perturbations of the perfect crystals. Next, we will introduce the related basic terminology and discuss the relevance of defects for explaining the mechanical properties. Later on during the course, we will also learn that defects can be important also when the electrical and thermal conductivities of materials are considered.

Lattice defects can be broadly divided into two categories. The first category are known as ***point defects*** that can be understood as localized point-like deviations from the ideal crystalline lattice. Defects that are no longer point-like, but rather extended

perturbations are known as ***extended defects***, that can be further divided into line defects, planar defects and bulk defects.

Next, we quickly introduce some of the most common types of point-like defects, which have been schematically illustrated in Fig. 3.7. A ***vacancy defect*** is a term for a lattice site that completely lacks an atom. When a vacancy defect is present, the missing atom is often re-located to a nearby or far-away location. When the atom is far away in a volume of lattice that looks perfectly crystalline, all possible lattice sites are occupied. In this case, the atom will find a location in the lattice that is normally unoccupied, and we speak of ***interstitial defects***. When the interstitial atom rests on a position to the vacancy, the vacancy and the interstitial defect form a so-called ***Frenkel pair***. Interstitial defects are often used to improve the mechanical properties of ***alloys***, where perhaps the most known example is ***carbon steel***.

In addition to the above point defects, atoms in a perfect lattice can also be replaced by atoms that are not identical to the atom normally occupying the lattice site. In this case, we speak of ***substitutional defects***. We will later learn in more detail how substitutional defects can be conventiently used to change the properties of semiconductors, particularly their electrical conductivities. In this case, we often speak of (semiconductor) ***doping***.

Now we are ready to move to line and planar lattice defects. Among the most important of such defects are ***dislocations*** and ***grain boundaries***. Dislocations are line-type defects where an abrupt irregularity exists in the lattice. Such an irregularity may be for example an extra sheet of atoms, as has been shown in Fig. 3.9. Dislocations are in practice very important in the formation of the plastic deformation regime of the solid, because such defects can move along the ***slip plane*** of the solid.[18]

We end our short introduction to the wonders of imperfect crystals by elaborating more on grain boundaries, which are essentially interfaces between two grains of crystalline solids (See Fig. 3.8). Grain boundaries are present in ***polycrystalline*** solids where they often decrease the electrical and thermal conductivities of such materials. Therefore, grain boundaries can be used to a degree to engineer such material properties.[19]

[18] ***Slip*** is a technical term that describes the displacement of part of a crystalline with respect to another part along a specific slipping plane.

[19] Recent and continuous advances in the fabrication methods to make large-area monocrystalline 2D mate-

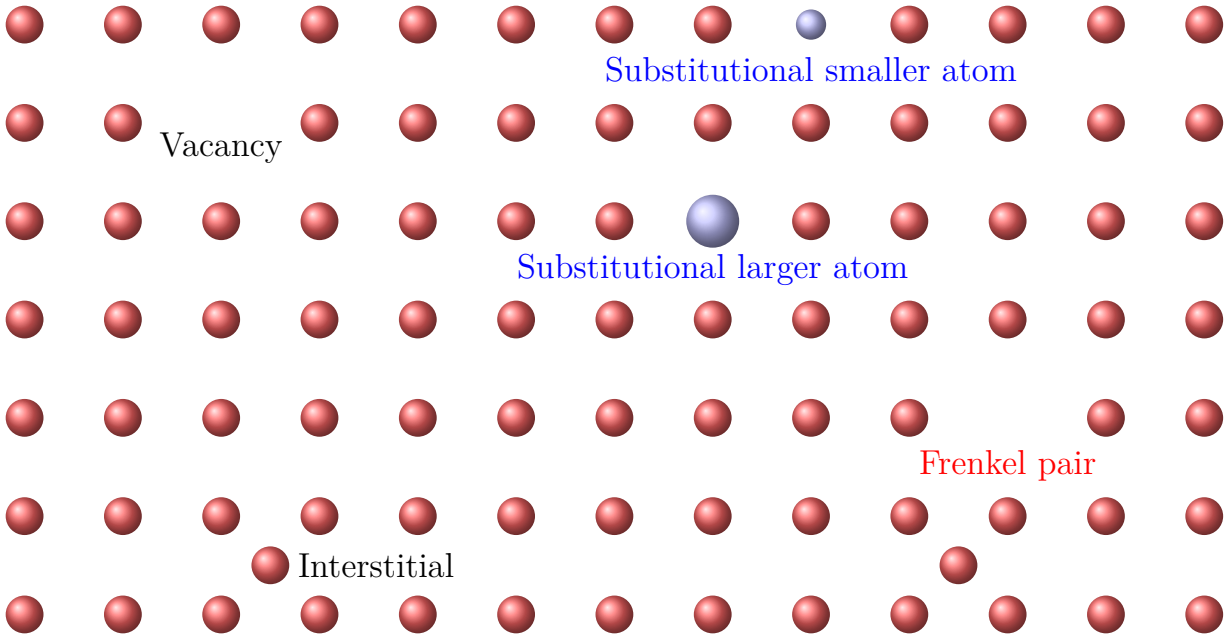

Fig. 3.7: Illustration of the most commont point defects occurring in a monatomic solid. Frenkel pair is a combination of a vacancy defect and an interstitial atom nearby the vacancy.

Furthermore, we note that grain boundaries and grains in general will play an important role when we turn our focus to understand magnetism and magnetic materials. In addition, they have a role in phenomena of ***creep***, where a solid is slowly displaced or deformed. Prime examples of creep include the unfortunately not-so-slow movements of ice glaciers during the ***Anthropocene***.[20]

3.3.3 The Role of Defects in Plastic Deformation

We are now ready to proceed in our efforts to understand the region of plastic deformation from the microscopic point of view, which we hold so dear during this course. Particularly, we will soon understand the reason for the surprising discrepancy between the estimated and measured yield shear stresses τ_Y.

We start by considering a crystalline solid where an edge dislocation, and subsequent slip plane, exists as is shown in Fig. 3.10. Upon application of shear stress, it is now possible that the gliding of atoms occurs one lattice column at a time (Fig. 3.10a–

rials, such as graphene, are an interesting practical example of the importance of grain boundaries. The first demonstrations of 2D graphene flakes, prepared using exfoliation of graphite, in practice demonstrated graphene flakes of few microns in sizes. Nowadays, monocrystalline flakes with cubic centimetre areas are routinely fabricated. This is big news for graphene electronics.

[20] Despite common knowledge, the deformed sagged shapes of old glass windows may not be due to creep at all.

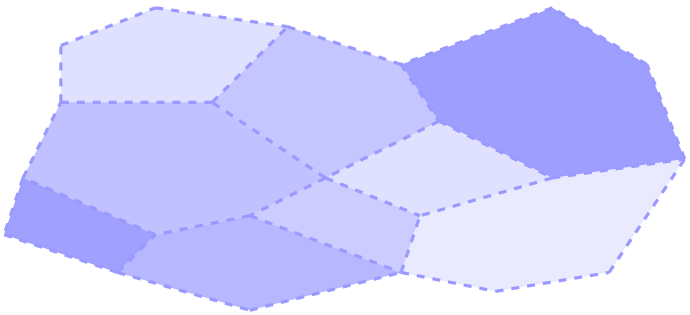

Fig. 3.8: In practice, most of solids are not monocrystalline, but rather polycrystalline materials. The sizes of crystallites vary, but are often of the order of few micrometers in length. Each monocrystalline grain (blue polygons) is separated from the adjacent grains by the grain boundaries (blue dashed lines).

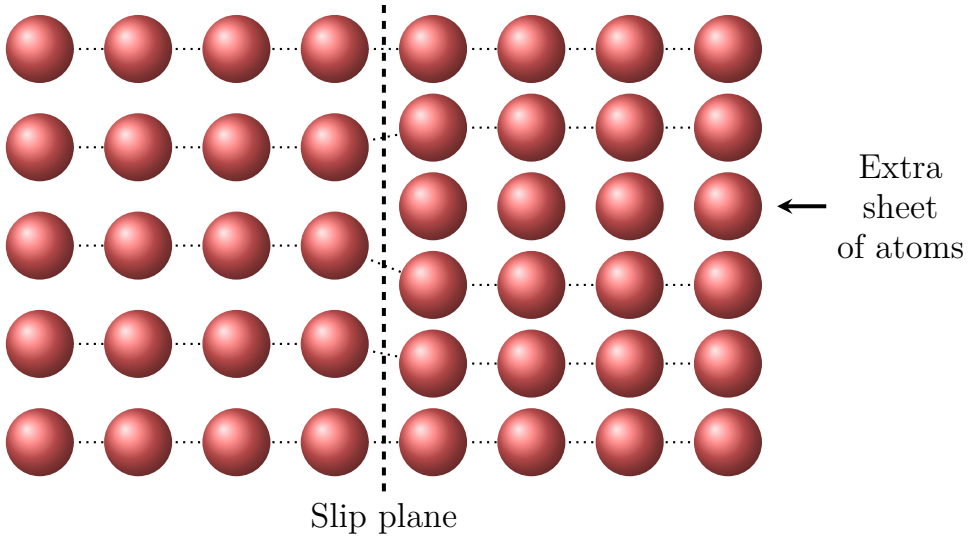

Fig. 3.9: An edge dislocation may form by addition of an extra sheet of atoms into a part of the lattice. The edge dislocation is associated with a slip plane, and can move along the slip plane.

c). Importantly, this process can occur at considerably lower values of shear stress τ, because considerably less atoms need to relocate themselves compared to the case of perfectly crystalline solid.

Interestingly, edge dislocations are quite abundant in real materials, making the above process quite probable and in fact the dominant mechanism that explains the experimentally measured low values of yield shear stress τ_Y. This yield shear stress value simply corresponds to a stress at which single (or only very few nearby the dislocation line) atoms start breaking and reforming their bonds with the nearby atoms. Alternatively, one could picture this process as the movement of the dislocations.

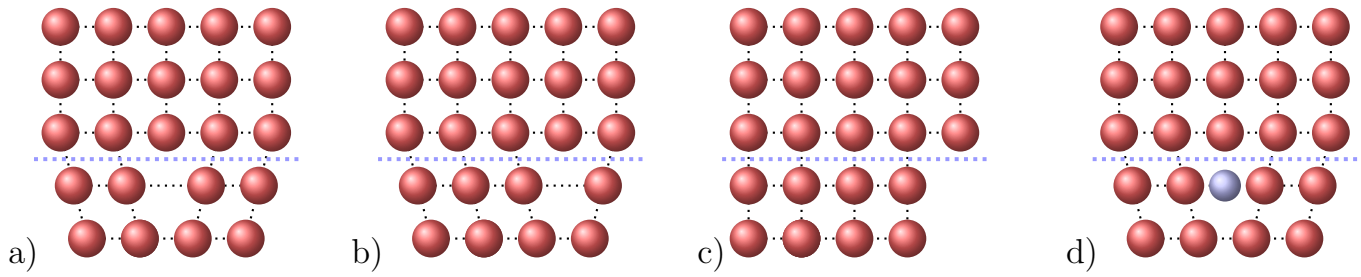

Fig. 3.10: a)–c) Shear-induced movement of an edge dislocation through the solid one row at a time. d) A point defect (interstitial blue atom) can freeze out the dislocation. Dotted blue line highlights the slip plane of the solid.

Understanding the microscopic origins behind shear stress τ is both intellectually rewarding and useful, because it provides us a methodology to increase the yield shear stress τ_Y of materials. We can now correctly reason, that by hindering the movement of dislocations we could increase the τ_Y values. In practice, the movement of dislocations is made more difficult by introducing new defects, such as interstitional atoms, into the solid (see Fig. 3.10d). When the vacancy defect is filled up by an interstitional defect, the dislocation cannot propagate as freely anymore.

This is also an important mechanism that describes how real materials, such as the already-mentioned carbon steel, are hardened in practice. Interestingly, an introduction of even a very small amount of impurities is often enough to visibly affect the mechanical properties of solids. This is due to the fact that impurities are not homogeneously dispersed within the solid, but instead strongly favor the vacancies present in dislocations.

The final region to consider in the plastic deformation regime of the stress–strain curve is the region of ***work hardening***, that occurs after the easy-glide region. In this region the curve becomes steeper, and when the applied stress is released, the material no longer returns exactly to its original length. The amount of strain that the solid does recover is known as the elastic strain ϵ_e, while the non-recoverable part is known as the plastic strain ϵ_p (Fig. 3.11). However, it is important to notice that the solid exhibits now a new yield stress value of σ'_Y that is larger than the original yield stress σ_Y of the solid. Therefore, the material can be hardened by stressing the material up to the region of work hardening.

Before moving on to the final part of the stress–strain curve, we briefly mention that the

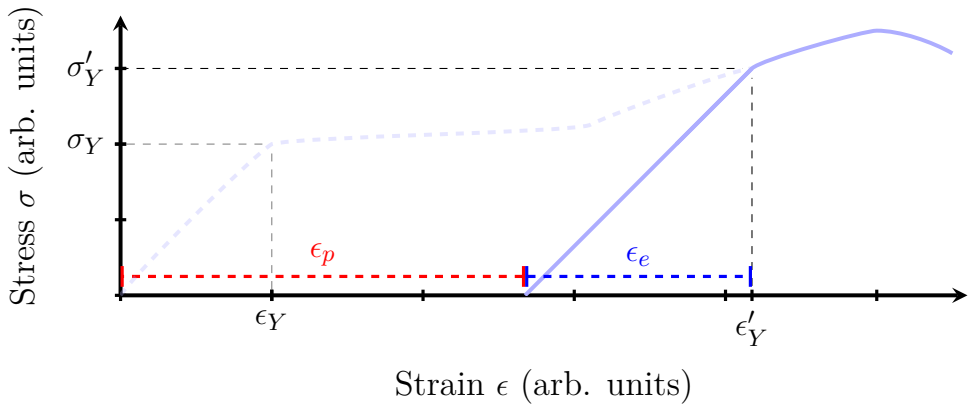

Fig. 3.11: Work hardening of a material. The original stress–strain curve of the material is shown in blue dotted line. Upon application of strong enough stress levels ($\sigma > \sigma_Y$), the material no longer returns to its original shape after the release of the applied stress. In this scenario, the non-recoverable part of the material strain ϵ is known as the plastic strain ϵ_p (red dotted line), while the elastic part of the strain is known as elastic strain ϵ_e (blue dotted line). Consequently, the original yield stress σ_Y of the material increases to a value σ'_Y.

temperature of the solid strongly affects its mechanical properties. Recalling the basic principles of thermodynamics and statistical physics, we can argument that generation and movement of defects becomes more important. The system is associated with a larger amount of energy, that particularly helps in overcoming the energy barriers associated with movements of defects. By elevating temperatures high enough, it is even possible to turn some brittle materials, such as many glasses, into ductile materials.

3.4 Fracture

The stress–strain curves end when the material ***fractures***. This occurs when the applied stress exceeds the ***ultimate tensile strength*** F_{tu}, which is the maximum amount of stress the material can withstand without breaking. Right before fracturing, the slope of the curve might even turn negative due to a process called ***necking***. At this stage, the applied axial tensile stress has considerably tapered the material somewhere between the points of applied stress. Because the stress σ is defined as the force F per cross-sectional area A, this phenomena results in increase of the local stress, that may surpass in strength the amount of applied stress. It is therefore important to distinguish between a so-called ***engineering stress***, which is equal to the applied stress, and the local stress which is also known as the ***true stress***, that is the instantaneous stress felt by the material in a given location. Similarly, two resulting types of strain and stress–strain curves exist. Unless otherwise specified, engineering stress–strain curves

are plotted, since that data is the one that can be straightforwardly measured using ***tensile testing***.

Necking can therefore be understood as a measurement artifact, where the decrease in engineering stress actually corresponds to a situation where the true stress at the tapered section of the material exceeds the (ultimate) tensile strength F_{tu}. In this area a heterogeneous neck is formed tapering the material further, which in overall results in a positive feedback of the microscopic fracturing process. Finally, the amplified process progressively results in the macroscopic fracture of the material.

The above considerations apply well to the case of ductile materials. However, the fracturing mechanism is different for brittle materials that were characterized by the lack of plastic deformation. This fracture mechanism is known as ***brittle fracture***. The microscopic details of brittle fracture are beyond the scope of introductory course, but it is worth to note that similar principles as introduced earlier can be used to understood the phenomena. The material exhibits planar defects, such as cracks, perpendicular to the direction of applied stress. The true local stress fields near these defects may be considerably stronger than the value of the applied engineering stress. As a consequence, a local fracture is formed. Again, these fracturing processes are associated with a positive feedback resulting in progressive fracturing phenomena. Exemplary cases of brittle fracturing phenomena include formation of ***crevasses*** in ice glaciers.[21]

[21] It is worth to note that pressure can also change the mechanical properties of materials. For example only the top layers of glaciers are rigid, forming so-called ***fracture zones***. The regions below these top layers are under heavy pressure, that subsequently turns the deeper regions of ice into plastic, slowly-flowing material.

4. THERMAL PROPERTIES OF THE LATTICE

In this Chapter, we will consider the lattice contributions to the thermal properties of solids. We note that the dynamics of the free electrons plays a strong role in thermal properties (mainly thermal conduction) of some solids (mainly metals). These solids and the related physics will be treated in the following Chapter focusing on metals and their properties.

We start our treatment by introducing to the reader lattice vibrations, also known as ***phonons***, their microscopic origins and main properties. The treatment takes use of the concepts of reciprocal lattice and the periodicity of solids, and at times assumes the reader to be already familiar with the basic principles of thermodynamics and statistical physics. In the latter part of this Chapter, we briefly outline the historical developments and experiments that gave birth to the modern and successful treatment of the problem. Finally we connect our microscopic picture with exemplary macroscopic phenomena such as thermal conductivity and thermal expansion of solids.

4.1 Lattice Vibrations

Atoms in a crystal can vibrate around their equilibrium position, similar to a mass attached to a spring. In the case of atoms, the restoring force F is caused by the interatomic potential ϕ that we discussed earlier [see Eq. (3.2.10)]. Here, we mainly consider the regime where the force F remains purely linear, allowing us to treat the atomic lattice vibrations as as system of many coupled ***harmonic oscillators***. This approximation is known as the harmonic approximation. We only note here that the nonlinear contributions to the lattice vibrations can also be important, because such processes may give rise to ***umklapp scattering*** processes that often limit thermal conductivities of materials.

4.1.1 A Simple Harmonic Oscillator

For completeness, we start our treatment by introducing the simple harmonic oscillator model for a single atom. Once the behavior of a single atom acting as a harmonic oscillator is understood, we can then generalize our approach to N atoms (Fig. 4.1), which would give rise to a system consisting of $3N$ independent harmonic oscillators.[1] Although this picture is simplistic, it provides us an excellent starting point to understand the properties of solids. Furthermore, such a toy model can be easily improved by making the oscillators to be coupled to each other.

[1]Each oscillator may oscillate along each spatial direction. Therefore, in three-dimensional space the number of possible oscillations is three-fold compared to a one-dimensional case.

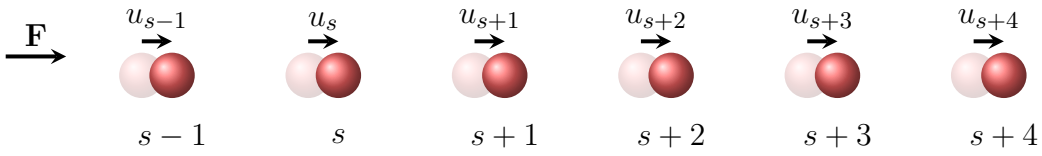

Fig. 4.1: N harmonic (uncoupled) oscillators oscillate independently of each other. Essentially, all oscillators behave identically, meaning that a driving force displaces each atom from its original equilibrium position (transparent atoms) by an equal amount $u_{s-1} = u_s = u_{s+1} = ... = u_{s+4} = u$.

So, we start by invoking the Newton's laws, and write the ***equation of motion*** for the atom simply as

$$M\frac{\partial^2 u}{\partial t^2} = M\ddot{u} = -\gamma u\,, \tag{4.1.1}$$

where M is the mass of the atom, and γ is the coupling constant arising from the interatomic potential ϕ, that takes into account the electrostatic interactions between the nearby atoms.[2] We also take use of the common convention of denoting the derivation of a variable over time using dots above the variable.[3] Since this problem has been solved many times during our earlier courses,[4] we outline here only the main results. The solution for the above partial differential equation is time-harmonic oscillation with oscillation frequency given by

$$\omega = \sqrt{\frac{\gamma}{M}}\,. \tag{4.1.2}$$

Furthermore, the total energy E of the system can be written as a sum of its kinetic and potential energies:

$$E = \frac{1}{2}Mv^2 + \frac{1}{2}\gamma u^2\,, \tag{4.1.3}$$

where v is the velocity of the atom. We can now connect this total energy E associated with the atomic harmonic oscillators to the ambient temperature T of the system, by recalling the ***equipartition theorem*** used in statistical physics/mechanics. This theorem states that in thermal equilibrium the total energy of the system is evenly distributed between the different forms of energy the system can sustain. In our case,

[2]Estimates of the coupling constant γ are beyond our current description but can be found from the book *Solid State Physics – An Introduction* by Philip Hofmann. In essence, the term γ can be connected with the elastic properties of solids, mainly with the Young's modulus Y of materials. A rule of thumb could therefore be given, where solids with high Y are expected to have also high values of γ.

[3]A single dot above the variable denotes one derivation over time, two two subsequent derivations and so on. Similarly, it is conventional to denote derivation over spatial coordinates using prime symbols, as for example in $\partial^2 f(x)/\partial x^2 = x''$.

[4]Perhaps the simplest way to solve this partial differential equation is by using a trial solution of a form of a time-harmonic plane wave.

the two forms of energy the system (N uncoupled oscillators) can sustain are the average kinetic energies and the potential energies.[5] Each form of energy can contribute an amount of $k_BT/2$ where k_B is the Boltzmann constant to the total energy of the system. Therefore, at the time when the displacement of atoms is maximized (kinetic energies minimized), the temperature T and the amplitude of the oscillation $u_{\max}$ are related by

$$\frac{1}{2}\gamma u_{\max}^2 = k_B T\,. \tag{4.1.4}$$

The above equation can now be solved for the amplitude

$$u_{\max} = \left(\frac{2k_BT}{\gamma}\right)^{1/2}\,, \tag{4.1.5}$$

allowing us to estimate the validity of the assumed harmonic approximation. For most solids and moderate temperatures, the displacement amplitude $u_{\max}$ is often small compared to the lattice constant a, justifying our linear approximation.

4.1.2 An Infinite Chain of Atoms

After solving our toy model, we now turn to the more realistic scenario, where the N oscillators are coupled to each other. For conceptual clarity, we restrict our treatment first to one dimension. We also first assume a monatomic lattice with a lattice constant a, where each unit cell consists of a single atom. In such one-dimensional system, the allowed lattice vibrations are ***longitudinal*** in nature, as has been clarified in Fig. 4.2. However, we note that already a two-dimensional system would support existence of ***transverse lattice vibrations*** as illustrated in Fig. 4.3.

4.1.2.1 One Atom Per Unit Cell

We consider a case where the atom s is connected to its nearest neighbors, i.e. atoms labelled as $s-1$ and $s+1$ (see Fig. 4.4). In equilibrium, the distance between each atoms is the lattice constant a, in which case we also set the displacements of each atoms from their equilibrium positions u_s to be identically zero. However, once the system is perturbed, the interatomic distances and respective atomic displacements u_s change. We assume that the forces on the atom s are caused only by its nearest neighbors.[6] Furthermore, we assume that harmonic approximation applies and the

[5] Recall that molecular gases may also store energy in form of molecular rotations.

[6] Although it would be a better approximation to include also coupling between the second (third etc.) nearest neighbors, not much new physics would emerge.

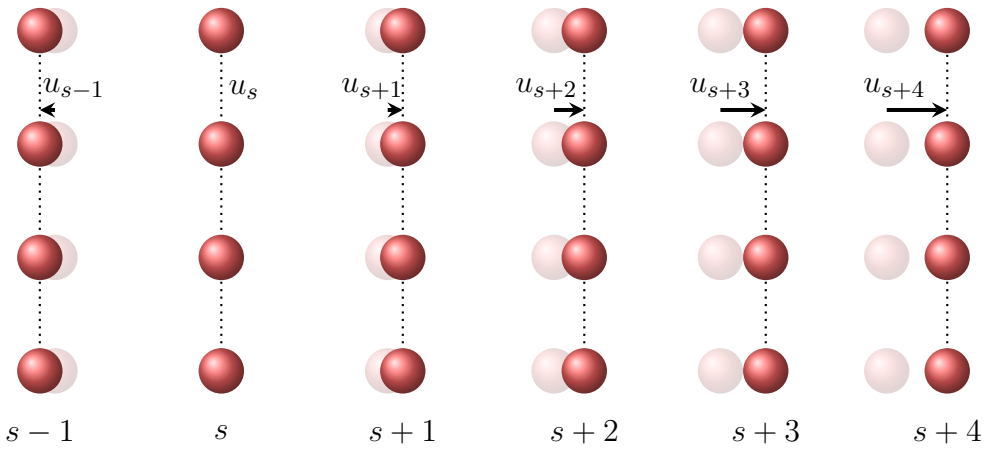

Fig. 4.2: Longitudinal phonons explained.

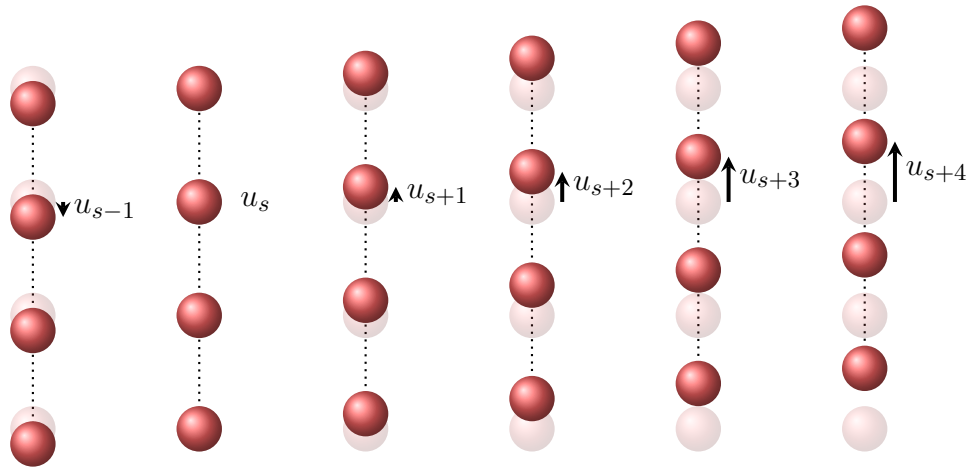

Fig. 4.3: Transverse phonons explained.

force is proportional to the difference of the atomic displacements $u_{s+1} - u_s$. The total force is therefore given by

$$F_s = \gamma\left(u_{s+1} - u_s\right) + \gamma\left(u_{s-1} - u_s\right) , \tag{4.1.6}$$

where γ is the force (or spring) constant between adjacent atoms.[7]

Applying the laws of Newton, we can now write the equation of motion for the s^{th} atom to be of form

$$M\ddot{u}_s = \gamma\left(u_{s+1} + u_{s-1} - 2u_s\right) . \tag{4.1.7}$$

This solution can be easily solved by looking for time-harmonic solutions, associated with time dependence of $\exp(-\mathrm{i}\omega t)$. In this case, the time derivation results in $\ddot{u}_s = -\omega^2 u_s$

[7]We note that the force constant is expected to be different for longitudinal and transverse lattice vibrations. Depending on the symmetry of the three-dimensional solid, it may also be expected to be different along different lattice planes of the solid, making the treatment quickly somewhat cumbersome.

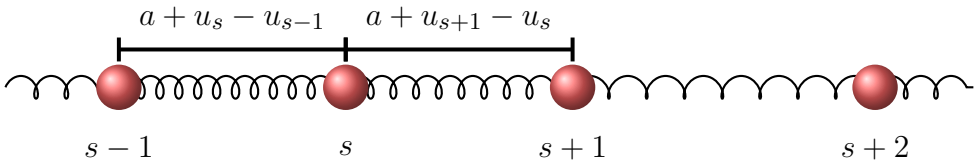

Fig. 4.4: N coupled harmonic oscillators.

allowing us to write the above equations into

$$-M\omega^2 u_s = \gamma\left(u_{s+1} + u_{s-1} - 2u_s\right) . \tag{4.1.8}$$

The above equation is now seen to be a ***linear difference equation***, allowing us to seek solutions of form

$$u_{s\pm1} = u\exp(\mathrm{i}ska)\exp(\pm\mathrm{i}ka) , \tag{4.1.9}$$

where k is the wavevector of the solution.[8] Note here that the value of a should also in the general case depend on the direction of the wavevector k.

Leaving the algebraic details of the rest of the problem to a homework exercise, we can now relate the frequency ω and the wavevector k to each other via equation given by

$$\omega^2 = (2\gamma/M)(1 - \cos ka) . \tag{4.1.10}$$

This equation is an important result, as it gives us the ***dispersion relation*** $\omega(k)$ of the system. However, the equation is conventionally given in a form where the frequency ω is not squared. This we can do conveniently by recalling the ***de Moivre's formula***, allowing to re-write the above equation as

$$\omega^2 = (4\gamma/M)\sin^2\frac{1}{2}ka , \tag{4.1.11}$$

which can be simply cast into the sought form

$$\omega = (4\gamma/M)^{1/2}\left|\sin\frac{1}{2}ka\right| . \tag{4.1.12}$$

This is our final form for the dispersion relation $\omega(k)$ of the solid in question and has been plotted in Fig. 4.5. Note, that when the number of atoms N is small value, the dispersion relation becomes discretized. In this picture/regime, it makes sense to

[8] Although many of the quantities in our one-dimensional case are actually scalars, we tend to call them vectors. Prime examples of such quantities are the wavevector k and the reciprocal wave vector $b = 2\pi/a$. This terminology has been chosen to facilitate the generalization of our treatment to the three-dimensional case, where such quantities are indeed vectors.

call any particular solutions fulfilling the above dispersion relation as normal modes or eigenmodes of the chain. Note also that the oscillations, or modes, of the lattice do not correspond to oscillations of particular atoms, rather such oscillations are collective oscillations of the whole lattice.

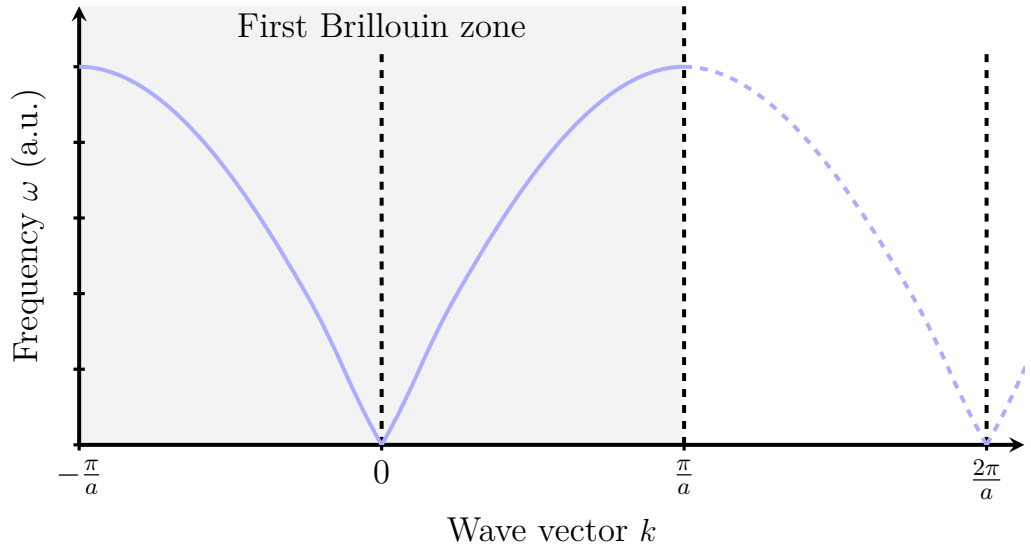

Fig. 4.5: Phonon dispersion relation. Note the very different behavior (the slope) of the dispersion relation near the points $k = 0$ and $k = \pi/a$.

Before moving forward in our treatment towards more complicated and also realistic lattices, we note that two very interesting limiting regions occur near the wavevector values of zero ($k = 0$) and $k = \pm\pi/a$ already for this simple monatomic one-dimensional lattice. First, we consider the case of $k = 0$, which corresponds to the case where the wavelength λ of the associated oscillation is very large. Consequently, the energy E associated with oscillation is small, which can be readily seen also by looking at the dispersion relation shown in Fig. 4.5. Physically, such lattice oscillation corresponds to a case, where nearby atoms all oscillate almost in phase with each other (see Fig. 4.6a). Importantly, we can in this case further simplify the dispersion relation above by noting that the argument of the sine function $1/2ka$ is now a very small quantity. In this case, the dispersion relation becomes simply

$$\omega(k) = \sqrt{\frac{\gamma}{M}}ak = vk\,, \tag{4.1.13}$$

where v can be understood as the (phase) velocity of the associated wave.[9] Furthermore, also the group velocity $v_g = \frac{\partial\omega}{\partial k}$ of the wave equals the velocity v. This result implies that

[9]Recall from earlier courses on electrodynamics that the phase velocity of a wave is given by $v = \omega/k$.

the propagation velocity of the waves do not depend on their frequency. Furthermore, by recalling that the propagation of sound in a medium is associated with the oscillations of the constituent atoms/molecules, we can argument that the velocity does in fact correspond to the velocity of sound in the solid!

Next, we consider the short-wavelength limit of the dispersion relation $\omega(k)$. The largest amplitude for the wavevector k occurs at $k = \pm\pi/a$, corresponding to a wavelength of $\lambda = 2a$. the lattice oscillation associated with this limit has been shown in Fig. 4.6b. Interestingly, we see that every second atom oscillates out of phase with respect to the adjacent atoms. Further insights into the limiting case is provided by considering the group velocity v_g of the associated wave, which physically corresponds to the slope of the dispersion relation $\omega(k)$. Because the dispersion curve flattens out near the points $k = \pm\pi/a$, we see that the group velocity of the associated solutions tends to zero implying that no energy is transferred by such waves. This finding agrees with the physical picture of adjacent atoms oscillating out of phase, i.e. forming a standing wave.

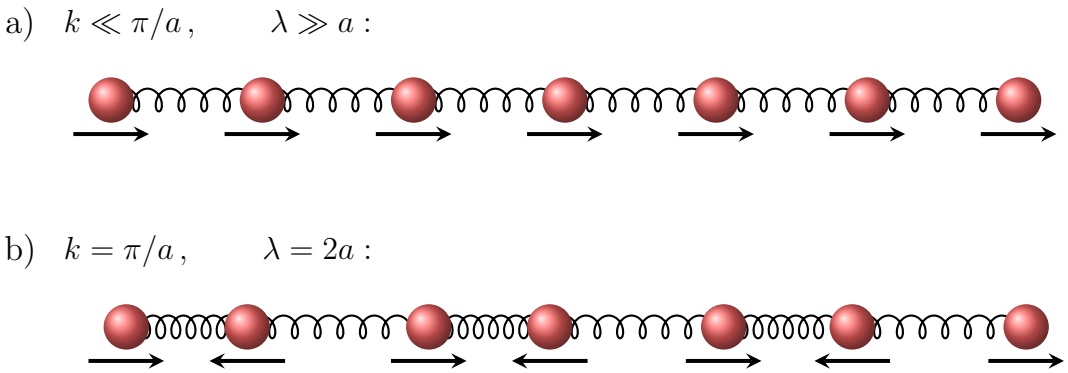

Fig. 4.6: Acoustic phonons associated with wavevector amplitudes a) $k \ll \pi/a$ and b) $k = \pi/a$.

4.1.2.2 The First Brillouin Zone

Next we discuss some of the most important consequences of the periodicity of the problem. Looking at the general solution of the dispersion relation (see Eq. (4.1.12), we see that it is a periodic function with periodicity corresponding to $2\pi/a$. Unsurprisingly, this periodicity corresponds exactly to the amplitude of the reciprocal wavevector $b = 2\pi/a$. Therefore, we see that physical outcome should not be affected if we add a reciprocal vector b to the wavevector k. In fact, the outcome would remain unaffected even if we would add any integer multiple of reciprocal vector b to the wavevector k.

Based on the above considerations, and by verifying them from the graphical representation of the situation in Fig. 4.5, we see that the situation can be entirely described by considering just the region of the dispersion relation where the wavevector ranges between $\pm\pi/a$. In principle, by taking use of the intrinsic symmetry of the system, just by considering the range $k = 0...\pi/a$ could be adequate.[10] This region of wavevector values $k = -\pi/a...\pi/a$ is known as the (first) ***Brillouin zone*** of the reciprocal lattice, and **it is a central concept in solid-state physics**. Interestingly, it consists of a single reciprocal lattice point, and can therefore be also understood as the Wigner–Seitz cell of the reciprocal lattice. Unfortunately, the generalization of the Brillouin zone into the three-dimensional reciprocal space (also known as the **k**-space) require some ***spatial abilities***.[11]

4.1.2.3 Two Atoms per Unit Cell

In this section, we extend our one-dimensional monatomic infinite chain of atoms into a diatomic infinite chain of atoms. We find out that in this case a new branch of lattice vibrations emerges, resulting in slightly more advanced phonon dispersion diagrams. This conceptual change is exactly what we would also see when performing experiments, giving us confidence towards our approach. In the next sections, we will subsequently generalize our treatment to periodic chains (not infinite one) and to the general three-dimensional case.

We start by considering a system of two masses, M_1 and M_1, connected to each other by springs (or equivalently forces similar to the Hooke's law). We denote the displacements of the s^{th} atoms at any given time t with symbols u and v, the first (last) corresponding to the mass M_1 (M_2). Again we assume, that the chain of atoms has periodicity of a. We can then write the equations of motion for the system as

$$M_1 \frac{\mathrm{d}^2 u_s}{\mathrm{d}t^2} = M_1 \ddot{u}_s = -\gamma \left[2u_s - v_{s-1} - v_s\right] , \tag{4.1.14}$$

$$M_2 \frac{\mathrm{d}^2 v_s}{\mathrm{d}t^2} = M_2 \ddot{v}_s = -\gamma \left[2v_s - u_s - u_{s+1}\right] . \tag{4.1.15}$$

Without going into the details of solving such as system of coupled linear differential

[10] Be warned here, that not all systems might exhibit such symmetry.

[11] In order to truly understand and appreciated solid-state physics, one needs to devote time and patience on this topic.

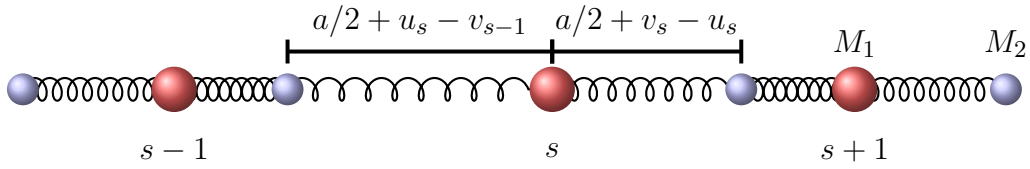

Fig. 4.7: One-dimensional, infinite chain of diatomic atoms that are connected to adjacent atoms by springs. The lattice constant of the chain is a. The mass of the red (blue) atom is M_1 (M_2). When the system is in motion, the s^{th} red (blue) atom is displaced by an amount u_s (v_s) from it equilibrium position.

equations, we consider trial solutions of more general form

$$u_s(t) = u\mathrm{e}^{\mathrm{i}(kas-\omega t)}\,, \tag{4.1.16}$$

$$v_s(t) = v\mathrm{e}^{\mathrm{i}(kas-\omega t)}\,, \tag{4.1.17}$$

that we and insert into the above two equations of motion resulting in

$$-\omega^2 M_1 u = \gamma v\left(1+\mathrm{e}^{-\mathrm{i}ka}\right) - 2\gamma u\,, \tag{4.1.18}$$

$$-\omega^2 M_2 v = \gamma u\left(\mathrm{e}^{\mathrm{i}ka}+1\right) - 2\gamma v\,. \tag{4.1.19}$$

The above system of linear equations can next be written as a matrix equation simply by rearranging the terms. This equips us with the tools of linear algebra to find a solution for the system. Mainly, by requiring the determinant of the formed coefficient matrix to vanish, we can find a unique solution for the system. Therefore, we seek conditions that fullfill the condition

$$\begin{vmatrix} 2\gamma-\omega^2 M_1 & -\gamma\left(\mathrm{e}^{-\mathrm{i}ka}+1\right) \\ -\gamma\left(1+\mathrm{e}^{\mathrm{i}ka}\right) & 2\gamma-\omega^2 M_2 \end{vmatrix} = 0\,. \tag{4.1.20}$$

From the above determinant equation we can finally find a relation between the frequency ω and the wavevector k. After a bit of tedious algebra, the relation turns out to be given by

$$\omega^2 = \gamma\left(\frac{1}{M_1}+\frac{1}{M_2}\right) \pm \gamma\left[\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4}{M_1 M_2}\sin^2\frac{ka}{2}\right]^{1/2}\,. \tag{4.1.21}$$

We rest our case here and turn our attention to the meaning of the above equation, that in essence provides us the dispersion relation $\omega(k)$ of the system of one-dimensional,

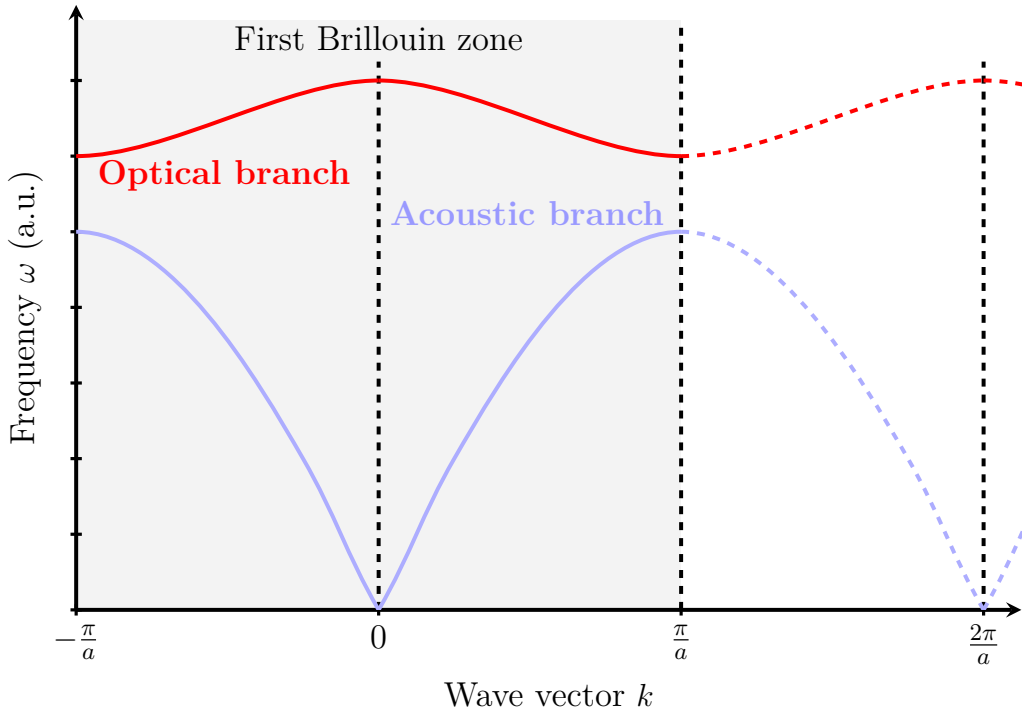

Fig. 4.8: Diatomic phonon dispersion.

periodic chain (lattice constant a) consisting of diatomic atoms of mass M_1 and M_2. The dispersion relation $\omega(k)$ of the system has been plotted in Fig. 4.8

Similar to the monatomic case we considered earlier, this dispersion relation also has a periodicity of $2\pi/a$, which is the amplitude of the reciprocal lattice vector b. As a consequence, we can fully describe the behavior of the system by only considering the first Brillouin zone, which fact has been also highlighted in Fig. 4.8.

Even more interestingly, we also see that the dispersion relation of Eq. (4.1.21) exhibits two separate solutions, i.e. branches, that occur because of the plus–minus ($\pm$) sign in front of the second term. This fact is also visible in the Fig. 4.8, where the two branches have been plotted using red and blue curves. The solution (blue curve) that tends to zero at $k = 0$ values is called the ***acoustic branch***, and behaves similarly as the solutions found for the case of the monatomic lattice (see again Fig. 4.5). As the name gives away, the solutions related to the acoustic branch can be understood as sound waves propagating through the lattice.

The other branch (red curve) is called the ***optical branch***, mainly because the associated solutions can be excited using incident electromagnetic fields. Similar to every

scattering process we can imagine, also such photon–phonon scattering processes are associated with the conservation of total momentum and energy of the system (here one photon and the ideal atomic lattice sustaining phonons). The photon momentum is given by $p_{\text{phot}} = \hbar k = h/\lambda$, and is in general quite small.[12] In practise, phonons exhibit such low values of momentum only near the $k = 0$ points of the phonon dispersion relations. However, the energies associated with such photons are not so small. As a consequence, in order for the process to conserve both the total momentum and the total energy of the system, the associated phonon energies should be non-zero. This can only occur at the upper, optical branches of phonon dispersion curves where the frequencies ω associated with the optical phonons do not tend to zero at $k = 0$.

It is also worth to note that the behavior of the two branches is very different near the $k = 0$ point also because the phonon dispersion relation of the optical branch flattens out. This latter fact again implies that in this limiting case the optical phonons do not transfer energy.[13]

4.1.3 A Finite Chain of Atoms

Next we improve our models developed so far to understand the lattice vibrations by changing the treatment from infinite chains to finite chains. This change is motivated by the fact that a material consisting of infinite number of atoms would also exhibit the capability to store an infinite amount of energy. Such a situation is clearly unphysical,[14] because it would for example result in a material exhibiting an infinite heat capacity. Therefore, we need to to change our description in order to avoid such issues. The solution is to consider finite, albeit very long chains of atoms, in which case we will soon see that the lattice vibrations will become quantized.

We start our description by introducing the concept of ***periodic boundary conditions***, which a one-dimensional chain of N atoms is mathematically stated as

$$u_{N+s} = u_s \,, \tag{4.1.22}$$

[12] In practise, phonon energies in materials are of the order of 1–100 meV, which corresponds to photon wavelengths of 12.4 μm – 1.24 mm. Note that such wavelengths are very long compared to the lattice constants a of periodic solids and correspond to infrared radiation, and not to visible light. In that sense, the term optical branch is a misnomer.

[13] Phonons, i.e. lattice vibrations, are called optical phonons, when they are associated with optical branch. Similarly, acoustic phonons correspond to lattice vibrations residing at the acoustic branch.

[14] We need to note here that infinite number of atoms would neither fit our universe, not to mention a small, e.g. a cubic centimeter volume of a solid.

where u_s is the displacement of the s^{th} atom. Often used visualization is to imaging that the 1D chain would be curved back to its starting point forming a ring of atoms.[15] Conveniently, the forms of the phonon dispersion relations $\omega(k)$ are not changed when moving the description to periodic boundary conditions. However, the allowed lattice vibrations no longer form a continuum, but rather a discretized set of solutions that follow the curves of ideal phonon dispersion relations.[16]The discretization of solutions can be understood as follows. We consider a periodic chain of N atoms exhibiting a lattice constant of a. Subsequently, the longest possible wavelength to consider is Na. This can be seen by requiring the above periodic boundary condition to hold resulting in condition

$$\mathrm{e}^{\mathrm{i}kas} = \mathrm{e}^{\mathrm{i}ka(N+s)}\,, \tag{4.1.23}$$

which is equivalent stating that

$$\mathrm{e}^{\mathrm{i}kNa} = 1\,. \tag{4.1.24}$$

By recalling the properties of the exponential function, we see that the above restriction is equal condition

$$k = \frac{2\pi}{aN} m\,, \tag{4.1.25}$$

where m is an integer. This equation which implies that the wavevector k is discretized. We now recall that the problem can be fully described by restricting the treatment to the first Brillouin zone (i.e. range $-\pi/a \leq k \leq \pi/a$). As a consequence, it is enough to consider how many values of wavevectors fit in to this zone. By looking at the above equation, we see that the only N different values of wavevector k (i.e. phonon modes) can fit the first Brillouin zone per dispersion branch. This fact will be used later on and is thus worth to memorize.

Next we consider, how many wavevectors k can fit the first Brillouin zone of a diatomic lattice.[17] Because the periodic boundary conditions are as above, we see that each branch (acoustic and optical) still can support N phonon modes. Therefore, in total the lattice of diatomic atoms supports up to $2N$ lattice modes.

This discretization of phonon modes has been illustrated in Fig. 4.9, where we have

[15]Because it is difficult to generalize such visualization to three-dimensional space, the lecturer dislikes this conceptualization.

[16]In other words, the actual phonon dispersion relations are no longer continuous lines, but rather densely-spaced collections of points.

[17]The ultimate generalization to three-dimensional solids follows in after the next sub-section.

considered the case of chain consisting of $N = 16$ atoms. Particularly, we see that the discretized wavevector k values and the corresponding frequencies ω coincide perfectly with the continuous phonon dispersion relation $\omega(k)$ corresponding the the case of infinite chain of atoms.

We conclude this section by emphasizing the conceptually important change we have introduced into our system. The finite but periodic chain of N atoms[18] can only support a finite amount of different lattice vibrations, that are often also called as the ***normal modes*** of the lattice. Therefore, the periodic lattice can also only support a finite amount of energy in lattice vibrations, subsequently resulting in physically reasonable, finite-valued material constants, such as thermal conductivities.[19] Furthermore, an arbitrary lattice vibration can be always decomposed into a superposition of normal modes of the lattice. In other words, the normal modes of the lattice form a ***basis*** that can be used to describe an arbitrary lattice vibration.

4.1.4 Phonons as Quantized Vibrations of the Lattice

This section is devoted to clarify the importance and consequences of the discretization process we have outlined above. We learn later in this Chapter that the discretized nature of lattice vibrations is an essential feature of the model that can explain the heat capacities associated with lattices.[20] Without going into the details here, we state that the energies associated with an ideal harmonic oscillator are given by

$$E_l = \left(l + \frac{1}{2} \right) \hbar\omega \,, \tag{4.1.26}$$

where l is an integer and E_l is the associated energy. Almost identical behavior is seen also in our system of periodic chain of N atoms. In our case, the associated energies

[18] that is each periodic chain consists of N atoms, and those periodic chains extend up until infinity.

[19] We note an interesting connection here to the branch of mathematics known as the ***set theory*** that is related to the ***cardinality of the continuum***. In fact, our system could still store an infinite amount of energy on it, because each lattice vibration can store l number of phonons, where l is in principle an unbound integer. We have only bound the allowed values of wavevector k to N values, whereas before we allowed k to be a continuous real-valued quantity ($k \in \mathbb{R}$). In other words, we have 'only' lowered the cardinality of the looming infinity from the cardinality of real numbers $\mathfrak{c}$ to the cardinality of natural numbers $\mathfrak{N}_0$. The change might be seemingly futile, since we're still troubled by the infinity. However, the size of the infinity has been dramatically lowered, because $\mathfrak{c} = 2^{\aleph_0} > \aleph_0$. We deem it appropriate to acknowledge here the ground-breaking contributions of ***Georg Cantor*** to the set theory.

[20] We note again, that in general the heat capacities of solids are due to two different contributions. One arising from lattice vibrations and the other from the movement of free (conduction) electrons. The former contribution dominates in dielectrics, notably in diamond, while the latter one explains the good thermal conductivities of most metals.

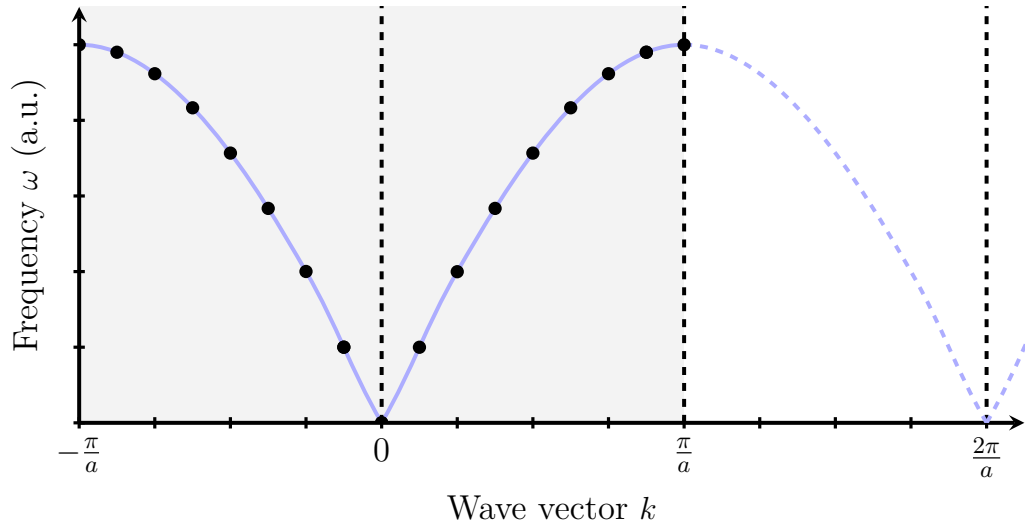

Fig. 4.9: Discretization of phonons. The blue curve shows the (acoustic) phonon dispersion relation $\omega(k)$ for a monatomic lattice for an infinite lattice. The allowed wavevector k values for the case of a periodic chain consisting of $N = 16$ atoms are shown as black dots. As is seen, the discretized values of k follow the dispersion relation of the infinite lattice. The first Brillouin zone has been highlighted in grey.

are given by

$$E_l(k) = \left(l + \frac{1}{2}\right)\hbar\omega(k)\,, \tag{4.1.27}$$

where we have replaced the continuous parameter ω with the discretized frequency associated with the phonon dispersion relation $\omega(k)$. Consequently, the discretized energy states of the system $E_l(k)$ become dispersive in nature, i.e. depend on the value of the wavevector k.[21] Although we do not adopt the following notation here, it might be helpful and less ambiguous to denote the energy states of the system as $E_{l,k}$ highlighting the fact that both parameters l and k are discrete. This notation might also help to appreciate the fact that the wavevector k associated with a very specific lattice vibration mode can be understood as a ***quantum number*** of the system.[22]

Before moving to the next section, we note that in the light of the above discussion, lattice vibrations, i.e. phonons, have very similar properties as bosonic ***elementary particles***. They are associated with discretized energy levels, obey the ***Bose–Einstein***

[21] Perhaps it is helpful to compare this result against other, more familiar systems. Consider, for example, the allowed energies of a single photon, propagating in free space. The energy of a freely-propagating photon **does not** depend on the direction of propagation., i.e. on its wavevector **k**.

[22] We note here the close analogy between the spherically symmetric case of ***atomic orbitals***, similarly associated with quantum numbers describing the basis of the orbitals. Here, our problem exhibits translational symmetry (of Cartesian coordinate system), and is similarly associated with quantum numbers that describe the basis consisting of the normal modes, that can be used to describe the physics of the system.

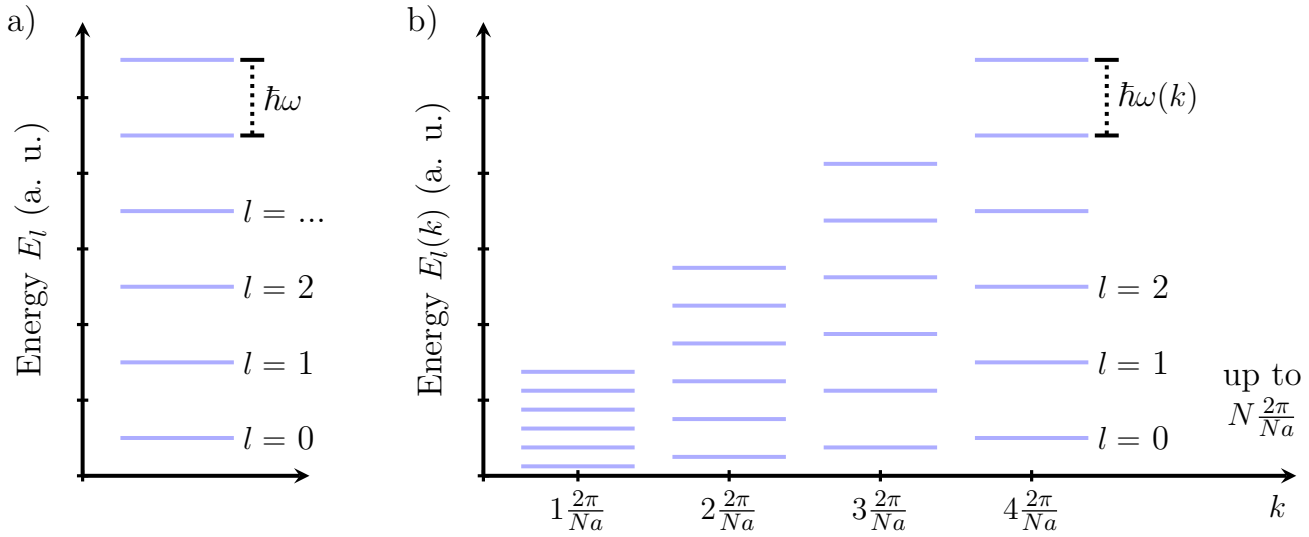

Fig. 4.10: Discretization and dispersion of energy states. 'a) Energy level diagram for one harmonic oscillator. b) Energy level diagram for a chain of atoms with one atom per unit cell and a length of N unit cells.'

statistics and even carry intrinsic spin momentum. The main difference between normal bosonic particles, such as photons, is that phonons do not carry proper momentum, but rather ***crystal momentum***, also called as ***quasimomentum***, arising from the fact that phonons are an ***emergent phenomena***, a collective excitation of the whole crystalline lattice, not associated with a single atom in the lattice.[23] Here, it is enough to remember that phonons are discretized, behave almost as particles,[24] and obey the Bose–Einstein statistics. The last property means that unlike fermionic electrons which obey the Pauli exclusion principle, many phonons can simultaneously exhibit the same quantum state. It is these properties that will be important in the upcoming treatment.

4.1.5 Generalization to Three-Dimensional Solids

Before we end our introduction to the lattice vibrations and the collective excitations known as phonons, we take a moment to outline how our so far one-dimensional treatment is generalized to three-dimensional lattices and space. Although the generalization is conceptually straightforward, it does require skills of abstraction and moderate spatial abilities. However, in order to truly understand and appreciate the behavior and properties of solid-state systems, it is paramount to understand the main features of the full treatment of three-dimensional solids. The following section might be hard to

[23] These points are admittedly subtle and complex. I suggest leaving them for future Chapters, once the concepts of collective excitations, such as phonons, have become more standard tools in your repertoire.
[24] Phonons can in fact be understood as ***quasiparticles*** or collective excitations (of the lattice).

grasp, but we ensure that the reward in the form of improved understanding is worth the effort!

First, we note that the number of phonon branches increase when we generalize the treatment to three-dimensional lattices. For example, monatomic lattices can support now three acoustic branches. One branch correspond to the longitudinal phonons we have already treated, while the remaining two branches correspond to transverse phonons, that we have only schematically highlighted in Fig. 4.3. Unsurprisingly, diatomic lattices can support six branches, three acoustic and three photonic.

The mathematical treatment becomes also more cumbersome, because some of the quantities, mainly the lattice vectors $\mathbf{a}$ and reciprocal vectors $\mathbf{b}$, become true vectors. As a consequence, the integer m associated with a specific normal mode of a periodic lattice becomes a set of integers, as is seen in the below generalization for the periodic boundary conditions for a cubic lattice with lattice constant a:

$$\mathbf{k} = \left(k_x, k_y, k_z\right) = \frac{2\pi}{aN}\left(n_x, n_y, n_z\right) = \left(\frac{n_x 2\pi}{L}, \frac{n_y 2\pi}{L}\frac{n_z 2\pi}{L}\right) . \tag{4.1.28}$$

A considerably harder concept to generalize than the ones mentioned above, is the range of wavevector $\mathbf{k}$ values that should be now considered. In the three-dimensional case, the range $-\pi/a \leq k \leq \pi/a$ we considered earlier becomes the Wigner–Seitz cell of the reciprocal lattice, which is better known as the first Brillouin zone. For the case a cubic fcc lattice, the first Brillouin zone is a truncated octahedron,[25] and has been illustrated in Fig. 4.11. Several often-encountered solids form an fcc lattice, such as aluminum (Al), copper (Cu), silver (Ag), gold (Au), sodium chloride (NaCl), diamond (C), silicon (Si) and germanium (Ge).

What makes the concept of Brillouin zones and the graphical representation of the phonon dispersion relations $\omega(\mathbf{k})$ particularly hard to grasp, is that it is difficult to represent a four-dimensional data set (ω as a function of k_x, k_y and k_z) in a three-dimensional world. Fortunately, the problem is not as formidable as first thought, since often it is enough to plot the dispersion relations only near a few specific ***points of high symmetry***, or via a few specific lines connecting these points. These important points of high symmetry for the first Brillouin zone for the fcc lattice are named as

[25] Interestingly, the Wigner–Seitz cell of a cubic bcc lattice is also a truncated octahedron.

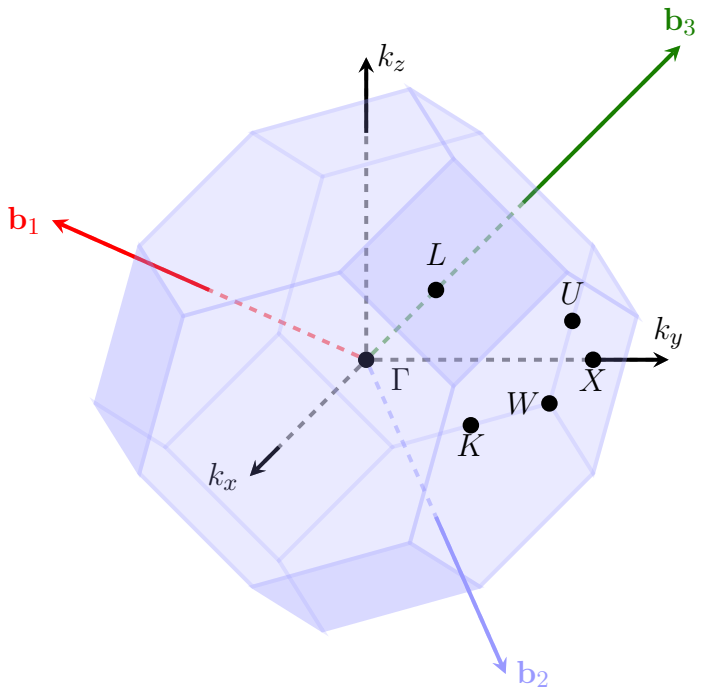

Fig. 4.11: The three-dimensional (first) Brillouin zone of an fcc cubic lattice along with the associated points of high symmetry. The Brillouin zone forms a truncated octahedron. The associated reciprocal lattice vectors $\mathbf{b}_1 = \frac{2\pi}{a}(-1,1,1)$, $\mathbf{b}_2 = \frac{2\pi}{a}(1,-1,1)$ and $\mathbf{b}_3 = \frac{2\pi}{a}(1,1,-1)$ have been also shown in red, blue and green arrows, respectively. The points of high symmetry are K, L, U, W and X, and the origo $\mathbf{k} = (0,0,0)$ is denoted with Γ.

K, L, U, W and X and are shown in the Fig. 4.11. The wavevectors $\mathbf{k}$ and sets of integers (n_x, n_y, n_z) are given in Table 4.1. In addition, the convention is to use the Greek letter Γ for the origo $\mathbf{k} = (0,0,0)$.[26] With the help of these symmetry points, it becomes possible to visualize the dispersion relations associated with three-dimensional solids already using regular two-dimensional plots.

Table 4.1: Points of high symmetry for the first Brillouin zone (truncated octahedron) of the cubic fcc lattice.

Symmetry point	(n_x, n_y, n_z)	(k_x, k_y, k_z)
Γ	$(0,0,0)$	$(0,0,0)$
X	$(0,1/2,1/2)$	$(0,2\pi/a,0)$
L	$(1/2,1/2,1/2)$	$(\pi/a,\pi/a,\pi/a)$
W	$(1/2,1,0)$	$(\pi/a,2\pi/a,0)$
U	$(1/4,1,1/4)$	$(\pi/2a,2\pi/a,\pi/2a)$
K	$(3/4,3/4,0)$	$(3\pi/2a,3\pi/2a,0)$

[26] We do not expect you to master this subtlety during introductory course. It is perfectly enough to just acknowledge that such conventions, terminology and concepts exist.

Two example visualizations have been given in Figs. 4.12 and 4.13.[27] Both cases correspond to a cubic fcc lattice, with the first Brillouin zone given by a truncated octahedron. Only the lines of high symmetry, which are formed by connecting the points of high symmetry, have been plotted. The first figure shows the phonon dispersion relation of a monatomic aluminum (Al). First we note that only the acoustic branch is visible, as should be the case for a monatomic lattice. Second, it is important to notice that the x-axis of the plot consists of several lines, that correspond to **different directions** in the reciprocal space. These lines are shown in Fig. 4.12b. It is paramount to understand the actual meaning of the x-axis, in order to correctly interpret such dispersion plots.

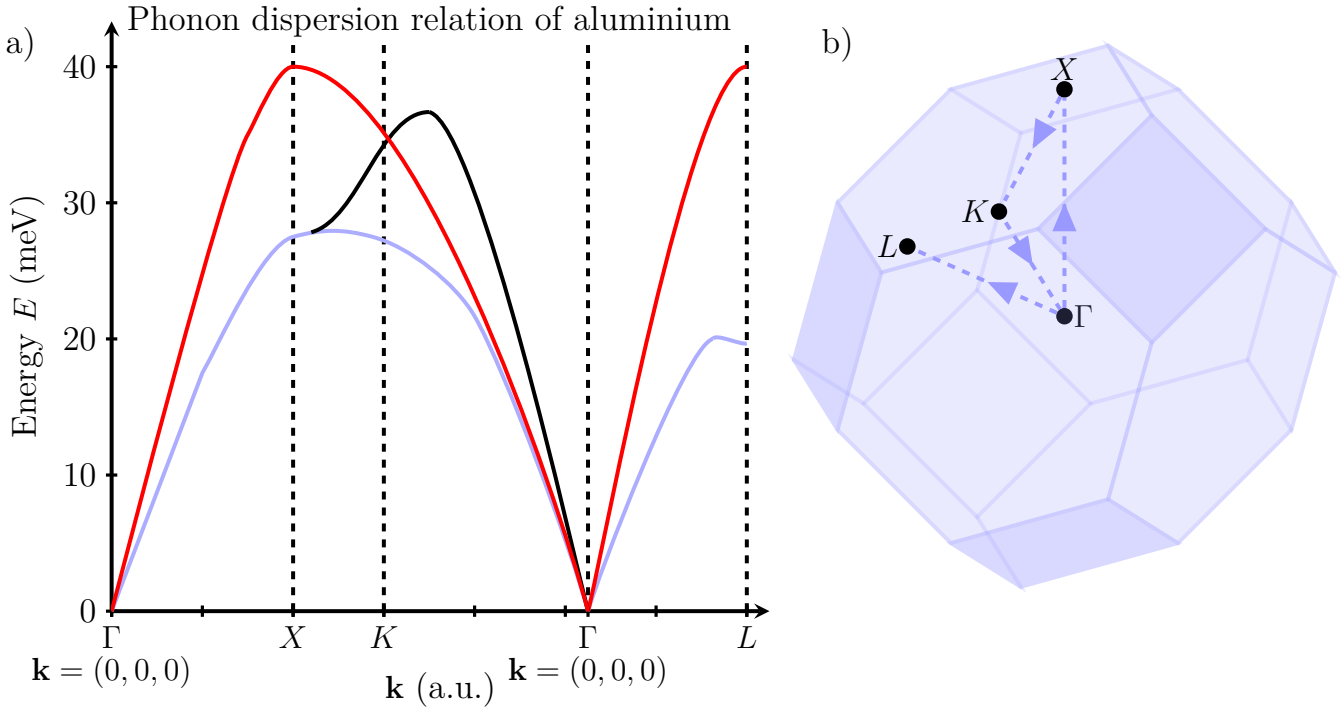

Fig. 4.12: a) Phonon dispersion relation of aluminium. b) Three-dimensional Brillouin zone of Al along with the relevant symmetry points Γ, X, K and L.

The latter phonon dispersion plot (see Fig. 4.13) corresponds to diamond. Diamond structure can be understood as a diatomic fcc lattice. Consequently, the optical branches of phonon dispersion should be visible, as is the case. We also see, that more phonon branches are visible compared to the monatomic case of Al. This agrees with our earlier discussion regarding the number of possible phonon branches in three-dimensional solids.

[27] We note for the interested reader that such beautiful data sets can be measured by performing inelastic neutron scattering experiments. During such inelastic scattering processes, phonons associated with specific wavevector $\mathbf{k}$ can be either emitted or absorbed by the solid. By then carefully controlling the wavevectors of the incident neutrons and recording the scattering directions of the output neutrons along with their associated kinetic energies, information of the phonon dispersion relations can be extracted.

We also note that in both cases, the dispersion of the acoustic phonon branches is linear near the Γ-point. In addition we note that the optical phonon branches of diamond are indeed quite flat near the Γ-point. Therefore, the limiting cases we considered for the one-dimensional lattices have moved unchanged into the three-dimensional case. This is reassuring, because it allows to introduce these important concepts and properties first using conceptually far simpler models.

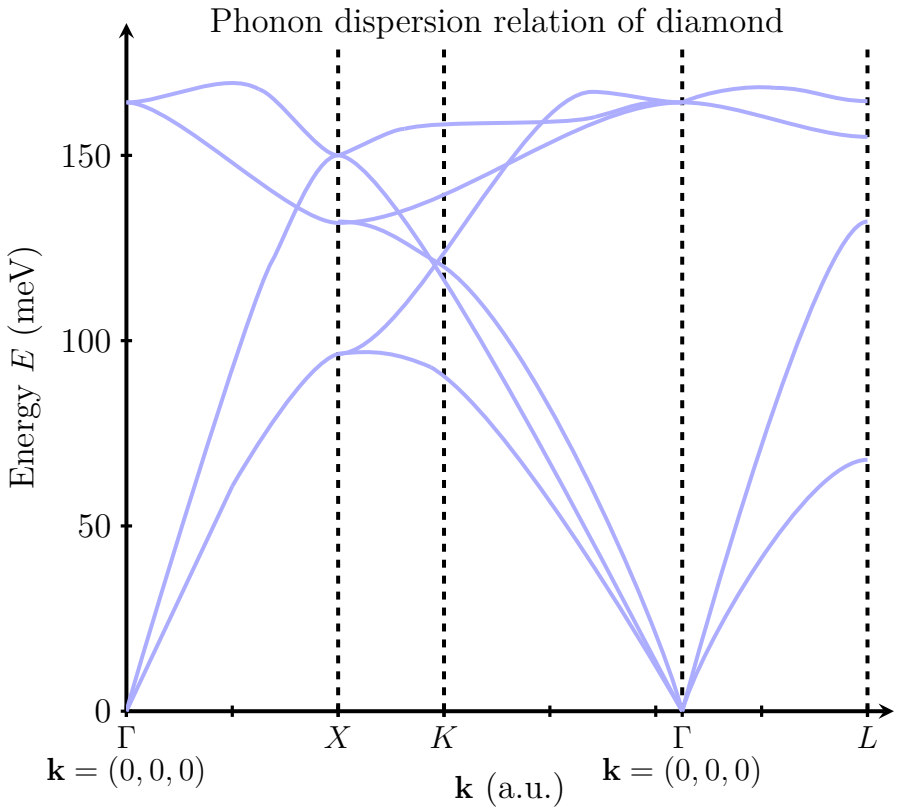

Fig. 4.13: Phonon dispersion relation of diamond.

4.2 Heat Capacity of the Lattice

From a historical point of view, the successful explanation of the heat capacities of various solids was a landmark achievement of the emerging quantum mechanical treatment of solids. Earlier puzzlement caused by the problem started turning into a firm belief into the quantum mechanics, forever changing many other branches of physics not to mention its importance in the birth of solid-state physics.

Although the historical perspective is interesting, here we will only introduce to the reader the first successful theory to explain the heat capacities of solids, which is known as the Debye model. The models that preceded the Debye model have been well described and treated in many other sources, such as in the book *Solid State Physics* –

An Introduction by Philip Hofmann.

4.2.1 Debye Model

Earlier models to explain the heat capacities had described the solids either from a purely classical point of view (Dulong–Petit) and were based on the statistical physics, or were principally based on quantum mechanics (Einstein model). Despite the success of Einstein model at the moderate and high temperature ranges, the model utterly failed at very low temperatures. ***Peter Debye*** was the first to realize that it was the band structure of solids that had been missed in the development of models. He formulated a theory, which treated the atoms in solids as coupled harmonic oscillators instead of uncoupled ones resulting in a model, an advancement that finally resulted in a working theory. In other words, Debye was the first to correctly take the effects of lattice vibrations, i.e. phonons, and their discretized and dispersive nature into account.

Next, we outline the derivation of the Debye model. We start by defining the heat capacity we seek to calculate as[28]

$$C \equiv \frac{\partial \langle E \rangle}{\partial T}\,, \tag{4.2.29}$$

where $\langle E \rangle$ is the mean energy of the associated oscillators of the system. and T is the temperature. In words, the heat capacity C corresponds to the rate of change of the mean energy when the temperature T of the system is changed. In order to calculate the mean energy $\langle E \rangle$, we need to calculate the mean phonon state occupancies $\langle l \rangle$, i.e. calculate how many phonons (and at which energy levels) in the matter have been excited. This can be achieved by taking use of Eq. (4.1.27) and re-writing it to describe the mean energy associated with the mean phonon number $\langle l \rangle$ as given by

$$\langle E \rangle = 3\left(\langle l \rangle + \frac{1}{2}\right)\hbar\omega(\mathbf{k})\,. \tag{4.2.30}$$

Here, the factor of 3 arises from the statistical nature of the system, because each oscillator exhibits three degrees of freedom.

[28]'We ignore the difference between heat capacities at constant volume and at constant pressure. For solids, this difference is usually quite small but not entirely negligible. Experimentalists usually like to measure the heat capacity at constant pressure, for example, at ambient pressure. Theorists, on the other hand, prefer calculations at constant volume because otherwise all the quantum mechanical eigenvalues have to be recalculated for every volume'. So, rigorously speaking we are calculating C_V.

In order to proceed, we note that the phonon number is given by the ***Planck distribution***[29]

$$\langle l \rangle = \frac{1}{\exp(\hbar\omega(\mathbf{k})/\tau) - 1}\,, \tag{4.2.31}$$

where the notation has be cleaned by using $\tau \equiv k_B T$ where k_B is the Boltzmann constant. By inserting the above equation into Eq. (4.2.30) results in

$$\langle E \rangle = 3\left(\frac{1}{\exp(\hbar\omega(\mathbf{k})/\tau) - 1} + \frac{1}{2}\right)\hbar\omega(\mathbf{k})\,. \tag{4.2.32}$$

We are ultimately interested in the derivative of the mean energy. Because the latter term is constant with respect to temperature T, it will play no role in our calculation of the heat capacity. Consequently, we will neglect this term in what follows. This allows us to write the mean energy simply as

$$\langle E \rangle = \frac{3\hbar\omega(\mathbf{k})}{e^{\hbar\omega(\mathbf{k})/\tau} - 1}\,, \tag{4.2.33}$$

which is almost the form we are pursuing. What is left is the summation over all oscillators and their degrees of freedom, that could be achieved by brute force summation as given by[30]

$$\langle E \rangle = 3\sum_{\mathbf{k}} \frac{\hbar\omega(\mathbf{k})}{e^{\hbar\omega(\mathbf{k})/k_B T} - 1}\,. \tag{4.2.34}$$

However, it is difficult to perform this summation analytically.[31] The problems in this approach are, that as we saw in Fig. 4.9, the phonon energies are discretized and dispersive making it non-trivial to perform the calculation. The first issue is often solved by assuming that the wavevectors $\mathbf{k}$ are not discrete, but a continuous parameter. Assuming this ***continuum limit***, the above summation can be turned into an integral over the frequencies $\omega(\mathbf{k})$:

$$\langle E \rangle = 3\int_0^{\omega_D} D(\omega, \mathbf{k}) \frac{\hbar\omega(\mathbf{k})}{e^{\hbar\omega(\mathbf{k})/k_B T} - 1}\,\mathrm{d}\omega\,, \tag{4.2.35}$$

where ω_D is known as the Debye frequency to be solved for. In adition, we had to introduce a new function $D(\omega, \mathbf{k})$, known as the ***density of states*** (DOS) that is

[29] Because of the elegance in the derivation of the Planck distribution, we leave this problem as a homework exercise.

[30] Note that the summation is in three-dimensional system actually three summations over all the allowed values of the wavevector $\mathbf{k}$.

[31] Estimate of the time it would take for a modern computer to perform $3 * 10^2 3$ multiplications..

essential in order for the integral to correspond approximately to the original summation in Eq. (4.2.34). Later on, we learn that the DOS is an essential concept almost in all of modern solid-state physics.[32] Because of its importance, we will next describe in more detail how the DOS is calculated in our case of three dimensions.

4.2.1.1 Density of States in Three Dimensions

We start by applying the periodic boundary conditions for our system consisting of N^3 atoms within a cubic volume of side length L. For this case the periodic boundary condition is given by

$$\mathrm{e}^{\mathrm{i}\left(k_x x + k_y y + k_z z\right)} = \mathrm{e}^{\mathrm{i}\left(k_x (x+L) + k_y (y+L) + k_z (z+L)\right)} \,. \tag{4.2.36}$$

As we have seen before, this condition is fullfilled by restricting the components of the wavevectors k_x, k_y and k_z to be

$$k_i = \left(-\frac{N\pi}{L}, \cdots, -\frac{4\pi}{L}, -\frac{2\pi}{L}, 0\,, \frac{2\pi}{L}\,, \frac{4\pi}{L}\,, \cdots, \frac{N\pi}{L} \right) \,, \tag{4.2.37}$$

where $i = \{x, y, z\}$. This means that a volume of $\left(2\pi/L\right)^3$ can occupy a single value of wavevector $\mathbf{k}$. This information can be used to calculate the total number of modes N for a given polarization state and for a given phonon branch. Noting, that the allowed wavevector $\mathbf{k}$ values are inside a sphere of radius $k \equiv \|\mathbf{k}_{\max}\|$, the total number of modes N can be written as

$$N = \left(\frac{L}{2\pi}\right)^3 \frac{4\pi k^3}{3} \,. \tag{4.2.38}$$

By applying the chain rule to the above equation, the DOS becomes

$$D(\omega, \mathbf{k}) \equiv \frac{\mathrm{d}N}{\mathrm{d}\omega} = \frac{V k^2}{2\pi^2} \frac{\mathrm{d}k}{\mathrm{d}\omega} \,, \tag{4.2.39}$$

where we have simplified the notation by expressing the cubic volume as $L^3 = V$. We note that the final term $dk/d\omega$ depends on the dispersion relation of the phonon branch, and remains to be solved for. Therefore, depending on the type of the problem at hand, the DOS may exhibit quite different range of values.[33]

[32] Not no mention other branches of physics, such as photonics.

[33] Lucid derivation for the general case for DOS has been given in the book *Introduction to Solid State Physics* by Charles Kittel.

4.2.1.2 Debye Model for the Density of States

Now, we are ready to solve the second problem associated with our original calculation of the mean energy. This problem was associated with the dispersive nature of the frequency $\omega(\mathbf{k})$. In general, the dispersion relation $\omega(\mathbf{k})$ is not a linear function of the components of $\mathbf{k}$. This fact we have already seen in all the figures showing us phonon dispersion relations of both ideal and real solids (such as aluminum of diamond).

At very low temperatures, however, the allowed states of the system are mostly only due to the acoustic phonon branches. As we have seen earlier, the dispersion of acoustic branches near the $\mathbf{k} = (0, 0, 0)$ point (known as the Γ-point), is approximately linear as given by[34]

$$\omega = v\|\mathbf{k}\| = vk\,, \tag{4.2.40}$$

where v is the velocity of sound. This is the approximation for the dispersion relation that is made in the Debye model, and it allows us to write the DOS of the system, given by Eq. (4.2.39), explicitly as

$$D(\omega, \mathbf{k}) = D(\omega) = \frac{V\omega^2}{2\pi^2 v^3}\,, \tag{4.2.41}$$

solving our second problem.

4.2.1.3 Debye Heat Capacity and the Debye Temperature

We are now ready to put all the pieces together and solve the value of the Debye frequency ω_D and the integral of Eq. (4.2.35) giving us the mean energy of the system. First, we calculate the Debye frequency ω_D, and then proceed to calculate the mean energy and the subsequent heat capacity of the considered solid.

Because the total number of states of the system equals to N,[35] we can take use of the DOS in order to find the value of ω_D by calculating the below integral:

$$N = \int_0^{\omega_D} D(\omega)\,\mathrm{d}\omega\,, \tag{4.2.42}$$

which is analogous to any integral concerning density parameters.[36] By performing the

[34] We neglect here the fact that the velocity of sound should in general depend on the type of the acoustic phonon (longitudinal vs. transverse) and the direction of the wavevector $\mathbf{k}$.

[35] Specifically, total number of states N for a given polarization and degree of freedom.

[36] Compare this integral against a familiar one giving us the mass m of an object of volume V: $m = \int_V \rho(\mathbf{r})\,\mathrm{d}V$ where ρ is the density of the object.

integral, we find the Debye frequency to be

$$\omega_D^3 = 6\pi^2 v^3 N/V \,. \tag{4.2.43}$$

We are now ready to perform the integral of Eq. (4.2.35) governing the mean energy of the system. Using the Debye DOS of Eq. (4.2.41) the mean energy becomes[37]

$$\langle E \rangle = 3 \int_0^{\omega_D} \frac{V\omega^2}{2\pi^2 v^3} \frac{\hbar\omega}{\mathrm{e}^{\hbar\omega/k_B T} - 1} \,\mathrm{d}\omega \,, \tag{4.2.44}$$

which can be further simplified by rearranging the terms into

$$\langle E \rangle = \frac{3V\hbar}{2\pi^2 v^3} \int_0^{\omega_D} \frac{\omega^3}{\mathrm{e}^{\hbar\omega/\tau} - 1} \,\mathrm{d}\omega = \frac{3Vk_B^4 T^4}{2\pi^2 v^3 \hbar^3} \int_0^{x_D} \frac{x^3}{\mathrm{e}^x - 1} \,\mathrm{d}x \,. \tag{4.2.45}$$

The latter equality in the above follows from notation $x_D \equiv \hbar\omega_D/k_B T \equiv \theta/T$, which has been adopted for brevity. The parameter θ is known as the ***Debye temperature***, and is explicitly written to be given by

$$\theta = \frac{\hbar v}{k_B} \left(\frac{6\pi^2 N}{V} \right)^{1/3} \,. \tag{4.2.46}$$

Using the Debye temperature θ we can further clean up the integral for the mean energy resulting in the final form

$$\langle E \rangle = 9Nk_B T \left(\frac{T}{\theta} \right)^3 \int_0^{x_D} \frac{x^3}{\mathrm{e}^x - 1} \,\mathrm{d}x \,. \tag{4.2.47}$$

We can now proceed to find the relation for the heat capacity C for the solid. This is done by inserting the above Eq. (4.2.47) into our definition for the heat capacity, Eq. (4.2.29), and performing the differentiation with respect to temperature T. After successful application of the chain rule, the result turns out to be

$$C = 9Nk_B \left(\frac{T}{\theta} \right)^3 \int_0^{x_D} \frac{x^4 \mathrm{e}^x}{\left(\mathrm{e}^x - 1\right)^2} \,\mathrm{d}x \,. \tag{4.2.48}$$

This is our final result and is also known as the Debye heat capacity. The importance of this equation is, that it was the first successful model that correctly reproduced both the high temperature limit known as the ***Dulong–Petit*** value, as well as the low

[37] Note that we have dropped here the implicit dependency of ω on the wavevector $\mathbf{k}$. We are integrating over frequencies, so this dependency doesn't affect the calculation.

temperature cubic dependence ($\propto T^3$) that had been experimentally verified to occur for vast majority of materials. Although more accurate calculations exist,[38] the treatment of Debye is both elegant and effective in its surprisingly good accuracy.

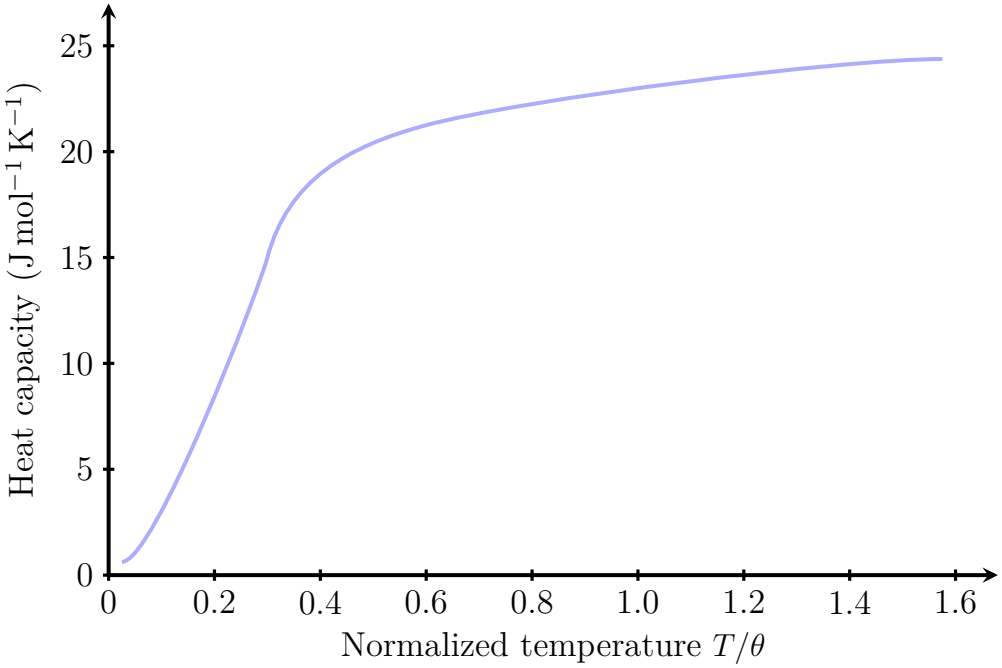

Fig. 4.14: Temperature-dependent heat capacity C of a solid, based on the Debye model. The temperature is normalized to the Debye temperature θ of the solid.

4.3 Thermal Conductivity

In this section we discuss the implications that the above mathematical treatment, particularly the Debye model, helped us form. In particular, we will consider how thermal energy is transported in the solid also as lattice vibrations. From the thermodynamic point of view this transferred energy is known as ***heat***. This process is described by the material property of thermal conductivity κ which we will soon introduce.

We start this section by pointing out that the above treatment of thermal capacities of solids results in remarkable agreement with some of the materials, and can for example perfectly explain the puzzlingly high thermal conductivity value $\kappa = 2300$ W/(m·K) of diamond.[39] However, the treatment is not yet adequate to completely describe the thermal properties of all types of solids. In particular, we are still entirely missing in

[38] Recall that we approximated the phonon dispersion relation to be entirely linear, which is a terrible assumption to make when generalizing the treatment to optical phonon branches and higher temperature ranges.

[39] We note here that diamond has a particularly high κ at the room temperature. What makes diamond even more amazing is that κ can be even further improved by using ^{12}C ***enriched diamond***, and by lowering the T of the system to cryogenic temperatures. At 104 K, ^{12}C enriched diamond boasts with $\kappa = 41\,000$ W/(m·K).

our treatment the possible, and in some cases quite important, contributions from free electrons of the solid. Especially in the case of metals, the effect of conduction electrons is usually dominant compared to contributions due to lattice vibrations, i.e. phonons. The treatment of free electrons in metals is left for the next Chapter. However, we will state here, that fortunately the above two contributions just simply add up in order to form the total thermal conductivity of a given material:

$$\kappa = \kappa_p + \kappa_e \,, \tag{4.3.49}$$

where κ_p (κ_e) is the thermal conductivity due to the lattice (electronic) contributions.

We will start our discussion by formally defining the meaning of the thermal conductivity κ of a given solid. For this purpose, we consider a rod-like solid with a cross-section of A (see Fig. 4.15). The first end is immersed in a heat bath associated with temperature T, while the other end of the rod is heated. The heating is assumed to be achieved by constantly heating the end with heating power of $\partial Q/\partial t$. After enough time has accumulated, the temperature profile of the solid stabilizes and the solid is said to be in a thermal equilibrium.

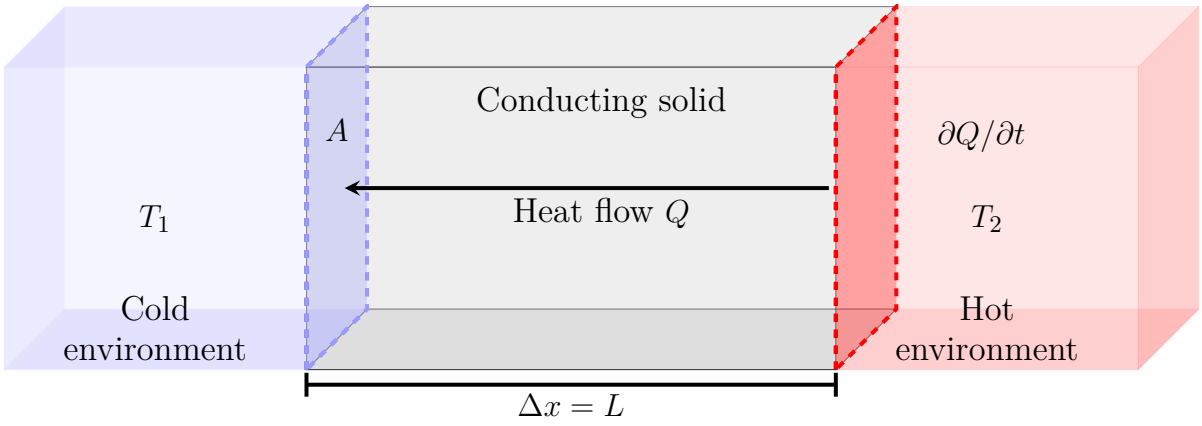

Fig. 4.15: Schematic for the heat conduction in a solid. The length of the solid is L, the cross-sectional area of the solid is A and the hot environment is heated with heating power of $\partial Q/\partial t$. The cold (hot) environment is associated with temperature T_1 (T_2).

In words, the thermal conductivity (coefficient) κ of a solid is defined as a proportionality constant between the ***flux*** of thermal energy j_E and the temperature gradient $\mathrm{d}T/\,\mathrm{d}x$

present in the solid, as given by[40]

$$j_E = \frac{1}{A}\frac{\partial Q}{\partial t} = -\kappa \frac{\mathrm{d}T}{\mathrm{d}x}\,, \tag{4.3.50}$$

The flux of thermal energy j_E is related to the energy transmitted across unit area per unit time, which fact gives us the second equality. Using Eq. (4.3.50), the thermal conductivity can be written to be[41]

$$\kappa = -\frac{1}{A}\frac{\partial Q}{\partial t}\frac{\mathrm{d}x}{\mathrm{d}T}\,. \tag{4.3.51}$$

The above equations define the thermal conductivity κ for a solid, and relates it with the amount of transferred thermal energy, i.e. heat Q, with the flux of thermal energy j_E. Note that many analogous parameters exist in physics, because the concept of ***flux*** is quite general in nature.

Next we proceed to qualitatively describe how this conduction of thermal energy can be understood to take place in a solid. Mathematically rigorous treatment from first principles would require some knowledge of statistical physics and ***kinetic theory of gases***, and is therefore beyond our scope.[42] In essence, we treat the excited phonons in the solid as energetic particles that can propagate through the solid. Although a single phonon is not affiliated with a large amount of thermal energy, physically meaningful solids are quite large in nature, in which case also the number of phonons present in the solid can be quite a large number. As a consequence, we treat phonons in the solid as a gas of particles which we call phonon gas. This treatment allows us to take use of the kinetic gas theory and write the associated (lattice) thermal conductivity as

$$\kappa_p = \frac{1}{3} C l_p v_p\,, \tag{4.3.52}$$

where C is the heat capacity of the solid, l_p the mean free path of the phonons and v_p the mean phonon velocity.[43] Here, we take the average phonon velocity v_p to be that

[40] More general definition would allow the thermal flux vector $\mathbf{j}_E$ to vary in the surface S perpendicular to the flow of thermal energy, i.e. the flux would be of form $\mathbf{J}_E = \iint_S \mathbf{j}_E \cdot \mathrm{d}\mathbf{A} = \iint_S \mathbf{j}_E \cdot \hat{\mathbf{n}}\,\mathrm{d}A = \partial\mathbf{Q}/\partial t$. In this case, the κ should be a second-rank tensor and $\frac{\mathrm{d}T}{\mathrm{d}x}$ be replaced with a proper gradient ∇T. Here, we simplify our notation and assume that the flux can be accurately enough described as a constant scalar resulting in equality $Aj_E = \partial Q/\partial t$, which is used in the second equality of Eq. (4.3.50).

[41] Note that our sign convention seem to differ from that of the treatment given in Hofmann's book. This difference doesn't affect the following general discussive treatment of the section.

[42] Interested reader is referred to the book *Introduction to Solid State Physics* by Charles Kittel.

[43] In order to facilitate the generalization of our treatment to vector quantities, we use the term '***velocity***' to describe the movement of the particles. Recall that velocity is associated with a direction, whereas 'speed' denotes merely the scalar amplitude of the velocity.

of the speed of sound in the solid. Because we already calculated the heat capacity C for the solid in the previous section, the only parameter left to understand is the mean free path of the phonons l_p, which will be discussed next.

In brief, many different ***phonon scattering*** mechanisms affect and limit the mean free path of phonons. Here, we will focus on two such mechanisms that are perhaps the most common ones. The first arises from scattering of phonons from lattice defects, and is known as ***phonon–impurity scattering*** while the second occurs due to ***phonon–phonon scattering***.[44] We continue our discussion with the former mechanism.

Interestingly, the propagation of phonons can be understood to be hindered by existence of lattice imperfections, i.e. defects, we discussed in Chapter III. Qualitatively, we can imagine that phonons, treated here as particles, can scatter from lattice defects, such as vacancies, interstitial defects, dislocations and grain boundaries. In the case of an idea monocrystalline solid, it is possible to reach a situation where the grain boundaries become the dominant source of scatter. However, most commonly occurring solids are not in this respect very pure. Instead, it is the other types of defects that also contribute to the mean free path of the phonons l_p.

The second mechanism limiting the thermal conductivities of materials is the phonon–phonon scattering. The process even becomes the dominant one at high enough temperatures. This can be understood by recalling that the mean phonon numbers associated with the solid depend on the temperature T of the solid. At high enough temperatures, many phonons are present in the solid at any given instance of time, making it increasingly probable for phonon–phonon scattering processes to take place.[45] Therefore, we see that the thermal conductivities κ of solids in general decrease when the temperature of the solid is increased. Furthermore, based on our earlier treatment of heat capacities C of solids, we can expect the thermal conductivities to also decrease at low enough values due to the decrease low-temperature behavior of C. As a consequence, materials are expected to have an optimal temperature range at which the thermal conductivities are maximized. These considerations are confirmed by looking at the measured thermal

[44] We note that such processes are actually nonlinear scattering processes, also known as ***Umklapp scattering*** processes. More details are found in book *Introduction to Solid State Physics* by Charles Kittel.

[45] We encourage an interested reader to picture this process also from the point of view where phonons are understood as collective wave-like excitations of the lattice, i.e. as small oscillatory perturbations of the atomic lattice from its original, near-zero temperature equilibrium position.

Table 4.2: Thermal conductivity κ for example metals and dielectrics at the room temperature. Note the exceptionally high value of κ for diamond. *This is the highest known value for a material at the room temperature.

Material	κ (W m^{-1} K^{-1})
Isotropically Pure Diamond	3320*
Diamond	2300
Copper	386
Aluminum	237
Steel	50
Quartz	10
Glass	0.8
Polystyrene	0.03

conductivity of silicon as a function of temperature $\kappa(T)$ presented in Fig. 4.16.

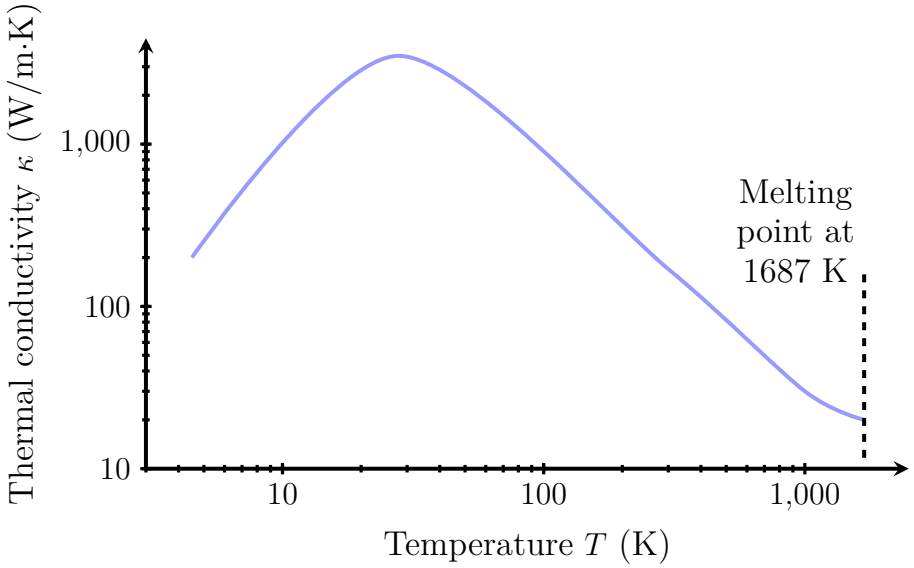

Fig. 4.16: Temperature-dependent thermal conductivity $\kappa(T)$ of silicon. Note that although the peak of the $\kappa(T)$ is quite high, the maximum occurs at relatively low temperatures. Therefore, at room temperature, the $\kappa(T)$ of silicon is relatively low.

Next we are ready to understand the very high thermal conductivity of diamond at the room temperature. First we note that due to the exceptional stability of the crystalline structure of diamond, lattice defects do not often contribute much to the mean phonon length l_p. In addition, diamond exhibits very similar temperature dependence of the thermal conductivity κ than the structurally similar silicon (Fig. 4.16). An important difference between the temperature dependencies is that the peak value for $\kappa(T)$ of diamond occurs at much higher temperatures T than for silicon.

Table 4.3: Thermal expansion coefficient α for representative materials. Note the very low value of the alloy Invar.

Material	$\alpha\left(10^{-5}\ \mathrm{K}^{-1}\right)$
Pb	2.9
Al	2.4
Cu	1.7
Steel	1.1
Glass	0.9
Invar	0.09

4.4 Thermal Expansion

In this final part of or Chapter IV on thermal properties of lattices, we will use our current approach to understand how and why solids in general undergo ***thermal expansion***, because this phenomenon takes place in many places of our everyday lives. Thermal expansion of solids has been taken into account in essentially all larger construction projects we can imagine. In addition, the phenomenon is also highly relevant in the finer inner workings of many technologically relevant instruments, ranging from mechanical clocks to satellites.

Again, we leave the detailed treatment of the problem for future excursion on the topic, and focus here mainly on the general principles. Essentially, thermal expansion of a solid is an exemplary process that can be only explained by generalizing the treatment of lattice vibrations to include ***anharmonic contributions***.[46]

We start by defining the coefficient of thermal expansion α by using equation

$$\frac{\Delta l}{l} = \alpha \Delta T\,, \tag{4.4.53}$$

where $\Delta l/l$ corresponds to the relative change of the length l of the solid and ΔT is a small change in the temperature. The thermal expansion coefficient α is often quite small-valued quantity ($\sim 10^{-5}$), as is seen in Table 4.3 that presents values of α for representative materials. However, when the solid of interest is of moderate size, or when even minuscule changes in the size of the solid become important, thermal expansion of a solid becomes non-negligible.[47]

[46] This aspect is also the principal reason we consider the process of thermal expansion here. Although anharmonicity of an oscillator may often be treated as negligible, such an assumption is not always valid.

[47] Consider for example why electric power transmission cables seem so loose during summer, or the ***bi-metallic thermometers*** in saunas.

A particularly interesting material alloy to discuss here is ***invar***. These nickel–iron alloys have some of the smallest thermal expansion coefficients α of all metallic compounds, and their discovery at the end of the 19$^{\text{th}}$ century resulted in breakthroughs in many precision measurements of that time. In fact, the inventor, Charles Édouard Guillaume was a sole receiver of the ***Nobel Prize in Physics in 1920*** for this discovery and his research on the anomalous thermal behavior of invar.

In order to understand the microscopic origins of thermal expansion, we turn our attention next on the interatomic potential ϕ. We recall from our earlier treatment (Chapter II) that ϕ can written using the Taylor expansion as

$$\phi(r) \approx \phi(a) + \frac{\phi'(a)}{1!}(r-a) + \frac{\phi''(a)}{2!}(r-a)^2 + \frac{\phi'''(a)}{3!}(r-a)^3 + \dots \tag{4.4.54}$$

During our earlier treatment we concluded that because the restoring force is proportional to the gradient of ϕ ($\mathbf{F} = -\nabla\phi$), mainly the third term gives rise to the elastic behavior of solids. Here, we make this notion more accurate by stating that also the higher-order contributions in the above series can affect the mechanical properties of the solid. Particularly, non-zero higher-order contributions in the series make ϕ and the subsequent potential well asymmetric with respect to the equilibrium position.[48] This point has been shown in Fig. 4.17, in which examples of both harmonic and anharmonic potential ϕ have been illustrated.

Next we qualitatively discuss how an anharmonic potential can be understood to give rise to thermal expansion.[49] We must next acknowledge that the solid is in general kept at a finite, non-zero temperature T. Recalling the equipartition theorem, we can state that the mean energy associated with the oscillator should be $k_B T$.[50] At low enough temperatures, the displacement oscillations of the oscillator are small. In this case the mean displacement, that can be associated as the equilibrium position, is close to the zero temperature ideal equilibrium position. At high enough temperatures T, this situation is changed due to the anharmonicity of the potential ϕ. In this case, the mean displacement of the oscillator shifts to longer values, as has been schematically

[48] In particular, the higher-order terms that are odd.

[49] To the best of lecturer's understanding, it is still difficult to quantitatively describe the processes of thermal expansion in solids, meaning that such processes are still poorly understood. Because lecturer is very fond of nonlinear physics, particularly of nonlinear optics, he finds solace in this fact that nonlinear physics seems in general somewhat difficult.

[50] One half of the mean energy is associated with the kinetic energy, while the other half with the potential energy of the oscillator.

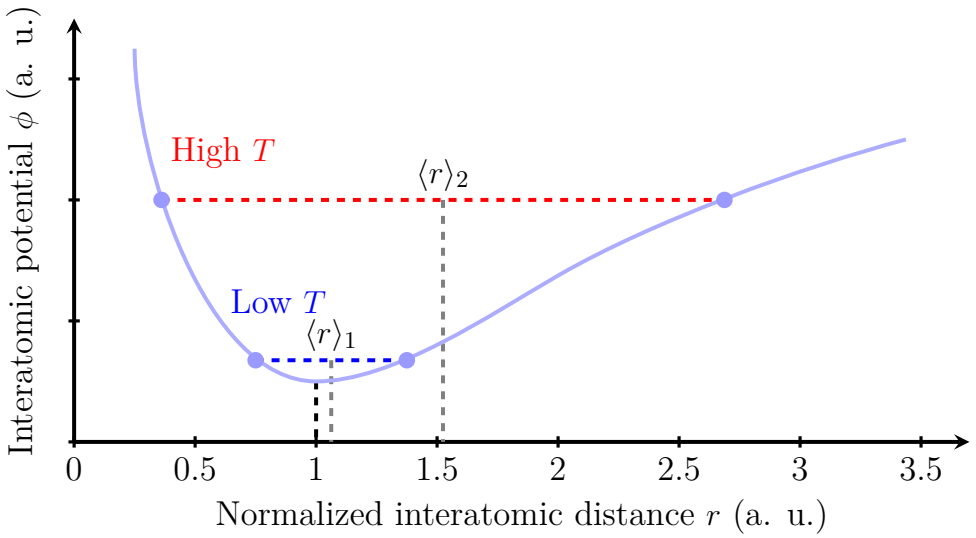

Fig. 4.17: 'Classical picture for the thermal expansion of a solid. The interatomic potential ϕ is shown as a function of interatomic distance. The gray line marks the temperature-dependent mean interatomic distance.'

shown in Fig. 4.17. Therefore we see that the interatomic distance a does in fact depend on the temperature T.[51]

We are now at the end of this Chapter that was devoted to understand the lattice contributions occurring in periodic lattices, and their role in the thermal properties of solids. Although this treatment provides us great understanding on how dielectric materials behave, we do not still know how another large class of materials known as metals could be understood to behave. Although the next Chapter focuses on the electronic properties of metals, it also elucidates to us these aspects on their behavior.

[51] We note that similar arguments could be used to qualitatively understand for example the ***piezoelectric effect***.

5. ELECTRONIC PROPERTIES OF SOLIDS: QUANTUM MECHANICAL APPROACH

In this Chapter, we generalize the concepts learned during the previous Chapter on the electronic properties of metals. In particular, our previous treatment focused on the ***Drude model***, that rested on the assumption that the electronic properties of metals are dominated by the responses due to the free electrons. We learned that many of the properties of metals can be understood by considering the model.

However, the Drude model has also its limitations. For example, the conductivities of metals at low temperatures were seen to be seriously and systematically underestimated. Furthermore, the Drude model fails entirely to describe the properties of alloys, that are an important class of metals from the engineering point of view. From the historical point of view, the biggest limitation of the Drude model was its disagreement with the ***Dulong–Petit*** values of many metals regarding their heat capacities. The puzzle was to understand why the free electrons did not seem to contribute much to the measured total heat capacities of solids. This mystery was only solved by generalizing the treatment to a quantum mechanical one, which will be outlined in this Chapter.

Before we start our rigorous treatment, we note that the outlined treatment can be further generalized in order to understand also the properties of non-metallic solids, mainly the semiconductors and dielectrics. These two important classes of solids will be treated in the upcoming Chapters.

5.1 Key assumptions in the Approach

In this Section, we outline the scope of our problem at hand and introduce the main assumptions and approximations we make in order to make the problem solvable.

In general, the problem at hand would be to find the quantum mechanical eigenstates associated with all the conduction electrons in the solid. In principle, we should be able to solve a many-body ***Schrödinger equation***, that would provide us with both the (many-body) electron ***wave functions*** and their associated energy eigenstates.[1] However, this problem is next to impossible, because in general the solution depends on the coordinates of all the electrons as well as all the ions.[2] We therefore need to simplify

[1] We note here that although the treatment of such many-body electron states is beyond the scope of our course, such states are paramount for some materials and related phenomena. In this case, one often speaks of ***strongly correlated materials***.

[2] In practise, this kind of stochastic problems are often solved by resorting to so-called ***mean-field theories***. Simplistically put, such theories try to find good approximations for the associated (effective) potential fields

our problem by resorting to approximations.

The first approximation we make is known as the ***Born–Oppenheimer approximation***. This approximation is quite general in nature, and is often used to simplify also other quantum mechanical problems. This approximation is also commonly used in other fields of physics.[3] In this approximation we assume that due to the huge differences in the masses of the electrons and the ionic cores, only the electrons are in noticeable motion. In other words, the ionic cores are assumed to be entirely stationary, or at least very slow, when compared against the quickly moving electrons.

The second assumption we make is that we neglect effects arising from electron–electron interactions.[4] This assumption allows us to restrict our treatment to wave functions that describe single electron states, in other words we do not consider correlated electrons. In this case, the effects due to other electrons and ions can be described by using an effective potential function $U(\mathbf{r})$,[5] allowing us the write a ***time-independent Schrödinger equation*** for our single-electron states as

$$-\frac{\hbar^2\nabla^2}{2m_e}\psi(\mathbf{r}) + U(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r})\,, \tag{5.1.1}$$

where E is the energy of the associated wave function $\psi(\mathbf{r})$. We will later see, that in many cases the potential $U(\mathbf{r})$ can be quite weak implying that in such cases electrons do indeed behave almost as free.

Finally, we note that we continue our focus on crystalline solids. For such solids, the atomic lattices forming the system are periodic, which for the cases of wave functions

due to interacting nearby objects. Once such fields are found, one can then often considerably reduce the complexity of the many-body problem. For example, in our current problem a mean-field theory could be utilized to find an effective potential field $U(\mathbf{r})$ taking into account the forces that positively charged ions have on the electrons. Particularly, the effects arising from **all of the charged ions** present in the solid, and not just the nearest or few nearest neighboring ones.

[3]Recall for example the Hamiltonian for the hydrogen molecule.

[4]We note that the ***superconductive phases*** that many solids exhibit at low enough temperatures suggest that this assumption often breaks apart at cryogenic temperatures. This experimental finding implies that (unsurprisingly) the electrons do interact with each other, but just very weakly so. An order-of-magnitude estimate for the interaction strength can be calculated by noting that many materials become superconducting at temperatures $T < 10$ K, corresponding to energies of 8.6×10^{-4} eV.

[5]The problem is turned into the problem of finding a working effective potential $U(\mathbf{r})$.

and potentials turns into periodic boundary conditions given by

$$\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{R}) \,, \tag{5.1.2}$$

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \,, \tag{5.1.3}$$

where $\mathbf{R}$ is the lattice vector. Consequently, we can simplify the mathematical treatment considerably by using lattice-periodic functions, and their respective ***Fourier series*** descriptions.

Equipped with these assumptions and respective approximations we could then proceed to solve the possible time-independent wave functions ψ_i that are associated with specific energies E_i. In order to clarify our terminology, we note that because our time-independent Schrödinger equation can be described as an ***eigenvalue problem***, we can associate the found wave functions ψ_i as ***eigenstates*** of the system, while the associated energies E_i are seen as the ***eigenvalues***.[6] Once we have solved the single-electron eigenstates for the system, we can proceed by taking use of the Pauli exclusion principle and fill up all the states using all the N electrons of our system. At zero temperature ($T = 0$ K) this approach would give us the correct state occupancies. However, in practise the systems are always at non-zero, i.e. at finite, temperatures ($T > 0$ K). As a consequence, it makes sense to treat the state occupancy distribution as a statistical quantity. Because electrons are fermions, their occupancies are governed by the ***Fermi–Dirac statistics***.

Before being carried away by the above outlined approach, we will discuss first qualitatively how the approach allows us to introduce an important concept in solid-state physics known as an ***energy band*** or as an ***electronic band***. This concept will be discussed more thoroughly later on, but already at this stage it is good to mention that the ***electronic band structure*** of the solid dictates whether the solid is a metal, semiconductor, semimetal or a dielectric.

5.2 The Idea of Energy Bands

Let's recall our earlier discussion in Chapter II regarding the bonding and anti-bonding states of a molecule. In that case, we considered a simple system consisting of only $N = 2$ electrons. In the case of a real solid, consisting of N electrons, the number of possible energy levels increases dramatically. In order to understand what happens,

[6] We note that many other very useful properties arise from this connection. The solutions are for example linear.

we schematically show in Fig. 5.1 how the number of bonding and anti-bonding states increase as a function of the number of considered electrons. For a large number of electrons,[7] the states become so closely spaced that they start forming a quasi-continuum of states known as an ***electronic band structure***.

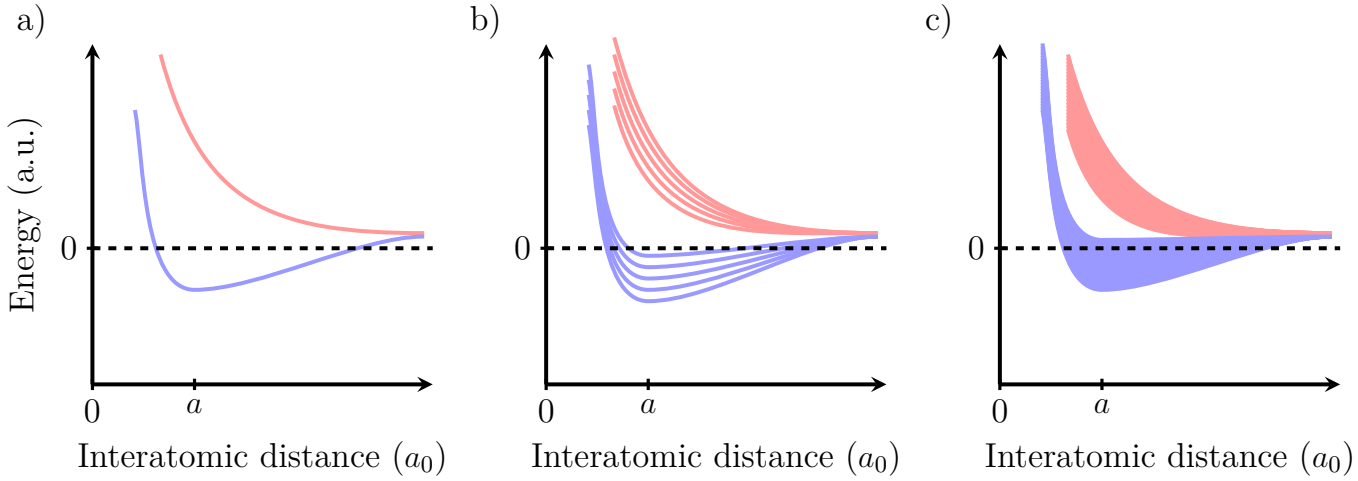

Fig. 5.1: Progression of band formation. a)–d).

For the case of real materials, the electronic band structures are admittedly often somewhat complicated to interpret during the first encounter. However, simplistic schematics of band structures can be drawn and understood already using our current knowledge, and have been plotted in Fig. 5.2. Looking at the figure we see that an energy level known as the Fermi energy E_F is an important quantity when the properties of solids are considered. What makes it such an important quantity is that it describes the level of electron state occupancy, which turns out to be a key defining feature which dictates whether a solid is a metal, semiconductor or a dielectric. If the Fermi energy E_F resides on a level where the solid exhibits an electronic band, then the solid will be a metal. If the E_F is inside a band gap, the solid is either a semiconductor or a dielectric.

5.3 Free Electron Model

The free electron model, also known as the ***Drude–Sommerfeld model***, is in essence the quantum mechanically correct version of the classical Drude model. The usefulness of the free electron model arises from its simplicity, allowing us to qualitatively understand the electronic properties of metals. In addition, the approach can be readily generalized to situations and materials where the electrons are not anymore entirely free, as we will learn during our next Section.

[7]Recall that a single cubic centimeter of solid matter already consists of $\sim 10^{22}$ electrons.

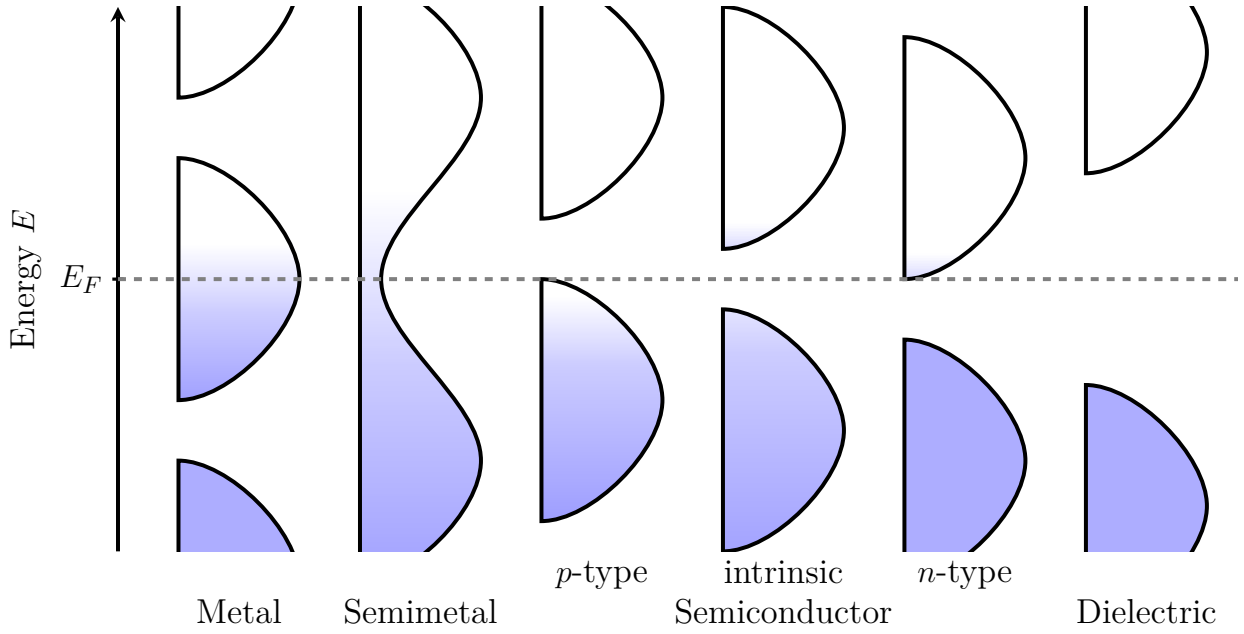

Fig. 5.2: Band picture of solids. Fermi energy E_F and the band structure of the solid dictate whether a solid is a metal, semimetal, semiconductor or a dielectric. If the Fermi energy E_F resides on an energy level where the solid exhibits an electronic band, then the solid will be a metal. If the E_F is inside a band gap, but the band gap is not very broad, the solid is either a semiconductor. For the case of broad band gap (> 2 eV), the solid is a dielectric.

In what follows, we will first calculate the quantum mechanical eigenstates associated with a free electron gas, restricted by periodic boundary conditions that our solid exhibits. We will also formally introduce the important concepts of Fermi energy E_F and the electronic density of states (DOS). Then we will use the developed approach to calculate the electronic contribution to the total heat capacity of the solid, discuss briefly of the historically important Wiedemann–Franz law and finally introduce to readers how applied electric fields are screened in metals.

5.3.1 Quantum Mechanical Eigenstates

We start by considering the standard quantum mechanical problem where a free particle is in a box. In practise, this means that the potential $U(\mathbf{r})$ is zero. In order to utilize this problem to understand the behavior of our periodic solid, we will again apply periodic boundary conditions to the above problem of solving the Schrödinger equation already encountered in Eq. (5.1.1). As before, we assume that the unit cell of our solid is a cube of side length L and volume $L^3 \equiv V$, resulting in the boundary conditions given by

$$\psi(\mathbf{r}) = \psi(x, y, z) = \psi(x + L, y, z) = \psi(x, y + L, z) = \psi(x, y, z + L)\,. \tag{5.3.4}$$

Leaving the gory details to a course focusing on quantum mechanics, the solution fulfilling the above conditions is found to be

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{V}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \,, \tag{5.3.5}$$

which is simply a plane wave propagating in the direction of the wavevector $\mathbf{k}$. Because of the imposed periodic boundary conditions, only discrete values of wavevector $\mathbf{k}$ of form

$$\mathbf{k} = \big(k_x, k_y, k_z\big) = \left(\frac{n_x 2\pi}{L}, \frac{n_y 2\pi}{L}, \frac{n_z 2\pi}{L}\right) \,, \tag{5.3.6}$$

are allowed. As before, n_x, n_y and n_z are integers. These above solutions (eigenstates) are affiliated with specific values of energy (eigenvalues) as given by

$$E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_e} = \frac{\hbar^2}{2m_e}\left(k_x^2 + k_y^2 + k_z^2\right) \,, \tag{5.3.7}$$

where m_e is the mass of the electron. Representative values of energy E have been plotted in Fig. 5.3 for a set of values n_x keeping $n_y = n_z = 0$.

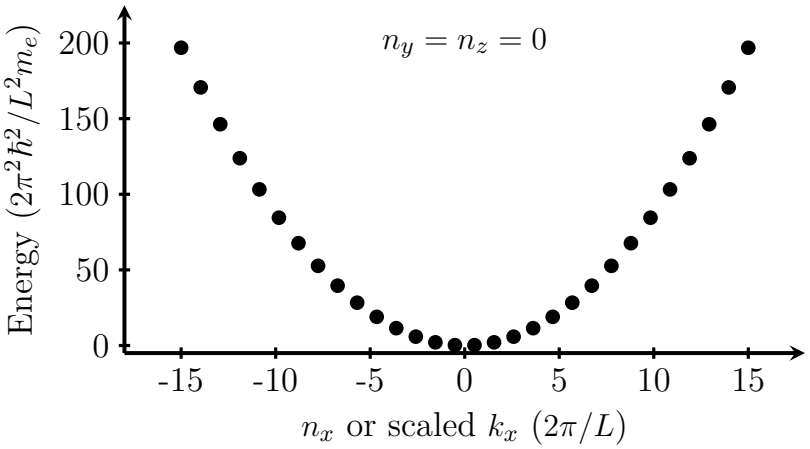

Fig. 5.3: Electronic states for a free electron in a box. as a function of integer n_x. For conceptual clarity, we restrict to energies with $n_y = n_z = 0$.

Looking at the above Eq. (5.3.7) we see that a single energy state fills a space of

$$\frac{\hbar^2}{2m_e}\left(\frac{2\pi}{L}\right)^2 \,. \tag{5.3.8}$$

Assuming the periodicity of our solid to be of the order of 1 Å, we see that the energy levels are indeed very finely spaced forming a quasi-continuum of states (i.e. an electronic band structure). Recalling the Pauli exclusion principle we can then start populating

these energy states of the system with our N electrons. Because two spin states can fit a single energy state, in total $N/2$ states need to be filled. Because the states are equidistantly and very closely spaced, the highest energy state $E_{\max}$ can be found to lie on the surface of a sphere of radius $n_{\max}$ (or equivalently $k_{\max}$). In this case, the below relation holds

$$\frac{N}{2} = \frac{4}{3}\pi n_{\max}^3 \,, \tag{5.3.9}$$

which can be solved for $n_{\max}$ resulting in

$$n_{\max} = \left(\frac{3N}{8\pi}\right)^{1/3} \,. \tag{5.3.10}$$

Now we can calculate the highest energy level $E_{\max}$ that is associated with the system. This energy level is better known as the ***Fermi energy*** E_F, and plays an important role in the band structure theories for solids, in particular for metals. The value of E_F is given by

$$E_F = \frac{\hbar^2 k_{\max}^2}{2m_e} = \frac{\hbar^2}{2m_e}\left(\frac{2\pi}{L}\right)^2 n_{\max}^2 \,. \tag{5.3.11}$$

By using values corresponding to typical metals, we see that E_F for metals is of the order of few electron volts. Recalling that we are currently focusing on solutions of the Schrödinger equation associated with a zero potential function $U(\mathbf{r})$, we see that the Fermi energy corresponds to the highest kinetic energy of the electrons present in the solid.[8]

Next, we can continue our efforts in order to calculate the DOS $g(E)$ associated with the free electron model (also known as the Drude–Sommerfeld model). Unlike for the case of lattice vibrations (Chapter IV), here we seek to give the electronic DOS as a function of the energy. In order to proceed, we take use of Eq. (5.3.10) in order to write Eq. (5.3.11) as

$$E(N) = \frac{\hbar^2}{2m_e}\left(\frac{3\pi^2 N}{V}\right)^{2/3} \,, \tag{5.3.12}$$

which essentially gives us the energy E of a state as a function of number of electrons N. The electronic DOS is defined as $g(E) \equiv \mathrm{d}N/\,\mathrm{d}E$. Therefore, after solving the

[8]It is worth to note here that we are treating the system at the ideal temperature of $T = 0$ K. The high values of kinetic energies are therefore solely due to the high number of electrons present in the system and the application of the Pauli exclusion principle in order to properly fill up the energy states.

above equation for $N(E)$:

$$N(E) = \frac{V}{3\pi^2}\left(\frac{2m_e E}{\hbar^2}\right)^{3/2}, \tag{5.3.13}$$

the electronic DOS $g(E)$ can be calculated by differentiating the above Eq. (5.3.13) with respect to N resulting in

$$g(E) = \frac{\mathrm{d}N}{\mathrm{d}E} = \frac{V}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2} E^{1/2}\,. \tag{5.3.14}$$

Note that the above relation will be used later in several occasions. The electronic DOS $g(E)$ holds for the temperature $T = 0$ K, and has been plotted in Fig. 5.4a. In essence, we see that the energy states are fully populated up until the Fermi energy E_F, where the the DOS abruptly falls to zero.

Next, we generalize our treatment to the more ordinary situations where the system is at a finite temperature ($T > 0$ K). In this case, some of the electrons of the system are thermally excited and will be at states that are associated with energies E above the Fermi energy $E > E_F$. Because electrons are fermions,[9] their occupation probability is governed by the Fermi–Dirac statistics, that can be taken into account by using the Fermi–Dirac distribution function given by

$$f(E,T) = \frac{1}{\mathrm{e}^{(E-\mu)/k_B T} + 1}\,, \tag{5.3.15}$$

where μ is known as the ***chemical potential***, T is the temperature and k_B the Boltzmann constant.

The real nature of the chemical potential μ is somewhat complex, but for metals the potential μ can be approximated to be equal to the Fermi energy ($\mu = E_F$). This can be seen perhaps most easily by writing a normalization condition for the system as

$$N = \int_0^\infty g(E) f(E,T)\,\mathrm{d}E\,, \tag{5.3.16}$$

which states that the total number of states N should be constant and related to the electronic DOS $g(E)$ and the Fermi–Dirac distribution $f(E,T)$ via the above integral. In

[9]This point is subtle. Precisely speaking, it is only the single-electron states that are fermionic. Electrons can also form many-body states, which may behave as bosons. For example, the quite fundamental phenomenon of ***superconductivity*** is often understood to arise from such a transition in the behavior of electrons.

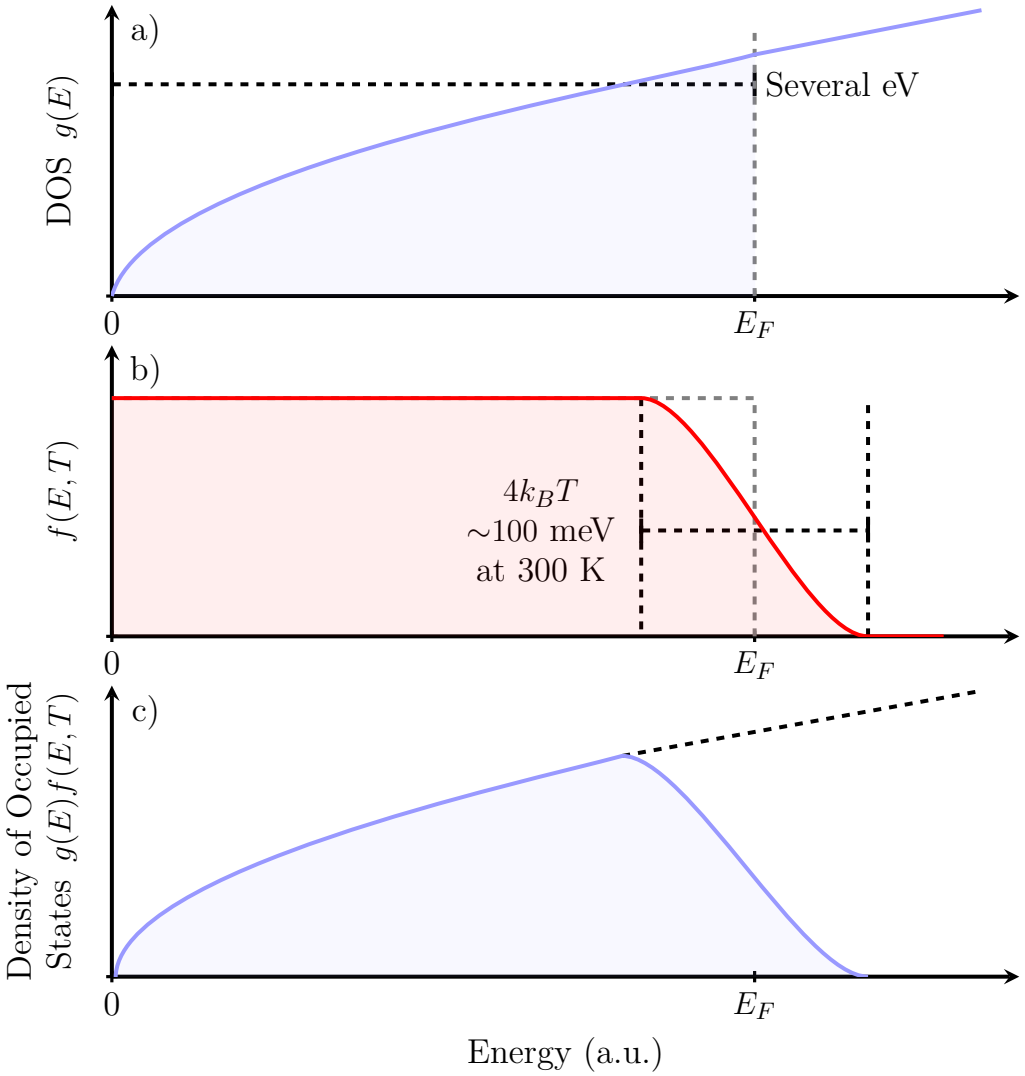

Fig. 5.4: Zero-temperature occupancies, finite temperature occupancies, DOS, Fermi–Dirac statistics.

other words, the above integral states that the total number of states N can be calculated by summing over all the occupied states. Looking at Eq. (5.3.16), we now see that at zero temperature the chemical potential is indeed equal to the Fermi energy. At finite temperatures the situation slightly changes. The range of the probability distribution $f(E,T)$ where the occupation probability drops from unity to zero is broadened to a width of $\sim 4k_BT$ (see Fig. 5.4b). However, close to room temperature this width is still very narrow ($\sim$100 meV) compared to the value of the Fermi energy (several eVs) validating our approximation that $\mu = E_F$ holds at all temperatures.

We have now all the ingredients to understand the electronic behavior of our system

consisting of free electron gas. Incidentally this model can be used to qualitatively understand the behavior of most metals. First, we clarify the main difference between our quantum mechanically correct Drude–Sommerfeld model and the entirely classical Drude model. In the classical Drude model the electron velocities (and therefore their associated energies) depend on the temperature T of the system. However, the temperature-dependence of the electron energies in the Drude–Sommerfeld model is quite weak, entering the governing equations only via the Fermi–Dirac distribution $f(E, T)$. This notion that the mean velocity of the free electrons is almost independent of the temperature T is fundamentally very different compared to the classical picture.

Second, we emphasize the importance of the region near the Fermi energy E_F in the description of the electronic behavior of metals. Recalling the Pauli exclusion principle, we can understand that the states far below the Fermi energy E_F are not very dynamic. In principle such states could be excited to higher energy levels. However, because the nearby levels have been already occupied, the excitation energies should be quite large (several eVs) in order to find an unoccupied state. Such energies could be easily introduced into the system optically, by shining light (or often UV radiation) into the material. However, such energies are not readily transferred thermally.[10]

The situation changes quite dramatically for states near the Fermi energy E_F. In this case, unoccupied states are abundantly available (see Fig. 5.4c) even for excitations associated with quite small transition energies of $\sim$100 meV. In this case, the electrons can be excited already by photons of thermal origin. In the upcoming section we will focus on this point of view, that allows us to understand the microscopic origins of the electronic contributions to the heat capacities.

5.3.2 Electronic Heat Capacity

In this section we use the introduced formalism to provide an order-of-magnitude estimate for the electronic contributions to the heat capacities of metals. In particular, we will find out that the contribution of the free electrons to the heat capacities is often quite small, and that the lattice contributions are often the ones that dominate the heat capacities of solids at moderate (i.e. at room) temperatures. This can be nicely seen by

[10] Recall the ***black-body radiation*** and the ***Planck's law*** close to room temperatures. The emissivity peaks near 17 µm (73 meV) and is already quite negligible at visible wavelengths. A rule of thumb is that objects at room temperature ($T = 300$ K) emit only a few visible wavelength photons per minute.

Table 5.1: Molar heat capacities of different metals at the boiling point of nitrogen and at room temperature. Note the close agreement of the room temperature values with the Dulong–Petit value of 24.9 J/K.

Material	77 (K)	273 (K)
Al	9.1	23.8
Au	19.1	25.2
Cu	12.5	24.3
Fe	8.1	24.8
Pb	23.6	26.7

looking at the tabulated molar heat capacities for different metals given in Table 5.1, where the room temperature values are seen to agree closely with the Dulong–Petit value of 24.9 J/K. We note, that at cryogenic temperatures the electronic contributions become often dominant.

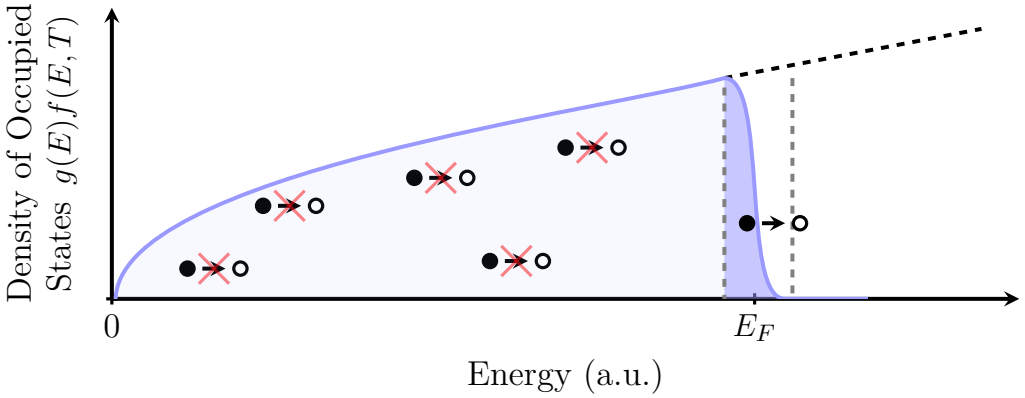

Fig. 5.5: Electron conductivity explained using occupied DOS.

We will next proceed by providing a back-of-the-envelope estimate for the electronic heat capacity C_e.[11] We start by stating that only the electrons near the Fermi energy E_F can participate to the heat capacity of the metal. The rest of the electrons simply cannot be thermally excited and therefore contribute to the heat capacity, because all the nearby states are already occupied by other electrons (see Fig. 5.5). The number of electrons that can participate to the heat capacity is then approximately given by $g\left(E_F\right) k_B T$. In three-dimensional case, each of these electrons exhibits three degrees of freedom and are therefore affiliated with a mean energy of $3/2 k_B T$. Combining these estimates, the mean energy for the relevant amount of electrons becomes

$$\langle E \rangle = \frac{3}{2} k_B T g\left(E_F\right) k_B T \,, \tag{5.3.17}$$

[11] Rigorous derivation of the electronic heat capacity is a nice homework.

which can be differentiated with respect to temperature T in order to find the heat capacity:[12]

$$C_e = \frac{\partial \langle E \rangle}{\partial T} = 3k_B^2 T g\left(E_F\right) \ . \tag{5.3.18}$$

Interestingly we see that the electronic contribution for the heat capacity is expected to depend linearly on the temperature T. This behavior is quite different from the contribution due to the lattice, which at low enough temperatures exhibited cubic dependence on T. This point has been clarified in Fig. 5.6, and suggests that the electronic contribution could become dominant at low enough temperatures. Furthermore, we see that the electronic DOS near the Fermi energy E_F plays an important role in the electronic contribution for the heat capacity. This aspect nicely exemplifies how these two concepts, the Fermi energy E_F and the electronic density of states $g(E)$, are of importance in solid-state physics.

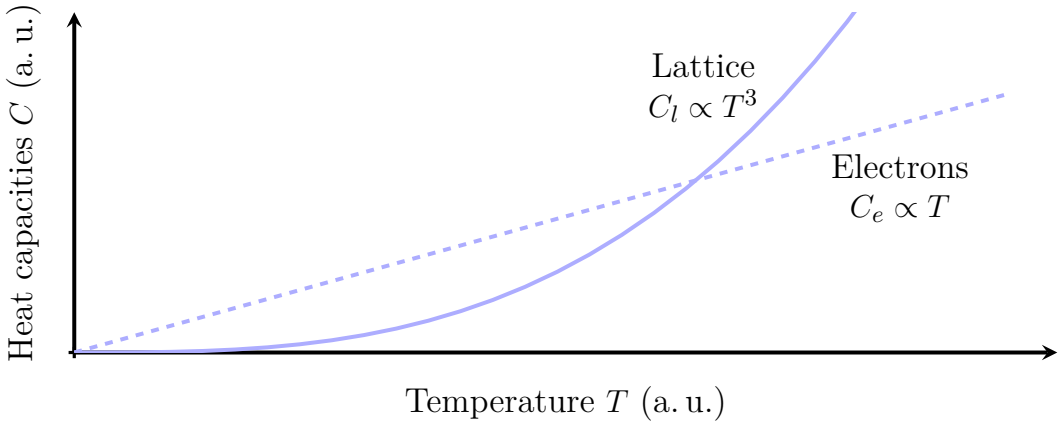

Fig. 5.6: At sufficiently low temperatures T, the lattice contributions to the heat capacity of solids C_l is a cubic function of T, whereas the dependency is linear for the case of the electronic contributions C_e.

5.3.3 Wiedemann–Franz Law

The original ***Wiedemann–Franz law*** was found during the nineteenth century by G. H. Wiedemann and R. Franz. Their experimental finding was that at a given temperature the ratio of the thermal conductivities κ and the electrical conductivity σ was found to be constant for all metals. Later, this constant was found to depend linearly

[12] This estimate is admittedly a bit rough, but results in surprisingly close agreement with the real value given by $C_e = \pi^2 k_B^2 T g\left(E_F\right)/3$.

on the temperature T resulting in an elegant equation of form

$$\frac{\kappa}{\sigma} = LT\,, \tag{5.3.19}$$

where the constant L is the Lorenz number, named to honor the discoverer ***L. Lorenz*** of the temperature dependence of the Wiedemann–Franz law. Derivation of the Lorenz number L is beyond our current scope, but follows straightforwardly from the Sommerfeld–Drude model. The result is given by

$$L = \frac{\pi^2}{3}\frac{k_B^2}{e^2}\,. \tag{5.3.20}$$

5.3.4 Screening of the Electric Field

In this section we will discuss how ***electric-field screening*** can take place in solids due to a presence of mobile charge carriers. In our current scope, we are particularly interested to understand how metals and semiconductors react to external electric fields, and aim to find a relation between the microscopic properties of the solid and the screening length, which quantifies how deep into the solid an external electric field can penetrate.

First, we explain qualitatively how electric-field screening takes place, and then proceed to the quantitative treatment. For conceptual clarity we consider here a metallic solid that exhibits a small region of positively charged particles acting as a material impurity. The volume of this region is V. This problem may seem very different from a case where an incident electric field is screened out in the surface of a conducting material, resulting in negligible field penetration into the material. Similar argumentation can be used, however, to understand also such situations.

Our impurity charge distribution $\rho_+(\mathbf{r})$ is associated with a Coulombic potential field $\phi(\mathbf{r})$ that in a dielectric environment is given by

$$\phi(\mathbf{r}) = \frac{q}{4\pi\epsilon}\iiint_V \rho(\mathbf{r}')\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\,\mathrm{d}V'\,. \tag{5.3.21}$$

where ϵ is the absolute permittivity of the environment,[13] $\mathbf{r}$ is the distance from the charge distribution and $\mathbf{r}'$ is the charge distribution coordinate. Locally, this potential

[13] Recall that the absolute permittivity of a material ϵ is related to the relative (dimensionless) permittivity ϵ_r via equation $\epsilon = \epsilon_r\epsilon_0$.

$\phi(\mathbf{r})$ can be seen to lower down the energy eigenstates of the nearby electrons by a factor of $e\phi(\mathbf{r})$,[14] as has been schematically illustrated in Fig. 5.7a. Consequently, it becomes energetically favorable that some of the far-away electrons present in the solid start drifting towards the impurity volume V in order to occupy the now free energy states that are associated with lower energies (see Fig. 5.7b). The number of drifted charges n can be approximately calculated to be

$$n \approx \int_{E_F - e\phi}^{E_F} f(E,T)g(E)\,\mathrm{d}E \approx g(E_F)e\phi(\mathbf{r})\,, \tag{5.3.22}$$

where we assume that the contribution from the Fermi–Dirac distribution $f(E,T)$ is negligible near room temperature, i.e. that $f(E,T) \approx 1$. In addition, we assume that the electronic DOS is constant over the integrated energy range. Using the above number of drifted charges, we can then proceed to write the associated charge distribution as

$$\rho_-(\mathbf{r}) = \frac{Q}{V} = \frac{-en}{V} = \frac{-e^2 g(E_F)\phi(\mathbf{r})}{V}\,, \tag{5.3.23}$$

This gives us the drifted charge density $\rho_-(\mathbf{r})$ as a function of the electronic DOS $g(E_F)$ and the potential function $\phi(\mathbf{r})$.

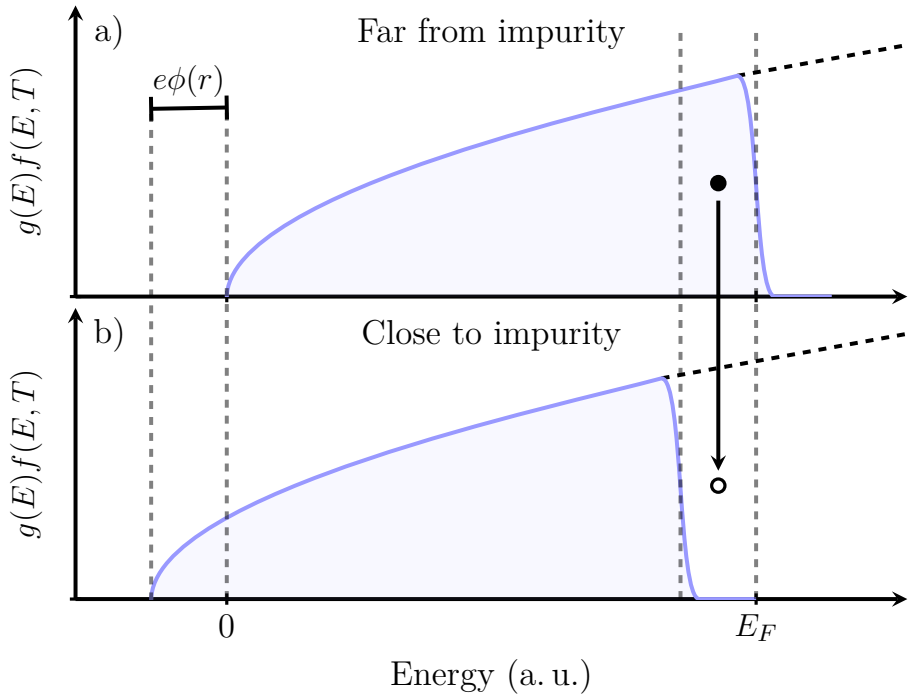

Fig. 5.7: Impurity effect on electronic DOS $g(E)$.

[14]Note that $e\phi(\mathbf{r})$ corresponds to the electric potential energy of the charge.

Next, we assume that the total potential $\phi_{\text{tot}}(\mathbf{r})$ felt by the charges away from the impurity is not strongly affected by the relocation of the charges. In this case, the two potentials are approximately equal $\phi_{\text{tot}}(\mathbf{r}) \approx \phi(\mathbf{r})$.[15] Then by noting that the total charge distribution is given by $\rho = \rho_- + \rho_+$, the potential is governed by the ***Poisson's equation***, given by

$$\nabla^2\phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon} = -\rho_-(\mathbf{r})/\epsilon - \rho_+(\mathbf{r})/\epsilon\,. \tag{5.3.24}$$

Interestingly, looking at Eq. (5.3.23) we see that the term $\rho_-(\mathbf{r})$ depends on the potential $\phi(\mathbf{r})$. In this case, it is more appropriate to modify the above equation into form known as the ***screened Poisson equation***:[16]

$$\left(\nabla^2 - \lambda^2\right)\phi(\mathbf{r}) = -\frac{\rho_+(\mathbf{r})}{\epsilon}\,, \tag{5.3.25}$$

where the notation has been simplified using

$$\lambda^2 = \frac{e^2 g(E_F)}{\epsilon V}\,. \tag{5.3.26}$$

The details for finding the solution to the above differential equation are left for interested readers, but follows straightforwardly by using the methodology based on ***Green's functions***.

The Green's function associated with our problem is found to be

$$G(\mathbf{r}-\mathbf{r}') = \frac{e^{-\lambda|\mathbf{r}-\mathbf{r}'|}}{4\pi\epsilon|\mathbf{r}-\mathbf{r}'|}\,, \tag{5.3.27}$$

where a point-like source is taken to be at $\mathbf{r}'$. The final solution for the potential field ϕ can now be calculated by solving an integral of form

$$\phi(\mathbf{r}) = \iiint_V G\left(\mathbf{r}-\mathbf{r}'\right)\rho_+(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}' = \frac{1}{4\pi}\iiint_V \frac{\mathrm{e}^{-\lambda|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\rho_+(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'\,. \tag{5.3.28}$$

[15] Note here that rigorously speaking we should take the effect of the drifted charges into account also in our calculation of the original potential ϕ, which would subsequently change the relocation of drifted charges. This change should then be also taken into account in a new calculation of the potential ϕ', where the prime denotes the fact that this potential is slightly modified from ϕ. This approach leads to a recursion of calculations, that is terminated once the condition $\phi_{\text{tot}}(\mathbf{r}) \approx \phi'(\mathbf{r})$ has been satisfactorily fulfilled. Related terminology clarifies this subtlety by stating that the original potential $\phi(\mathbf{r})$ is not yet self-consistent, whereas $\phi'(\mathbf{r}) = \phi_{\text{tot}}(\mathbf{r})$ is a self-consistent quantity.

[16] Note the similarity of this equation with the ***inhomogeneous Helmholtz equation***. Essentially the only difference between the equations is the different sign in front of the screening term. As we well know from optics and electrodynamics, the eigensolutions of the Helmholtz equation are propagating plane waves.

In order to make our treatment more intuitive to grasp, we will next only consider a point-like source of charge q at origo. In this case, the below integral simplifies into

$$\phi(\mathbf{r}) = \frac{q}{4\pi\epsilon}\frac{\mathrm{e}^{-\lambda|\mathbf{r}|}}{|\mathbf{r}|}\,, \tag{5.3.29}$$

which essentially describes an evanescent potential, because $\lambda > 0$. It is conventional to describe the situation using the inverse value of the parameter λ, which is also known as the ***Thomas–Fermi screening length*** r_{TF}, given by

$$r_{\mathrm{TF}} \equiv \frac{1}{\lambda} = \sqrt{\frac{\epsilon V}{e^2 g(E_F)}}\,. \tag{5.3.30}$$

For usual metals, this screening length is of the order of 1 Å implying that the screened potential we have just solved decays much faster than the bare Coulomb potential. We conclude this section by looking at the above Eq. (5.3.30), which verifies to us that indeed we have found a relation that gives us the screening length associated with a conductive material as a function of the electronic density of states. This is quite an interesting result, because it can be readily applied to quantitatively understand the electric-field screening taking place in many metals and in semiconductors.

5.4 Nearly Free Electron Model

We are now in a stage where we can generalize our treatment in order to understand the properties of an even broader class of solids. So far, our approach has rested on the assumption that the electrons in the solid have been free, meaning that the potential U felt by these electrons has been taken to be zero. Although our previous section discussing the electric-field screening does justify our assumption of free electrons, the assumption is by no means always valid.

This section is devoted to develop our mathematical treatment to a stage that allows us to move our description beyond the free electron model. In essence, we will next develop a treatment where the electron moves in a periodic potential field U. In this case the form of the electron wavefunction ψ is given by the ***Bloch theorem***, which we will next introduce. After this we will introduce a model known as the ***Kronig–Penney model*** which allows to treat the electrons present in the solid as almost free. This improvement turns out to be the final missing piece required to complete our description. Using this model, we can understand almost all of the properties metals qualitatively starting from a microscopic description of solids.

5.4.1 The Bloch Theorem

We start this section by re-introducing to the reader the time-independent Schrödinger equation

$$-\frac{\hbar^2\nabla^2}{2m_e}\psi(\mathbf{r})+U(\mathbf{r})\psi(\mathbf{r})=E\psi(\mathbf{r})\,, \tag{5.4.31}$$

where $\psi(\mathbf{r})$ is the single-electron wave function, $U(\mathbf{r})$ the effective potential function and E the total energy associated with the wave function. The above is our master equation for this section. The solutions of this equation have provided us all the understanding we have formed so far, and will continue to do so in what follows. Up until this point we have studied the case of a zero potential $U(\mathbf{r})=0$, which is a restriction we now seek to remove.

The above restriction can be lifted by introducing the celebrated Bloch theorem, which in words is given by:

> The eigenfunctions ψ of the wave equation for a periodic potential U are products of a plane wave $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r})$ and a function $u_{\mathbf{k}}(\mathbf{r})$ that exhibits the periodicity of the crystal lattice.

Turning the above statement into an equation, the Bloch theorem turns into a condition

$$\psi_{\mathbf{k}}(\mathbf{r})=\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r})\,, \tag{5.4.32}$$

where $\psi_{\mathbf{k}}$ is a single-electron eigenfunction known as the ***Bloch function***, or Bloch state, and $u_{\mathbf{k}}(\mathbf{r})$ is a lattice-periodic function fulfilling

$$u_{\mathbf{k}}(\mathbf{r})=u_{\mathbf{k}}(\mathbf{r}+\mathbf{R})\,, \tag{5.4.33}$$

where $\mathbf{R}=n_1\mathbf{a}_1+n_2\mathbf{a}_2+n_3\mathbf{a}_3$ is the Bravais lattice vector. Alternatively, the two above equations can be combined resulting in:

$$\psi_{\mathbf{k}}(\mathbf{r}+\mathbf{R})=\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{r})\,, \tag{5.4.34}$$

which equation is sometimes also referred to as the Bloch theorem.

In order to appreciate the meaning of the Bloch theorem we hurry next to derive it. First, we note that the functional form of the wave function $\psi_{\mathbf{k}}$ needs to comply with the periodicity of our system. As before, we consider a cubic volume of side length of L,

which is the periodic unit cell forming our macroscopic solid. These periodic boundary conditions result in the discretization of the allowed wavevector values $\mathbf{k}$ as we saw in Eq. (5.3.6). Subsequently, any lattice-periodic wave function $\psi(\mathbf{r})$ can be written using a Fourier series as

$$\psi(\mathbf{r}) = \sum_{k_x} \sum_{k_y} \sum_{k_z} c_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{k}} c_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \,, \tag{5.4.35}$$

where the coefficients $c_{\mathbf{k}}$ define the form of the solution. Similar to the wave function, also the potential U of the system must be lattice-periodic, implying that the general form of the potential is given by[17]

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} \mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}} \,. \tag{5.4.36}$$

Because the potential U is a real-valued function, its Fourier series description is a ***Hermitian function***

$$U_{-\mathbf{G}} = U_{\mathbf{G}}^{*} \,. \tag{5.4.37}$$

We proceed by inserting the two above Fourier series descriptions into the Schrödinger equation resulting in

$$\begin{aligned}
&\left(-\frac{\hbar^2 \nabla^2}{2m_e} - E + U(\mathbf{r}) \right) \psi(\mathbf{r}) = 0 \,, \\
\Longrightarrow \quad &\left(-\frac{\hbar^2 \nabla^2}{2m_e} - E + U(\mathbf{r}) \right) \sum_{\mathbf{k}} c_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = 0 \,, \\
\Longrightarrow \quad &\left(-\frac{\hbar^2 \nabla^2}{2m_e} - E + \sum_{\mathbf{G}} U_{\mathbf{G}} \mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}} \right) \sum_{\mathbf{k}} c_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = 0 \,,
\end{aligned} \tag{5.4.38}$$

which at first sight may look formidable. However, the picture can be considerably simplified. First, we look at the kinetic energy term, i.e., the first term. Because the coefficients $c_{\mathbf{k}}$ do not depend on spatial coordinates, the Laplace operator acts only on

[17] Note here that equally valid notation would have been to use a symbol $\mathbf{G}'$ to describe the wavevector $\mathbf{k}$ associated with the electron wave function. The important point here is to understand that these vectors comply with the periodicity of the lattice.

the plane wave term resulting in

$$-\frac{\hbar^2\nabla^2}{2m_e}\sum_{\mathbf{k}} c_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{k}} -\frac{\hbar^2}{2m_e}c_{\mathbf{k}}\nabla^2\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$$
$$= \sum_{\mathbf{k}} -\frac{\hbar^2(\mathrm{i}k)^2}{2m_e}c_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{k}} \frac{\hbar^2k^2}{2m_e}c_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,. \tag{5.4.39}$$

Next we look at the potential term, i.e., the summation terms over the reciprocal lattice vector $\mathbf{G}$ and the wavevector $\mathbf{k}$:

$$\sum_{\mathbf{G}} U_{\mathbf{G}}\mathrm{e}^{\mathrm{i}\mathbf{G}\cdot\mathbf{r}}\sum_{\mathbf{k}} c_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \sum_{\mathbf{G}}\sum_{\mathbf{k}} U_{\mathbf{G}}c_{\mathbf{k}}\mathrm{e}^{\mathrm{i}(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}}$$
$$= \sum_{\mathbf{G}}\sum_{\mathbf{k}'} U_{\mathbf{G}}c_{\mathbf{k}'-\mathbf{G}}\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}\,, \tag{5.4.40}$$

where the latter summation is to be performed over the shifted wavevector $\mathbf{k}' \equiv \mathbf{G}+\mathbf{k}$. We now note that physically the shifted wavevector $\mathbf{k}'$ is equivalent to the original wavevector $\mathbf{k}$, because adding an integer multiple of the reciprocal lattice vector $\mathbf{G}$ does not result in any physical changes in the system. This realization allows us to change the summation index back to the original $\mathbf{k}$. Together with Eq. (5.4.39) this realization allows to simplify Eq. (5.4.38) into a form

$$\sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\left\{\left(\frac{\hbar^2k^2}{2m_e} - E\right)c_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}}c_{\mathbf{k}-\mathbf{G}}\right\} = 0\,. \tag{5.4.41}$$

Now we recall that plane waves form an ***orthogonal basis***. Therefore, the above summation can be zero only if each of the terms inside the brackets vanish giving us a relation known as the ***central equation***:[18]

$$\left(\frac{\hbar^2k^2}{2m_e} - E\right)c_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}}c_{\mathbf{k}-\mathbf{G}} = 0\,, \qquad \text{for every } \mathbf{k} \neq 0\,, \tag{5.4.42}$$

which is a regular algebraic linear system of equations and not the formidable differential equation we started with. The treatment of this whole system of equations giving us the full wave function $\psi(\mathbf{r})$ will be given in our next Chapter. Then we will for example

[18] It is interesting to see that the original Schrödinger equation, which is a differential equation, has been cast into an algebraic form. Similar mathematical tricks are often performed with many differential equations, such as when a wave equation is transformed into a ***Helmholtz equation***.

learn that the emergence of electronic band structures can be seen to arise due to the above central equation.

Next we focus on completing our derivation of the Bloch theorem. After successfully solving the above central equation, we can write for a Bloch function associated with a specific one wavevector $\mathbf{k}$ a Fourier series description (see Fig. 5.8) as[19]

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}} \,. \tag{5.4.43}$$

Re-arranging the exponential the above equation becomes

$$\psi_{\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{r}} \left(\sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}} \right) \equiv \mathrm{e}^{\mathrm{i}\mathbf{k}\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \,, \tag{5.4.44}$$

which is the form of the Bloch theorem. We are left to verify that the above defined

$$u_{\mathbf{k}}(\mathbf{r}) \equiv \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}} \,, \tag{5.4.45}$$

is indeed a lattice-periodic function, which is now easy via direct calculation. Let's recall that the Bravais lattice vector of our system is $\mathbf{R}$, in which case the lattice remains invariant under any translation of the lattice with $\mathbf{R}$. This means that in order for $u_{\mathbf{k}}(\mathbf{r})$ to be lattice-periodic, the equality

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) \,, \tag{5.4.46}$$

should hold for all $\mathbf{R}$. Taking use of Eq. (5.4.45) we see that

$$\begin{aligned} u_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) &= \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})} \\ &= \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{R}} = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}} = u_{\mathbf{k}}(\mathbf{r}) \,, \end{aligned} \tag{5.4.47}$$

where we have used the identity $\mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{R}} = 1$, which we have proved earlier during this course.[20] We have therefore proved that the Bloch theorem holds.

We conclude this section by discussing the physical significance of the Bloch theorem. First, we note that the Bloch theorem and the associated Bloch functions allow to

[19] We now seek to find a specific component of the total wave function. In this case, only the selected few $\mathbf{k}$ components are being summed over in the Fourier series. These selected components are the sequence given by $\{..., \mathbf{k}-2\mathbf{G}, \mathbf{k}-\mathbf{G}, \mathbf{k}, \mathbf{k}+\mathbf{G}, \mathbf{k}+2\mathbf{G}, ...\}$.

[20] Recall the exercise 2 of the HW1.

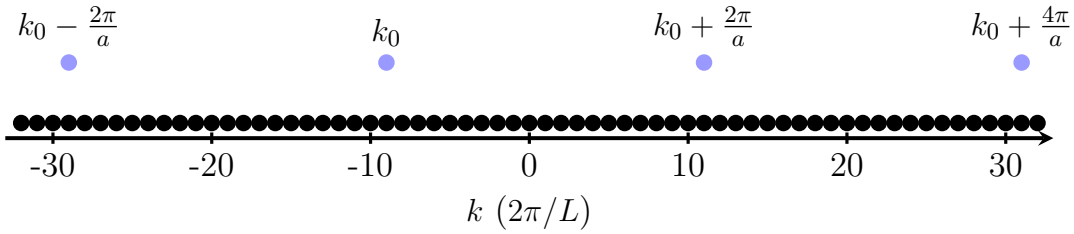

Fig. 5.8: One-dimensional illustration of the periodicity of the wavevector k, associated with a given Bloch function ψ_k. The wavevector values $k' = k_0 \pm N\frac{2\pi}{a}$, where $k = k_0 = -9\,(2\pi/L)$ (blue dots) for a periodic system consisting of N primitive cells, with a lattice constant a.

solve the single-electron wave functions associated with the Schrödinger equation for a periodic potential. Spectacularly, we see that the electron wave function ψ can be decomposed into eigenstates, here called as the Bloch functions, that closely resemble travelling plane waves. This realization justifies our earlier assumption where we simply assumed that the electrons can be treated as free. Furthermore, we even see that in this respect the electrons remain 'free' even in the case of a non-zero periodic potential. This result is very surprising, but helps us to improve our understanding of the electronic properties of solids.

Finally we discuss the meaning of the wavevector $\mathbf{k}$ associated with a given Bloch function. Looking at the above Eq. (5.4.34) we notice that we only need to multiply the Bloch function with a phase term $\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{R})$ in order to translate the Bloch function from $\mathbf{r}$ to $\mathbf{r}+\mathbf{R}$. The wavevector $\mathbf{k}$ is also the quantity that enters the conservation laws associated with electron scattering processes in the lattice.[21] Because of this fact, the quantity $\hbar\mathbf{k}$ is called the ***crystal momentum*** of an electron.

5.4.2 Kronig–Penney Model

In this section we will introduce the Kronig–Penney model, which provides us a simple enough example case where the central equation outlined in Eq. (5.4.42) can be analytically solved. Importantly, the model is a ***nearly free electron model*** that naturally gives rise to dispersive solutions, i.e., formation of electronic band structures and associated band gaps. These two concepts are very important in the description of electronic properties of solids.

[21] For example, we could write a momentum conservation law for a scattering process where an incident electron, associated with a wavevector $\mathbf{k}$ is scattered with a lattice phonon associated with a wavevector $\mathbf{p}$. In this case, the momentum is conserved up to a multiple of reciprocal lattice vector $\mathbf{G}$ resulting in $\mathbf{k}+\mathbf{p}=\mathbf{k}'+\mathbf{G}$, where the scattered electron state is associated with a wavevector $\mathbf{k}'$.

In order to demonstrate how the central equation can be utilized to solve problems associated with more complicated periodic potentials, we will first describe the Kronig–Penney model in the real space. Once the electronic band structures, i.e. the dispersion relations $E(k)$ for the single-electron wave functions have been found, we will then proceed to repeat the derivation in the reciprocal space using the central equation.

5.4.2.1 Kronig–Penney Model in Real Space

Here we derive the wave functions associated with the Kronig–Penney model. For conceptual clarity, we restrict our treatment to one dimension. We seek to find solutions that fulfill the one-dimensional Schödinger equation

$$\left(-\frac{\hbar^2}{2m_e}\frac{\mathrm{d}^2}{\mathrm{d}x^2}+U(x)\right)\psi(x)=E\psi(x)\,, \tag{5.4.48}$$

where $U(x)$ is our periodic potential and E the energy eigenvalue associated with the wavefuntion $\psi(x)$. The potential $U(x)$ is assumed to be rectangular and has been visualized in Fig. 5.9.

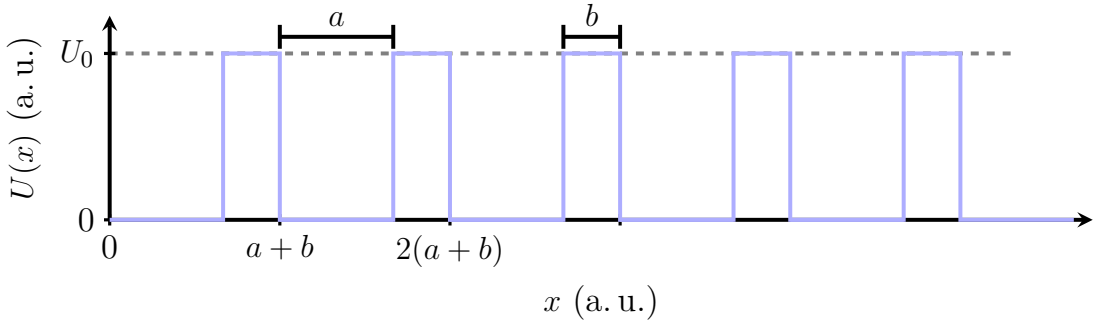

Fig. 5.9: The Kronig–Penney model and its periodic square-well potential $U(x)$. The height of the potential is U_0, width of the well is a and the width of the potential barrier is b.

From our earlier encounter to such (non-periodic) problem, we recall that the solutions in the area of the well $0 < x < a$ should be linear superpositions of two travelling plane waves as given by

$$\psi = A\mathrm{e}^{\mathrm{i}Kx}+B\mathrm{e}^{-\mathrm{i}Kx}\,, \tag{5.4.49}$$

where A and B are constant to be determined. The associated wavevector of the solution is K and can be related with the energy eigenstate through

$$E=\hbar^2K^2/2m\,. \tag{5.4.50}$$

In the region inside the well ($-b < x < 0$) the solutions are given by

$$\psi = C\mathrm{e}^{Qx} + D\mathrm{e}^{-Qx}\,, \tag{5.4.51}$$

where again C and D are constant to be determined, and Q is the associated (imaginary part of the) wavevector in this interval.[22] Next, we impose the periodic boundary conditions to our problem. In other words, we seek solutions that comply with the Bloch theorem and are of the form of Eq. (5.4.34). This implies that the solution should exhibit the following periodicity

$$\psi(a < x < a + b) = \psi(-b < x < 0)\mathrm{e}^{\mathrm{i}k(a+b)}\,, \tag{5.4.52}$$

where k is the wavevector associated with the Bloch function.

The constants A, B, C and D along with wavevector K can now be determined. We do so by noting that the solutions ψ along with their derivatives $\mathrm{d}^2\psi/\,\mathrm{d}x^2$ should be continuous at points $x = 0$ and $x = a$. For the case of $x = 0$, this results in a set of equations

$$\begin{aligned} A + B &= C + D\,, \\ \mathrm{i}K(A - B) &= Q(C - D)\,, \end{aligned} \tag{5.4.53}$$

while the point $x = a$ gives us

$$\begin{aligned} A\mathrm{e}^{\mathrm{i}Ka} + B\mathrm{e}^{-\mathrm{i}Ka} &= \left(C\mathrm{e}^{-Qb} + D\mathrm{e}^{Qb}\right)\mathrm{e}^{\mathrm{i}k(a+b)}\,, \\ \mathrm{i}K\left(A\mathrm{e}^{\mathrm{i}Ka} - B\mathrm{e}^{-\mathrm{i}Ka}\right) &= Q\left(C\mathrm{e}^{-Qb} - D\mathrm{e}^{Qb}\right)\mathrm{e}^{\mathrm{i}k(a+b)}\,. \end{aligned} \tag{5.4.54}$$

We have encountered such linear systems of equations already before, and know that a unique solution to the problem is only found when the determinant of the resulting coefficient matrix vanishes. After some straightforward algebra we end up with[23]

$$\left[\left(Q^2 - K^2\right)/2QK\right]\sinh Qb \sin Ka + \cosh Qb \cos Ka = \cos k(a + b)\,, \tag{5.4.55}$$

which doesn't yet look very transparent. However, by approaching a limiting case of $b = 0$ and $U_0 = \infty$, and using a notation $P \equiv Q^2ba/2$, we see that $Q \gg K$ along with $Qb \gg 1$. These assumptions allow us to rewrite the above equation as relatively understandable ***transcendental equation*** of form

$$(P/Ka)\sin Ka + \cos Ka = \cos ka\,, \tag{5.4.56}$$

[22] The term Q could be also called the damping constant.

[23] According to Kittel's book *Introduction to Solid State Physics* by Charles Kittel, the algebra is admittedly rather tedious.

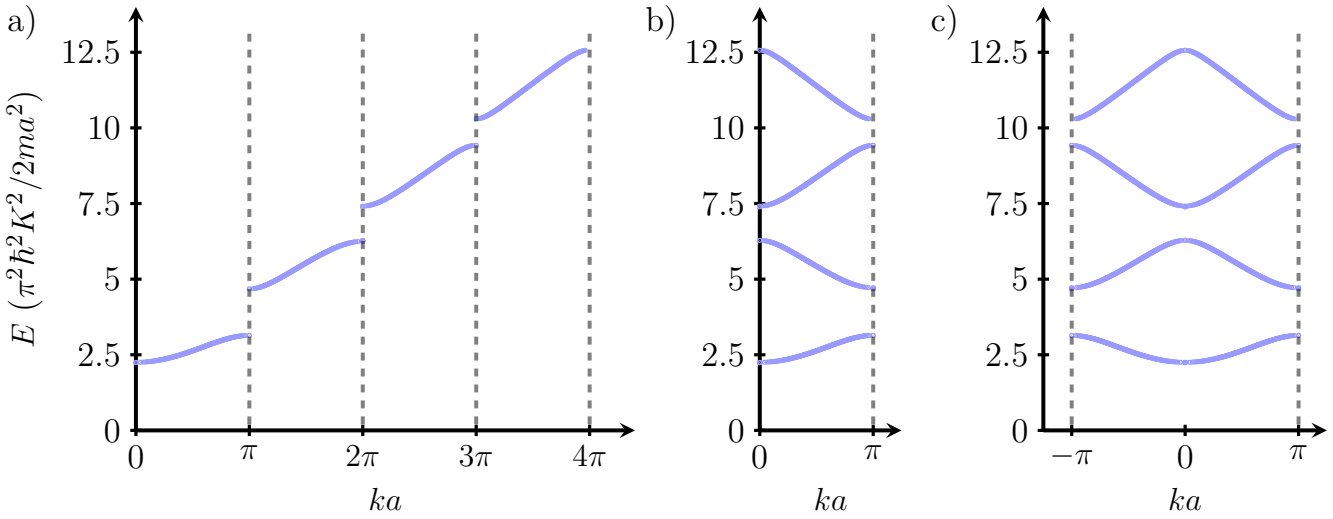

Fig. 5.10: The Kronig–Penney dispersion relation $E(k)$ and the emergence of band gaps at wavevector values $k = m\pi$, where m is an integer. a) Unfolded dispersion relation. Due to flexibility in choosing the k values, it is conventional to either b) fold the dispersion relation to (half of) the first Brillouin zone or c) plot the dispersion in the first Brillouin zone.

where the wavevector K is related to the energy of the state via Eq. (5.4.50). Therefore, we see that the above equation acts as a dispersion relation $E(k)$ associated with our problem and is our final result for this sub-section. The dispersion relation is visualized in Fig. 5.10a for the case of $P = 3\pi/2$.

5.4.2.2 Kronig–Penney Model in Reciprocal Space

In this section we use the central equation of Eq. (5.4.42) in order to solve the problem of having a single electron in a periodic rectangular well. In other words, we find a solution for the Kronig–Penney model using a reciprocal space description, validating to ourselves the usefulness of the approach.[24] The end result is to verify the dispersion relation $E(k)$ given by Eq. (5.4.56).

We start by writing the periodic potential function $U(x)$ as a Fourier series given by

$$U(x) = 2\sum_{G>0} U_G \cos Gx\,, \tag{5.4.57}$$

where $G = 2\pi/a$ is our reciprocal lattice vector and a is the lattice spacing of our periodic lattice. The main feature of the Kronig–Penney model is that the potential

[24] Note here that the reciprocal space approach could be used to solve Schrödinger equations with associated with periodic potentials of more complicated shape that the simple square-well potential treated here. Simple brute force approaches fail in this respect.

function is a periodic rectangular square well, in which case the above Fourier series turns into[25]

$$U(x) = Aa \sum_m \delta(x - ma)\,, \tag{5.4.58}$$

where A is a constant to be determined and m is an integer. The Fourier coefficients U_G of this potential are easy to find via direct calculation using the inverse transform[26]

$$\begin{aligned} U_G &= \frac{1}{a}\int_0^a U(x)\mathrm{e}^{-\mathrm{i}Gx}\,\mathrm{d}x = A\int_0^a \sum_m \delta(x-ma)\mathrm{e}^{-\mathrm{i}Gx}\,\mathrm{d}x \\ &= A\int_0^a \delta(x-a)\mathrm{e}^{-\mathrm{i}Gx}\,\mathrm{d}x = A\mathrm{e}^{-\mathrm{i}Ga} = A\mathrm{e}^{-\mathrm{i}\frac{2\pi}{a}a} \\ &= A\,. \end{aligned} \tag{5.4.59}$$

We are now ready to solve our one-dimensional version of the generally three-dimensional central equation of Eq. (5.4.42). In order to simplify our notation, we define $\lambda_k \equiv \hbar^2 k^2/2m$ giving us a one-dimensional central equation of form

$$(\lambda_k - E)\, c_k + A\sum_n c_{k-nG} = 0\,. \tag{5.4.60}$$

We proceed by adopting a clarifying short-hand notation

$$f_k \equiv \sum_n c_{k-nG} = f_{k-nG}\,, \tag{5.4.61}$$

where the latter equality arises from the periodicity of the coefficients c_k. We use the above notation to write the coefficient c_k as

$$c_k = \frac{-Af_k}{\lambda_k - E}\,. \tag{5.4.62}$$

Next we seek to cancel out the f_k from the above equation. This can be managed by writing an equation for the coefficient c_{k-nG}:

$$c_{k-nG} = \frac{-Af_{k-nG}}{\lambda_{k-nG} - E} = \frac{-Af_k}{\lambda_{k-nG} - E}\,, \tag{5.4.63}$$

where we have used the periodicity of f_k (i.e. $f_k = f_{k-nG}$) for the second equality. Next

[25]Note that a periodic distribution of Dirac functions is also known as the ***Dirac comb***.
[26]See also the Fourier coefficients associated with a Dirac comb.

we can sum both sides of the above equation over integers n resulting in

$$\begin{aligned} \sum_n c_{k-nG} &= \sum_n \frac{-Af_k}{\lambda_{k-nG} - E} \\ \implies \qquad f_k &= -\sum_n \frac{Af_k}{\lambda_{k-nG} - E} \\ \implies \qquad \frac{1}{A} &= -\sum_n \frac{1}{\lambda_{k-nG} - E}\,. \end{aligned} \tag{5.4.64}$$

Recalling the adopted notations for $\lambda_k \equiv \hbar^2 k^2/2m$ and wavevector K [see Eq. (5.4.50)], we can multiply both sides of the equation by term $\hbar^2/2m$ giving us

$$\frac{\hbar^2}{2mA} = -\sum_n \frac{1}{(k-nG)^2 - \frac{2mE}{\hbar^2}} = -\sum_n \frac{1}{(k-nG)^2 - K^2}\,. \tag{5.4.65}$$

Again, after some tedious but straightforward algebra given as a homework assignment, we can cast the above equation into a form[27]

$$\frac{mAa^2/2\hbar^2}{Ka} \sin Ka + \cos Ka = \cos ka\,, \tag{5.4.66}$$

which is our final result. Equating $P = mAa^2/2\hbar^2$ we see that the result is indeed identical to our earlier solution given by Eq. (5.4.56). Therefore, we see that the method based on the central equation and the reciprocal space (i.e. Fourier) description of the relevant functions works. Importantly, this methodology can be easily generalized to treat more complicated forms of potentials, allowing us to improve the accuracy of performed numerical calculations.

5.4.3 Three-dimensional Electronic Band Structures

We are now ready to generalize our treatment in order to understand the electronic properties of metals from the conceptually clear one-dimensional case to the more complicated three-dimensional case. As we have already seen for the case of lattice vibrations, known as phonons, it is enough to restrict our description only to the first Brillouin zone (see Fig. 5.10). The Brillouin zones beyond the first one can always be folded back to the first one simplifying the treatment. In general, the dispersion relation of the electrons, also known as the electronic band structure, generalize similarly

[27] Decompose the term in the sum using ***partial fraction decomposition***, take use of the relation for cotangent $\cot x = \sum_{n=-\infty}^{\infty} \frac{1}{n+x}$, and simplify further using known trigonometric identities. Getting somehow rid of the extra multiplier of 4 that riddles the readers of Kittel's book.

as the dispersion relations of phonons. The first Brillouin zone becomes a volume (the Wigner–Seitz cell of the reciprocal lattice), while the dispersion relations become four-dimensional quantities. As before, their visualization is somewhat non-trivial, and in the more general case relies on plotting ***isosurfaces*** of the dispersion relations.[28] A particularly important isosurface is known as the ***Fermi surface***, which in essence is the generalization of the concept of the Fermi energy E_F to the three-dimensional solid.

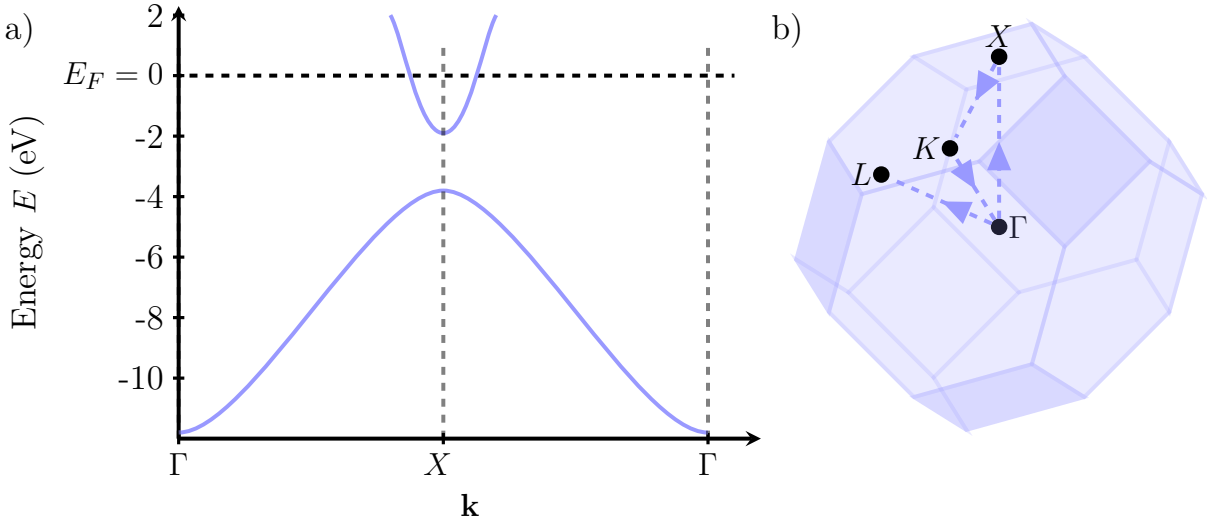

Fig. 5.11: a) Electronic band structure of aluminum near the X point exhibits a local band gap. b) The first Brillouin zone of Al and the associated critical points.

Interestingly, the nearly free electron model gave rise to formation of band gaps, which has been illustrated in Fig. 5.11a), which demonstrates the electron band structure near the X point of the First Brillouin zone. However, the concept of band gap becomes more complicated in the general case of three-dimensional solids. Although a band gap may form into the dispersion relation plotted along a single line in the Brillouin zone, it is not always the case that a band gap would remains in place when the whole dispersion relation is plotted. This point will be discussed more thoroughly in our next Chapter devoted to semiconductors and dielectrics, materials that in fact do exhibit such complete band gaps.

The symmetry properties of any particular solid remains unchanged whether we are dealing with phonons or electrons. Therefore, also in the case of electrons it is often enough to focus only on the dispersion near the critical points of the Brillouin zone. In Fig. 5.12 we plot the electronic band structure of aluminum near and between these

[28] We recommend an interested reader to check the Chapter 9 in *Introduction to Solid State Physics* by Charles Kittel.

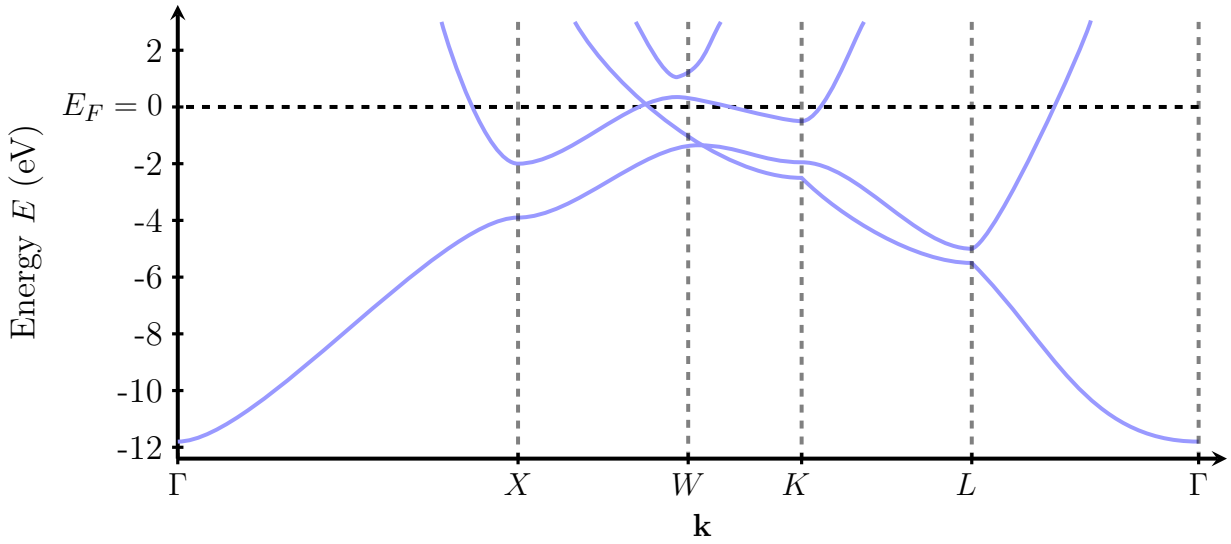

Fig. 5.12: Electronic band structure of aluminum near the critical points. Note that no complete band gap exists, because every wavevector value **k** is associated with at least a single energy state E. The Fermi level E_F marks the level of state occupancy for zero temperature $T = 0$ K.

points of high symmetry,[29] while Fig. 5.11b) reminds to the reader that the first Brillouin zone of Al is a truncated octahedron.[30] It is worth to note that no complete band gap is present in the band structure of Al. This is the main property that makes metals behave as metals, and should be contrasted against the full electronic band structures of dielectrics and semiconductors, which will be treated in our next Chapters. We also remind the reader, that in the beginning of this Chapter, a schematic of band structures has been plotted (see Fig. 5.2) clarifying exactly this point.

Before moving forward to describe the electron transport properties of metals, we note that in three-dimensional band structures of metals the Fermi level is usually set to zero $E_F = 0$ (Fig. 5.2). Recalling the meaning of the Fermi level E_F, we see this is the level up to which the electron states are occupied at zero temperature $T = 0$ K. At finite temperatures the Fermi–Dirac statistics smoothens the state distribution a little bit, but this fact doesn't usually noticeably shift the Fermi level (or more correctly the chemical potential μ).

[29] We note for the experimentally oriented readers that such data sets can be measured by performing ***angle-resolved photoemission spectroscopy*** (ARPES).
[30] Recall that Al exhibits a cubic fcc lattice.

5.4.4 Transport Properties

We can now take use of the formalism we have developed in order to understand how electric charges are transported in metals giving rise to electric currents. We note that similar considerations could be also formulated to describe the thermal transport properties of metals/solids, however we deem such considerations beyond our current scope. We will also restrict our treatment to a qualitative level, because the full description of the problem is quite complicated.

By recalling the Bloch theorem and the resulting Bloch functions describing the single-electron wavefunctions in a periodic solid, we see that the electrons can be expected to propagate freely inside the material. In a perfect periodic lattice, the electrons are not expected to scatter at all, implying to us that the electrical conductivity of such a material should be infinitely large. Because experiments with real materials don't show such ideal behavior, some electron scattering processes can be expected to take place in real solids. Similarly to the case of lattice vibrations (see Chapter 4), such scattering processes can be quantitatively taken into account by introducing a relaxation time τ, and will be more formally introduced at the end of this section.

We now develop a simple semi-classical mathematical framework to qualitatively understand ***electric conduction***. Similar to the case of lattice vibrations, instead of picturing electrons as delocalized wave-like objects (using Bloch functions) we treat them as particles propagating in the solid. This shift in our treatment can be justified by assuming that such particle-like states, or ***wave packets***, could be formed by proper superpositions of Bloch functions.[31] Because such wave packets necessarily consist of several wavevectors $\mathbf{k}$, the velocity of the pulse is given by the ***group velocity*** v_g of the wavepacket instead of the phase velocity v_p associated with the velocity of a single wave component.

We start by considering a particle associated with a velocity v_g propagating in an electric field $\mathcal{E}$.[32] The charge of the particle is assumed to be $-e$, in which case the present electric field $\mathcal{E}$ gives rise to a ***Lorentz force*** $F = -e\mathcal{E}$, that affects the motion of the

[31] Mathematically this picture is analogous to the case of pulsed light, which can be understood to consist of appropriate superpositions of time-harmonic waves, all oscillating at slightly different frequencies. Here, we don't work at the interface of the time–frequency domain, but rather in the space–momentum domain.

[32] Recall that we can easily relate the electrostatic electric field $\mathcal{E}$ to a voltage difference ΔV_{AB} across two points A and B. Under the electrostatic approximation, the relation is given by $\Delta V_{AB} = -\int_A^B \mathcal{E} \cdot \mathrm{d}\mathbf{s}$.

particle. When the particle moves a distance of s, the mechanical work W done for the particle is given by

$$W = Fs = -e\mathcal{E}s\,. \tag{5.4.67}$$

We can next calculate the rate of change of the mechanical work W, which in our case is in fact entirely in the form of kinetic energy E. By taking the time derivative of both sides we end up with

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -e\mathcal{E}\frac{\mathrm{d}s}{\mathrm{d}t} = -e\mathcal{E}v_g\,. \tag{5.4.68}$$

Next, we recall that the time derivative over the kinetic energy can be re-written using the chain rule into a form

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\partial E}{\partial k}\frac{\mathrm{d}k}{\mathrm{d}t}\,, \tag{5.4.69}$$

where the partial derivative term vaguely reminds us of the earlier relation found for the group velocity v_g for a wavepacket. Specifically, the group velocity is defined through[33]

$$v_g \equiv \frac{\partial\omega(k)}{\partial k} = \frac{1}{\hbar}\frac{\partial E(k)}{\partial k}\,, \tag{5.4.70}$$

which we can now combine with the above two equations in order to write

$$\begin{aligned} \frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\partial E}{\partial k}\frac{\mathrm{d}k}{\mathrm{d}t} &= \hbar v_g\frac{\mathrm{d}k}{\mathrm{d}t} = -e\mathcal{E}v_g \\ \implies \qquad \frac{\mathrm{d}k}{\mathrm{d}t} &= -\frac{e\mathcal{E}}{\hbar}\,, \end{aligned} \tag{5.4.71}$$

which tells us that the electric field $\mathcal{E}$ gives rise to a time-dependent wavevector k, and that the associated rate of change is proportional to the amplitude of the electric field.

Next, we proceed by introducing the concept of ***effective mass***, which turns out to be a very important concept in solid-state physics, particularly in the upcoming treatment of the electronic properties of semiconductors. This is most easily achieved by noting that the acceleration a of the particle can be related to the group velocity v_g via

$$a \equiv \frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial E(k)}{\partial k} = \frac{1}{\hbar}\frac{\partial^2 E(k)}{\partial k^2}\frac{\mathrm{d}k}{\mathrm{d}t}\,, \tag{5.4.72}$$

where the first equality takes use of Eq. (5.4.70), while the latter equality follows from the application of the chain rule. By using Eq. (5.4.71), the above can be further modified into

$$a = -\left(\frac{1}{\hbar^2}\frac{\partial^2 E(k)}{\partial k^2}\right)e\mathcal{E}\,. \tag{5.4.73}$$

[33] In the general three-dimensional case the group velocity would be given by $\mathbf{v}_g \equiv \nabla_{\mathbf{k}}\omega(\mathbf{k})\,.$

Now we note that by dividing both sides with the term in the parenthesis, the above equation turns into a classical equation of motion. In other words we can define the term

$$(m^*)^{-1} \equiv \frac{1}{\hbar^2}\frac{\partial^2 E(k)}{\partial k^2}\,, \tag{5.4.74}$$

as the inverse of our effective mass m^*. Although our derivation might seem abstract at first, it provides us interesting insights into the problem at hand. We can now understand that the electronic band structure affects the movement of electrons. Particularly, the curvature of the dispersion relation $E(k)$ affects the movement via the effective mass m^*. At regions where the band curvature is small, the effective mass is large and vice versa.[34]

We can now take use of classical Drude formula for the electrical conductivity, where we replace the rest mass of the free electron m_e with the effective mass. Particularly, we know that the classical conductivity is given by

$$\sigma = \frac{ne^2\tau}{m_e}\,, \tag{5.4.75}$$

where n is the electron density, τ is the relaxation time and m_e the rest mass of the electron. By making the replacement we end up with

$$\sigma = \frac{ne^2\tau(E_F)}{m^*(E_F)}\,, \tag{5.4.76}$$

where using notations $m^*(E_F)$ and $\tau(E_F)$ we have emphasized the fact that we specifically need to calculate the above values near the Fermi level E_F, because that is the relevant region.

We are now ready to understand qualitatively how charge transport takes place in a solid. First, we consider a solid that is not under an external electric field. This situation has been schematically shown in Fig. 5.13a which depicts that the electron states are filled up until the Fermi level E_F. Furthermore, the state distribution is symmetric with respect to the wavevector k. This situation is changed when an applied electric field $\mathcal{E}$ is present in the solid. The electrons start drifting as suggested by Eq. (5.4.71), making the state distribution asymmetric as has been visualized in Fig. 5.13b. After a given relaxation time τ, the drifted electrons relax back to available states. On average

[34]This classical approach fails in the extreme case of linearly-dispersive materials, such as for graphene. This point will be discussed later when we focus on 2D materials.

the electrons start moving in the direction opposite of the applied field giving rise to a macroscopic ***electric current***.

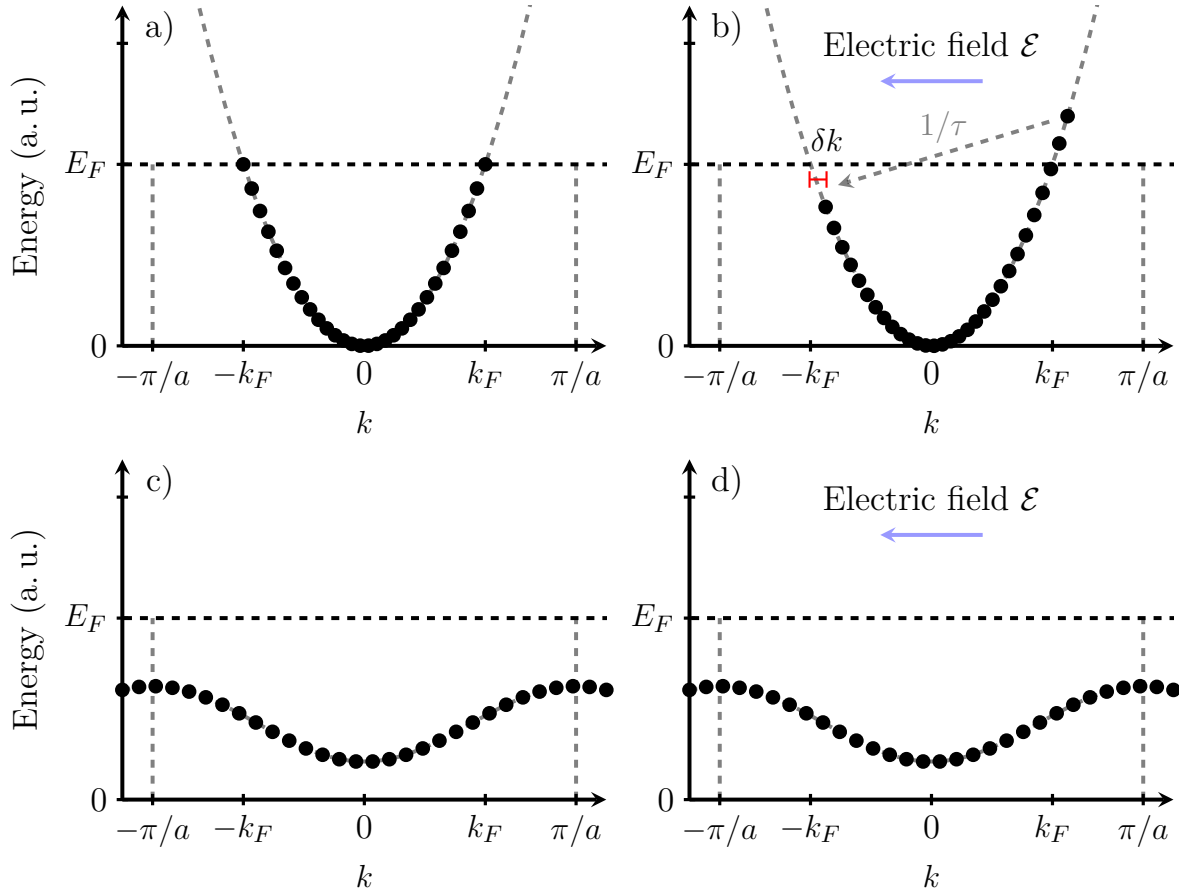

Fig. 5.13: Charge transport in solids. Circles denote filled electron states, reaching up until the Fermi level E_F. (a) A partially filled band under no applied electric field. (b) For a non-zero applied field $\mathcal{E} \neq 0$, the electron states start drifting. After accumulation of relaxation time τ the electron states relax back to free electron states. The asymmetric in the situation gives rise to an electric current. c) A fully filled band under no applied electric field. d) In the case of a non-zero applied field $\mathcal{E} \neq 0$, nothing happens in the first Brillouin zone, because electrons outside the Brillouin zone drift at the same velocity as the electrons inside the Brillouin zone. The situation remains fully symmetric and no macroscopic electric current is formed.

This picture allows us to also understand why solids with entirely filled bands are not good electrical conductors. The band structure of such a solid has been visualized in Fig. 5.13c. The relevant point here is that the situation is not physically changed under an applied electric field, which case has been visualized in Fig. 5.13d. The electrons will still drift in the filled band. However, the electrons beyond the first Brillouin zone move in similar fashion making the situation in the first Brillouin zone unchanged.

We conclude our qualitative discussion by highlighting that in this picture the key parameters are the effective mass m^* and the relaxation time τ of the charge carrier. Particularly, these parameters should be calculated near the Fermi level E_F, because that is the energy region where the relevant carriers reside (Fig. 5.13b). The role of the effective mass m^* will be discussed in more detail later, but at this point we already note that m^* allows to extend our treatment to ***charge carriers*** that are not simple electrons. For many materials, the electrical conductivity can be seen to be partly/mostly driven by movement of electron state vacancies in the electron bands that are almost filled. This also turns out to be the case for our example metal of aluminum, and provides us an interesting motivation to continue our treatment to understand the electronic properties of solids. Although the main principles are becoming clear, we notice that many things are not yet perfectly understood.

Finally we discuss the meaning of the relaxation time τ for the electric conductivity. Similarly to the case of lattice vibrations, the relaxation time is decreased by electron scattering processes, which fundamentally originate from the imperfections of the crystal lattice. In general, the total relaxation time τ can be estimated by using the ***Matthiessen's rule***

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{impurities}}} + \frac{1}{\tau_{\text{defects}}} + \frac{1}{\tau_{\text{lattice}}} + \frac{1}{\tau_{e-e}} + \cdots , \tag{5.4.77}$$

showing that the different contributions can be simply summed up. Again, a few main contributions can be highlighted, where the first two arise from lattice impurities and defects (impurity atoms, grain boundaries, dislocations etc.) and the associated electron scattering processes. The third contribution originates from lattice vibrations that make our initial assumption of the immobility of the ionic cores invalid.[35] The number of phonons present in the lattice depend strongly on the temperature T of the system, while the defects are virtually unaffected by the temperature. As a consequence, the impurity scattering usually dominates at low temperatures, while the ***electron–phonon scattering*** becomes the dominant process at moderate and high temperatures. Finally we note that at high enough electron densities n and/or incident electric fields ($\sim 10^5$ V/m) also electron–electron scattering processes may start taking place.[36] As with the case

[35] Recall that this assumption is also known as the ***Born–Oppenheimer approximation***.

[36] However, we note the electron–electron scattering processes are still quite weak. For metals the mean free paths for electron-electron collisions can be longer than 1 µm at room temperature and longer than 10 cm at 1 K. These distances are huge when compared against typical ionic distances of 1 Å.

of phonon–phonon scattering, after this regime has been reached, the electronic properties of solids become in principle nonlinear giving rise to new interesting family of processes.[37]

[37] Such is the nature of physics. We always start our treatment by assuming everything behaves linearly, and end up in allowing anharmonic contributions to take place in order to see effects the linear treatment fails to predict. Only in rare cases when we cannot use such ***perturbative description***, we need to resort to ***non-perturbative*** approaches.

6. SEMICONDUCTORS

Our earlier Chapter focused on introducing quantum mechanically sound treatment to understand the electronic properties of metals. This Chapter extends the introduced treatment to semiconductors and dielectrics. Although the difference between semiconductors and dielectrics is somewhat indeterminate, for conceptual clarity we will restrict our discussion on this Chapter only to semiconductors, while dielectrics will be discussed in the following Chapter.[1]

This Chapter introduces first some basic terminology regarding semiconductors. Then the discussion proceeds to the energy band structures, and particularly of the band gaps of semiconductors, focusing on the most important semiconductor of silicon (Si). Other notable semiconductors include germanium (Ge) and gallium arsenide (GaAs), but will not be convered much during our limited scope on the topic. We will also elucidate the role of the temperature on the conductivities of semiconductors, and how the properties of semiconductors can be engineered by semiconductor doping introducing to the reader the most fundamental concepts regarding these wonderful solids.

6.1 General Points Regarding the Terminology

From the macroscopic point of view it might sound feasible to categorize materials between metals, semiconductors and dielectrics based on their electrical conductivity. However, such efforts are futile because the differences in the conductivities of solids range almost 32 orders of magnitude! It is also worth to note here that the conductivity of a solid depends highly on its temperature. For the cases of metals and semiconductors the temperature-dependence is often quite dramatic, as we will see later on.

Instead, a working solution is to categorize solids based on their band gap E_g values, which vary only between 0–11 eV.[2] As we already saw in our previous Chapter, the band gap value E_g is the energy difference between the lowest energy state of the conduction band and the highest energy state of the valence band. These points are sometimes called the conduction band edge and valence band edge, respectively. The often used threshold value for the band gap (3 eV), which distinguishes solids between semiconductors and dielectrics is not arbitrarily chosen. This value corresponds to the case, when at room

[1] Semiconductors and dielectrics both exhibit a complete band gap. A common approach is to label materials with band gap values larger than ≈3 eV to be dielectrics. However, so-called ***wide-bandgap semiconductors*** are associated with band gaps as broad as 2–4 eV. Therefore such materials act as borderline cases between the conventional semiconductors and dielectrics.

[2] ***Magnesium fluoride*** (MgF_2) is a dielectric which exhibits a band gap of $E_g = 10.8$ eV. To the best of the author's knowledge, this is the largest E_g value. Consequently, MgF_2 is a highly transparent material down to very short UV wavelengths.

temperatures a small but measurable number of highest occupied electron states can be thermally excited to the lowest unoccupied states.[3] Representative semiconductor materials along with the band gap values E_g have been listed in Table 6.1.

The most common ***elemental semiconductors*** are silicon (Si) and germanium (Ge), and belong to the ***group IV***,[4] also known as the ***carbon group*** of the elementary table.[5] Many ***compound semiconductors*** are ***isoelectronic*** with Si and Ge. ***Silicon carbide (SiC)*** is a notable example of such a semiconductor consisting of elements from the same carbon group.[6] Other examples include semiconductors formed by combining a trivalent element from the group III with a pentavalent element from group V. Such semiconductors are known as the ***III–V semiconductors*** or just III–V's. The most notable examples are ***gallium arsenide (GaAs)*** and ***indium phosphide (InP)***.

Table 6.1: Band gaps of materials.

		E_g (eV)	
Crystal	Gap[a]	0 K	300 K
Diamond[b]	i	5.4	
Si	i	1.17	1.11
Ge	i	0.744	0.66
SiC[c]	i	2.4	2.36
InP	d	1.42	1.27
GaP	i	2.32	2.25
GaAs	d	1.52	1.43
Tc	d	0.33	–
HgTe[d]	d	−0.30	
CdS	d	2.582	2.42
CdSe	d	1.840	1.74
CdTe	d	1.607	1.44
Cu_2O	d	2.172	–

[a] i = indirect band gap, d = direct band gap.
[b] Diamond is here as a reference dielectric.
[c] Bands overlap, crystal is a semimetal.
[d] SiC forms many different compounds, where the E_g ranges between 2.4–7 eV.

Another important class of semiconductor compounds is formed by combining one element from the group II with another one from the group VI, giving rise to the ***II–VI semiconductors***. Exemplary II-VI semiconductors include cadmium telluride (CdTe) and cadmium sulfide (CdS). In general, II-VI semiconductors exhibit larger band gaps than III-V's making them principally more suitable for short wavelength (UV–VIS) optoelectronic applications.

[3] Recall the Fermi–Dirac distribution $f(T, E)$.
[4] The new convention calls this group as the group 14. However, it is the old nomenclature which makes sense with the following terminology regarding compound semiconductors.
[5] Kittel writes it with insight: '*Carbon gives biology, but silicon gives geology and semiconductor technology.*'
[6] Silicon carbide is another technologically very important material. Particularly, it is a very hard material and exhibits an extremely high sublimation temperature of ∼3000 K. Furthermore, it is a prime example of a wide-bandgap semiconductor (2.4 eV $< E_g <$ 7 eV).

The properties of above-mentioned binary semiconductors can be greatly altered by adding a third element into the compound. Such semiconductors are called ***ternary semiconductors***. For example, GaAs is a pure binary semiconductor. Because the bond lengths associated with Ga in the crystal are of similar length that what aluminum (Al) would exhibit, it is possible to form a ternary compound known as aluminum gallium arsenide (AlGaAs), whose chemical formula can be written as $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$, where $x \in \{0, 1\}$. The semiconducting properties of AlGaAs can be greatly modified by changing the chemical composition. For example, the value of the band gap E_g can be tuned between the range of 1.42 eV (GaAs) to 2.16 eV (AlAs). In addition, the type of the band gap can be changed from so-called ***indirect band gap*** into a ***direct band gap***, by using compounds with $x < 0.4$.

Final point to make regarding the terminology is that we call a semiconductor an ***intrinsic semiconductor***, if it only consists of the main constituents. If the semiconductor has been doped by adding additional elements to it, we call the solid a ***doped semiconductor*** or ***extrinsic semiconductor***. Unlike with ternary semiconductors, the amount of added dopants is small, and is not shown in the chemical formula.[7]

6.2 Intrinsic Semiconductors

We will first discuss the electronic properties of intrinsic semiconductors. Particularly, we will discuss the temperature dependence of the electronic properties. After this we move our discussion to the technologically more relevant doped semiconductors, also known as extrinsic semiconductors.

6.2.1 Electron States in the Conduction Band

We start by looking for the density of occupied electron states n in the conduction band. We label the energy at conduction band edge with E_C and this particular electronic DOS with g_e. At finite temperatures ($T > 0$ K) and at thermal equilibrium, the electron states are governed by the Fermi–Dirac distribution. The density n in a periodic cubic volume $V = L^3$ of side length L is then given as a familiar integral of form

$$n = \frac{1}{V} \int_{E_C}^{\infty} g_e(E) f_e(E, T) \, \mathrm{d}E \,, \tag{6.2.1}$$

[7] We note here that in practise even the so-called intrinsic or pure semiconductors always exhibit some amount of impurities. Technology to purify germanium has been developed to perfection. Even for ultrapure Ge semiconductors, the highest purity samples exhibit around one impurity atom per 10^{10} Ge atoms. The density of the Ge lattice is around 4.42×10^{22} cm^{-3}.

where $f_e(E,T)$ is the associated Fermi–Dirac distribution, known to be

$$f_e(T,E) = \frac{1}{\mathrm{e}^{(E-\mu)/k_BT}+1}\,. \tag{6.2.2}$$

The term μ is the chemical potential of the system while k_B is the Boltzmann constant.

The problem of Eq. (6.2.1) is that depending on the form of the electronic DOS g_e, it cannot be solved analytically.[8] Therefore, we need to resort to approximations. First, we will restrict our treatment to moderate temperatures, for which $(E-\mu) \gg k_BT$ is valid. In this case the Fermi–Dirac distribution $f_e(E,T)$ is approximately given by[9]

$$f_e(T,E) \approx \mathrm{e}^{(\mu-E)/k_BT}\,. \tag{6.2.3}$$

Next, we assume that the electron band structure near the conduction band is parabolic,[10] allowing us to find a simple dispersion relation for the associated energies

$$E(k) = E_C + \hbar^2k^2/2m_e^*\,, \tag{6.2.4}$$

where m_e^* is the effective mass of the electron. Equipped with the above assumption, and recalling that the wavevector amplitudes k are discretized for free-electron-like states as given by

$$k^2 = \left(\frac{2\pi}{L}\right)^2\left(\frac{3N}{8\pi}\right)^{2/3}\,, \tag{6.2.5}$$

where N is the total number of states. Inserting the above equation into Eq. (6.2.4) and solving for N, we can calculate the electronic DOS to be

$$g_e(E) \equiv \frac{\mathrm{d}N}{\mathrm{d}E} = \frac{V}{2\pi^2}\left(\frac{2m_e*}{\hbar^2}\right)^{3/2}\big[E(k)-E_C\big]^{1/2}\,. \tag{6.2.6}$$

We have now made the necessary simplifications to proceed in calculating the integral of Eq. (6.2.1):

$$\begin{aligned} n &= \frac{1}{V}\int_{E_C}^{\infty}\frac{V}{2\pi^2}\left(\frac{2m_e^*}{\hbar^2}\right)^{3/2}\big[E(k)-E_C\big]^{1/2}\,\mathrm{e}^{[\mu-E(k)]/k_BT}\,\mathrm{d}E \\ &= \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3}\int_{E_C}^{\infty}\big[E(k)-E_C\big]^{1/2}\,\mathrm{e}^{[\mu-E(k)]/k_BT}\,\mathrm{d}E\,. \end{aligned} \tag{6.2.7}$$

[8] Physisicts love analytical formulas and their derivation, because simple equations often provide superior intuition regarding the problem and its solutions. Numerical calculations provide unparalleled accuracy, but often lack in the intuition they can provide.

[9] For $(E-\mu) \gg k_BT$, the denominator turns into $\mathrm{e}^{(E-\mu)/k_BT}+1 \approx \mathrm{e}^{(E-\mu)/k_BT}$.

[10] For the case of real 3D solids the band structures can actually be quite complicated turning the calculations of effective masses m_e^* quickly non-trivial.

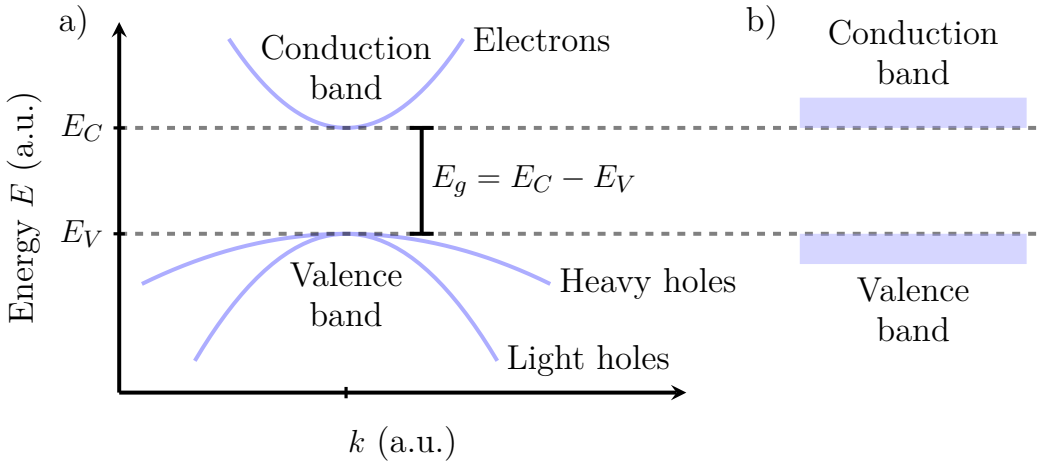

Fig. 6.1: a) The conduction bands and valence bands are often approximated to be parabolic. In many real semiconductor band structures, the Fermi surface of the valence bands near the relevant critical points are ellipsoidal giving rise to two effective masses for holes. Subsequently, the associated holes are often called light and heavy holes. b) In schematic representations, the bands are often simplified further and drawn as flat.

Substituting $X_g \equiv \left[E(k) - E_C\right]/k_BT$ into above equation, changing the variable of integration along with the limits of integration, we get

$$n = \frac{(2m_e^*)^{3/2}}{2\pi^2\hbar^3}\,(k_BT)^{3/2}\,\mathrm{e}^{(\mu-E_C)/k_BT}\int_0^\infty X_g^{1/2}\mathrm{e}^{-X_g}\,\mathrm{d}X_g\,. \tag{6.2.8}$$

The above integral is equal to $\sqrt{\pi}/2$, finally giving us our quantity of interest

$$n = \frac{1}{\sqrt{2}}\left(\frac{m_e^*k_BT}{\pi\hbar^2}\right)^{3/2}\mathrm{e}^{(\mu-E_C)/k_BT} = N_{\mathrm{eff}}^C\mathrm{e}^{(\mu-E_C)/k_BT}\,, \tag{6.2.9}$$

where

$$N_{\mathrm{eff}}^C \equiv \frac{1}{\sqrt{2}}\left(\frac{m_e^*k_BT}{\pi\hbar^2}\right)^{3/2}\,, \tag{6.2.10}$$

can be seen as an effective number density of states in the conduction band.

6.2.2 Electron Vacancy States (Holes) in the Valence Band

Next we repeat the above calculation for the electron vacancy states, which are also simply called as holes. In other words, we will calculate the density of existing hole states p in the valence band for a cubic volume $V = L^3$. We start with an integral very similar to the one we started our previous calculation. We label the energy at the valence band edge with E_V and the electronic DOS associated with holes with g_h. The Fermi–Dirac distribution $f_e(E,T)$ gives us the probability that the electron is in

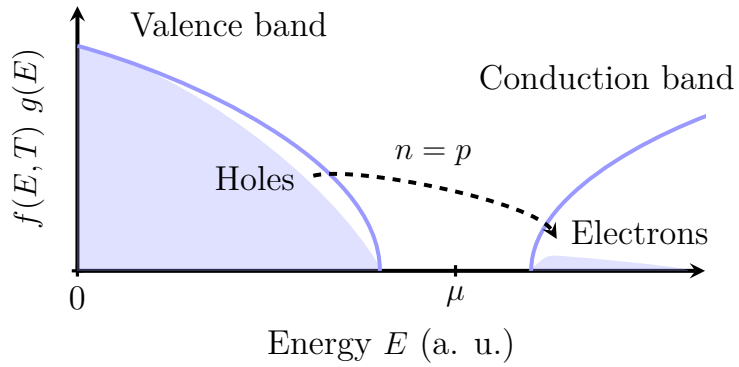

Fig. 6.2: Band occupations for intrinsic semiconductors.

a specific excited state. Here, we need to calculate the probability that the electron is not in a specific state $f_h(E,T)$, which is given by the complement probability given by $f_h(E,T) = 1 - f_e(E,T)$.[11] Combining the above considerations our integral of interest turns to be

$$p = \frac{1}{V}\int_{-\infty}^{E_V} g_h(E) f_h(E,T)\,\mathrm{d}E = \frac{1}{V}\int_{-\infty}^{E_V} g_h(E)[1 - f_e(E,T)]\,\mathrm{d}E\,. \tag{6.2.11}$$

Again, we are interested in the behavior of the system at moderate temperatures, in which case our Fermi–Dirac distribution turns into[12]

$$f_h(E,T) = 1 - \frac{1}{\mathrm{e}^{[E(k)-\mu]/k_BT}+1} = \frac{1}{\mathrm{e}^{[\mu-E(k)]/k_BT}+1} \approx \mathrm{e}^{[\mu-E(k)]/k_BT}\,. \tag{6.2.12}$$

We next assume that also the relevant part of the valence band is approximately parabolic resulting in a simple dispersion relation given by[13]

$$E(k) = E_V - \frac{\hbar^2 k^2}{2m_h^*}\,, \tag{6.2.13}$$

where m_h^* is the effective mass of the hole.[14] Using Eq. (6.2.5) the DOS associated with the holes turns out to be

$$g_h(E) = \frac{V}{2\pi^2}\left(\frac{2m_h^*}{\hbar^2}\right)^{3/2}\left[E_V - E(k)\right]^{1/2}\,, \tag{6.2.14}$$

[11] In other words, the electron can be in any other state than the one associated with the probability $f_e(E,T)$. This probability is the $1 - f_e(E,T)$.

[12] The Fermi–Dirac distribution can be simplified using $1 - \frac{1}{\mathrm{e}^x+1} = 1 - \frac{\mathrm{e}^{-x}}{1+\mathrm{e}^{-x}} = \frac{\mathrm{e}^{-x}+1}{\mathrm{e}^{-x}+1} - \frac{\mathrm{e}^{-x}}{\mathrm{e}^{-x}+1} = \frac{1}{\mathrm{e}^{-x}+1}$.

[13] Note that the derivation in Hoffman's book uses notation where $E_V = 0$ and $E_C = E_g$, because $E_g = E_C - E_V$.

[14] Note that in many cases the effective mass of the hole m_h^* can be made to include the details of the real band structure. Provided that the values for m_h^* and their possible dispersion is known, our simplistic treatment gives already surprisingly good and accurate results.

which can be inserted together with Eq. (6.2.12) into our integral of Eq. (6.2.11) resulting in

$$p = \frac{1}{\sqrt{2}} \left(\frac{m_h^* k_B T}{\pi \hbar^2} \right)^{3/2} \mathrm{e}^{(E_V - \mu)/k_B T} = N_{\mathrm{eff}}^V \mathrm{e}^{(E_V - \mu)/k_B T} \,, \tag{6.2.15}$$

where

$$N_{\mathrm{eff}}^V \equiv \frac{1}{\sqrt{2}} \left(\frac{m_h^* k_B T}{\pi \hbar^2} \right)^{3/2} \,. \tag{6.2.16}$$

We note here that the derived conduction electron concentration n and the hole concentration p both look almost like classical ***Boltzmann distributions***. The only difference is that the effective number densities, N_{eff}^C and N_{eff}^V show here temperature dependence.

6.2.3 Law of Mass Action

We now have the temperature-dependent conduction electron and hole concentrations for an intrinsic semiconductor. Next we will combine these relations in order to find a way to describe the electron (and the hole) concentration at a given temperature T without any information of the chemical potential μ, here taken to be synonymous to the Fermi level E_F.

We start by noting that the associated band edge energies are related by

$$E_g = E_c - E_v \,. \tag{6.2.17}$$

Using the above equality, the Eqs. (6.2.1) and (6.2.11) can be multiplied together resulting in

$$\begin{aligned} np &= \left(N_{\mathrm{eff}}^C \mathrm{e}^{\frac{\mu - E_C}{k_B T}} \right) \left(N_{\mathrm{eff}}^V \mathrm{e}^{\frac{E_V - \mu}{k_B T}} \right) = N_{\mathrm{eff}}^C N_{\mathrm{eff}}^V \mathrm{e}^{-E_g/k_B T} \\ \implies \qquad &= 4 \left(\frac{k_B T}{2\pi \hbar^2} \right)^3 (m_e^* m_h^*)^{3/2} \, \mathrm{e}^{-E_g/k_B T} \,. \end{aligned} \tag{6.2.18}$$

This relation is known as the ***mass action law***, and turns out to be a highly useful quantity in semiconductor physics. Importantly, the above equation shows that at given temperature T, the product of electron and hole concentrations does not depend on the location of the chemical potential μ.

Because in intrinsic semiconductors the number of conduction electrons is equal to

the number of holes,[15] we can take use of the above mass action law in order to find potential-independent relations for the carrier concentrations. These turn out to be

$$n_i = p_i = \sqrt{np} = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} (m_e^* m_h^*)^{3/4}\, \mathrm{e}^{-E_g/2k_B T}\,, \tag{6.2.19}$$

which are in general quite low for intrinsic semiconductors are moderate temperatures. Subsequently, intrinsic semiconductors are quite poor electrical conductors.

The concentration Eqs. (6.2.9) and (6.2.15) can be also used in order to find a relation for the chemical potential μ. By calculating their ratio, and solving for the potential μ, we find an equation of form[16]

$$\mu = \frac{E_g}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_h^*}{m_e^*}\right)\,, \tag{6.2.20}$$

where the chemical potential μ has been given as a function of the band gap energy E_g and the effective masses. Looking at the above equation, we can instantly verify to ourselves that at zero temperature, the chemical potential does indeed lie exactly at the center of the band gap. Furthermore, this relation allows us to calculate the location of the chemical potential μ also for finite temperatures ($T > 0$ K) once we know the values of the effective masses (see Table 6.2).

Table 6.2: Effective masses of semiconductors.

Material	m_e^*/m_e	m_h^*/m_e
Ge	0.60	0.28
Si	0.43	0.54
CdSe	0.13	0.45
InSb	0.015	0.39
InAs	0.026	0.41
GaAs	0.065	0.50

6.3 Doped Semiconductors

In this section we generalize our treatment to doped semiconductors. We learn that two different kinds of doping exist, and can be used to control the properties of semiconductors to an amazing degree. After the brief introduction to the interesting world of doped semiconductors, we proceed to derive the equations giving the carrier densities for doped semiconductors. Although these concepts are very general, we will restrict our scope to doping of silicon semiconductors.

6.3.1 Electron Donors and n Doping

Impurities in semiconductors can markedly affect their electronic properties. For example, at cryogenic temperatures (3 K) the conductivity of pure Si semiconductors can be

[15] This fact is sometimes also called the requirement for charge neutrality.
[16] Note that we are implicitly setting here $E_V = 0$, which gives us $E_C = E_g$.

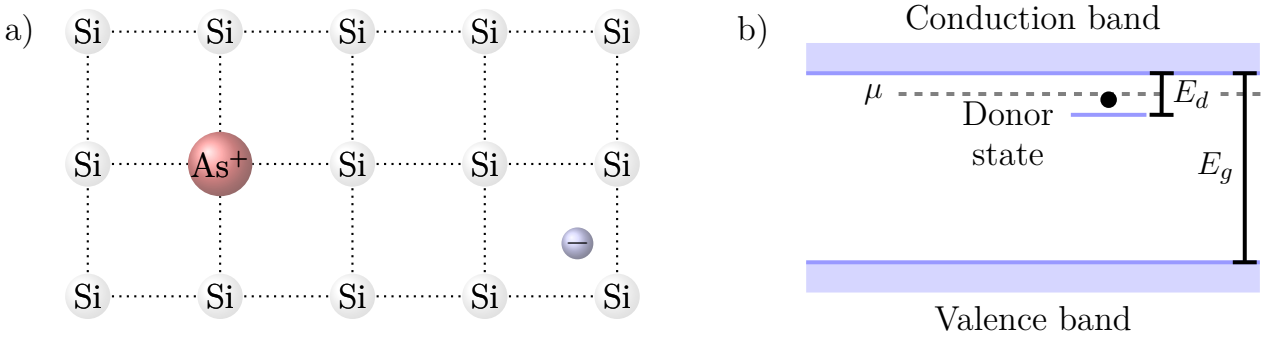

Fig. 6.3: a) Silicon can be n-doped using arsenic (As), which acts as an electron donor. b) 'The donor ground state is placed just below the conduction band minimum. Ionization of a donor atom corresponds to the transfer of the extra electron into the conduction band.'

increased thousand-fold by adding one boron atom to every 10^5 Si atoms. This sensitivity to impurities is remarkable. When the impurities are intentionally added into the semiconducting material, the process is called ***(semiconductor) doping***.

Silicon has a valency of four, subsequently crystallizing in the diamond structure. Recalling our Chapter II, the four neighbouring atoms in Si and their covalent bonds therefore are bonded in a three-dimensional tetrahedral configuration. For conceptual clarity, this configuration is often drawn simply as a planar configuration as shown in Fig. 6.3.

When a pentavalent impurity atom (e.g. arsenic, phosphorus or antimony) is added into Si crystal, it substitutes a normal Si atom. Doping of Si with one arsenic (As) atom is shown in Fig. 6.3a.[17] Because the valency of the doped atom (As) is higher than the valency of the host semiconducting material (Si), one valence electron remains free after the doped atom has accommodated the crystal. Such impurity atoms are called ***donors*** and the type of doping as ***n doping***.

We will next make a simple calculation in order to estimate the energies associated with the accommodation of the donor. Particularly, we aim to provide an order of magnitude estimate for the ionization energy, after which the extra electron indeed can be taken to behave as a free electron and thus occupy the conduction band of the semiconductor.

We start by noting that the situation of the accommodated impurity atom resembles to a degree the situation of a hydrogen atom. The four valence electrons of As have

[17]Experiments have shown that doped atoms do not favor forming interstitional defects.

formed covalent bonds with the neighboring Si atoms, and the accommodated As atom is left with a single positive charge, which binds the final electron to the As. The main differences between our situation and the case of a hydrogen atom is that our As atom is placed in a dielectric medium, and that the atom is in a periodic crystal. The latter gives rise to the fact that the electron mass should be replaced with an effective mass arising from the structure of the crystal.

For a quick calculation, it is adequate to use the classical ***Bohr model*** for a hydrogen. Taking use of the above modifications, the ionization energy for the As core turns into[18]

$$E_d = \frac{e^4 m_e}{2\left(4\pi\epsilon_r\epsilon_0\hbar\right)^2}\,, \tag{6.3.21}$$

where we have replaced the electron rest mass m with the effective mass m_e and the vacuum permittivity ϵ_0 with the permittivity of the medium $\epsilon = \epsilon_r\epsilon_0$.[19] The effective masses for Si are of the order of $m_e \approx 0.2m$, while the relative permittivity $\epsilon_r = 11.7$.

Table 6.3: Ionization energies for impurities in Si.[a]

Impurity	Type[b]	E_d	E_a
P	d	45	–
As	d	49	–
Sb	d	39	–
B	a	–	45
Al	a	–	57
Ga	a	–	65
In	a	–	157

[a] Energies are in meV.
[b] d = donor, a = acceptor.

With these changes in place, we see that the ionization energy is reduced dramatically by a factor of $m_e/m\epsilon_r^2 = 1.5 \times 10^{-3}$ from the case of a hydrogen atom. This reduction results in donor ionization energy of 20 meV. Admittedly, our calculation was simplistic. However, it is not much off from a proper calculation taking into account the band structure (and thus the quantum mechanics) of the situation. These rigorous calculations would give us an ionization energy of $E_d = 29.8$ meV for silicon, while the experimentally measured ionization energy is 49 meV. Other experimentally measured ionization energies for Si have been listed in Table 6.3, which are seen to agree quite well with the calculations.

The important point here is that the ionization energy E_d for Si is similar to the thermal energies associated with room temperatures ($k_BT \approx 26$ meV). This means that the donor electrons are quite readily ionized from the potential of the As core already at

[18] Compare with ionization energy of hydrogen $-e^4m/2\left(4\pi\epsilon_0\hbar\right)^2$, which results in energies of the order of $\sim$13.6 eV.
[19] We use notation ϵ_r for the relative permittivity in order to avoid confusion. Some authors simply use ϵ.

room temperatures, which explains the dramatic changes in electron concentrations and conductivities due to n doping.

6.3.2 Electron Acceptors and p Doping

Next we consider the case when an impurity atom of lower valency has been added to the semiconducting host crystal. In what follows, we will see that the impurity atom gives rise to an additional hole, that is loosely bound to the impurity. In this case, the impurity atoms are called ***acceptors***, and the type of doping as ***p doping***.

For the case of silicon, the impurity atoms acting as acceptors are trivalent, such as aluminum, boron, gallium or indium. Once the acceptor atom has been accommodated into the Si crystal, and extra electron vacancy state, i.e. an electron hole, is created. The situation has been clarified in Fig. 6.4a for the case of adding an Al atom into an otherwise pure Si. The process of hole formation (i.e. the ionization of the acceptor atom) requires energy. In energy diagrams, electron states receiving energy become elevated. The opposite happens for holes in the energy diagrams, which sink when gaining energy (see Fig. 6.4b).

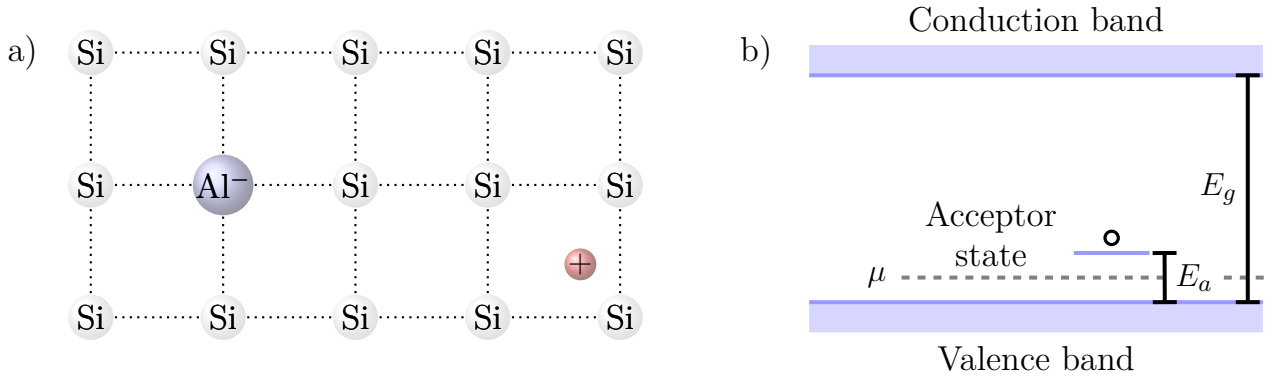

Fig. 6.4: a) Si can be p-doped using aluminum (Al), which acts as an electron acceptor. b) 'The acceptor level is placed just above the valence band maximum. Ionization of an acceptor atom corresponds to accepting (taking) an electron from the valence band and thereby to generating a mobile hole.'

Again we can provide an order-of-magnitude estimate for the ionization energy E_a by using similar argumentation as for the case of As donor atom.[20] The experimentally measured ionization energies for acceptor impurities in Si have been listed in Table 6.3, and are of the same order of magnitude in strength as for the case of donors. Therefore,

[20]There are complications in the calculation due to degeneracy in the hole states at the top of the valence bands, but such subtleties in the effective masses are deemed to be beyond our introductory level.

also by p doping a semiconductor its electronic properties and conductivities can greatly modified. Therefore, other considerations play a stronger role, when deciding which kind of doping is used when making specific semiconductor devices. We leave such considerations also beyond the scope of our current course.

6.3.3 Carrier Density

The calculations of the carrier densities are standard exercises of statistical physics and/or thermodynamics. We start by calculating the conduction electron concentration n, assuming first a case when no acceptors are present in the material. At low (cryogenic) temperatures $k_B T \ll E_d$, the concentration is approximately given by[21]

$$n \cong (n_0 N_d)^{1/2} \exp\left(-E_d/2k_B T\right) , \tag{6.3.22}$$

where $n_0 \equiv 2\left(m_e k_B T/2\pi\hbar^2\right)^{3/2}$ and N_d is the concentration of donors. Similar equation holds for acceptors, assuming that no donors are present.

The more general case of having a semiconductor doped both with acceptors and donors is somewhat complicated. In practise, one solves such problems numerically/iteratively. For such calculations, we need to take use of the mass action law (Eq. (6.2.18), $np =$ const). Increasing the number of donors results in increase of electron concentration n together with a decrease of the hole concentration p. However, the sum of concentrations $n + p$ increases. Assuming that the carrier mobilities are equal, increased amount of carriers results in increased electrical conductivity. In other words, in order to maximize the electrical conductivity, it is preferable to dope the system using either n- or p-type doping, over doping with both types simultaneously. As a consequence, a doped semiconductor usually exhibits either electrons or holes in great excess compared to the other carrier type. The more abundant carriers are called ***majority carriers***, while the less abundant ones are known as ***minority carriers***. In n-type semiconductors the electrons are the dominant carriers while the holes are the minority ones, and vice versa for p-type semiconductors.

The physical origin behind the mass action law is a process known as ***carrier recombination***, where a conduction electron state and a hole annihilate each other. The process is intuitively quite clear. An electron vacancy is filled by relaxation of a more

[21] Assume the system to be in chemical equilibrium. Maybe this could be a nice homework assignment?

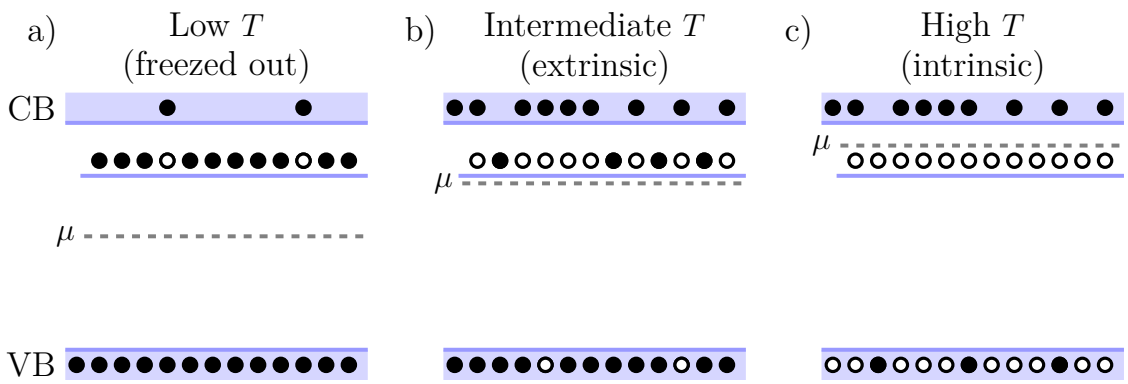

Fig. 6.5: Schematic electron density at different temperature levels for an n-doped semiconductor. a) At very low temperatures ($T < E_d/k_B$) both the valence band (VB) and the donor band are mostly occupied leaving the conduction band (CB) to be mostly unoccupied. b) At intermediate temperatures ($T \geqslant E_d/k_B$) most of donors have been ionized populating the CB with electrons. This region is known as the extrinsic region. c) At very high temperatures ($T > E_g/k_B$) the electrons in the VB start populating the CB. This region is known as the intrinsic region.

energetic conduction electron state. However, the carrier recombination is a dynamical process that can occur via several channels making its estimation beyond our scope.[22]

Before moving on in our treatment, we qualitatively discuss the temperature dependence of the charge carrier density. In essence, three different temperature ranges can be found, as have been illustrated in Fig. 6.5 for an n-type semiconductor. The first region is known as the freeze-out region and takes place at very low temperatures ($T < E_d/k_B$). In practise, for most doped semiconductors, this occurs well below the room temperature. In this region both the valence band (VB) and the donor band are mostly occupied leaving the conduction band (CB) almost empty of electrons.

The second region takes place at intermediate temperatures ($T \geqslant E_d/k_B$), at which most of donors have been ionized populating the CB with electrons. This region is known as the extrinsic region.[23] The final region corresponds to very high temperatures ($T > E_g/k_B$), at which also the electrons in the VB start populating the CB. This region is known as the intrinsic region.[24]

Understanding the origins behind these three different temperature regions provides

[22] Again, the process can be quantitatively described by using the concept of relaxation time. However, the devil is in the details to estimate the relaxation times associated with different mechanisms.
[23] Recall that doped semiconductors are also known as extrinsic semiconductors.
[24] Recall that undoped semiconductors are also known as intrinsic semiconductors.

us with a good intuition for the behavior of doped semiconductors. Furthermore, this picture provides us an intuition to understand how the chemical potential μ of the doped semiconductor system can be expected to behave as a function of temperature T. At low temperatures, the chemical potential μ is approximately in the center of the band gap energy E_g. At intermediate temperatures the potential is raised close to the donor state energies $E_g - E_d$, and further raised for high temperatures. Because a semiconductor device in a thermal equilibrium exhibits only a single chemical potential μ level (also called as Fermi level), the possibility to shift the position of the chemical potential μ by doping the semiconductor is an extremely important capability.

6.4 Conductivity of Semiconductors

6.4.1 Mobility

The electrical conductivity in semiconductors differs from that of metals mainly due to the possibility that both carriers participate on the process. For most of metals, the conductivity was mostly associated with the conduction electrons residing in the conduction band.[25] As a consequence, we need to modify our way to calculate the electrical conductivity σ. When both carriers play a role in the conductivity, it becomes useful to introduce a new quantity known as ***mobility***. The mobility is defined as

$$\mu \equiv \frac{|v|}{\mathcal{E}}\,, \tag{6.4.23}$$

where v is the drift velocity of the carrier and $\mathcal{E}$ an applied field. The total conductivity σ of the solid can then be written as

$$\sigma = (ne\mu_e + pe\mu_h)\,, \tag{6.4.24}$$

where μ_e (μ_h) is the electron (hole) mobility, e electric charge and n and p the electron and hole concentrations, respectively. Looking at the above equation, we see that both electrons and holes participate similarly to the overall conductivity of the solid.

Table 6.4: Carrier mobilities at room temperature.[a]

Crystal	μ_e	μ_p
Diamond	1800	1200
GaAs	8000	300
Si	1350	480
GaSb	5000	1000
Ge	3600	1800
InSb	800	450
PbSe	1020	930
InAs	30000	450
InP	4500	100
AgCl	50	–
AlAs	280	–
SiC	100	10–20

[a] Mobilities in units of $\mathrm{cm}^2/\mathrm{V{\cdot}s}$.

Further insight is gotten by noting, that the carrier mobilities μ can be connected with

[25] Notable and historically very interesting exception is aluminum, where holes dominate the electrical conductivity.

the associated relaxation times and effective masses as given by

$$\mu_e = e\frac{\tau_e}{m_e}\,, \qquad\qquad \mu_h = e\frac{\tau_h}{m_h}\,. \tag{6.4.25}$$

The above equations suggest that for most semiconductors a higher conductivity can be achieved by n doping. This occurs because the effective masses of electrons m_e are often smaller than the effective masses of holes m_h, which is attributed to the band structure of the solid.[26]

6.4.2 Hall Effect

> *Rowland now advised me, in repeating this experiment, to use gold leaf mounted on a plate of glass as my metal strip. I did so, and, experimenting as indicated above, succeeded on the 28th of October in obtaining, as the effect of the magnet's action, a decided deflection of the galvanometer needle. In short, the phenomena observed were just such as we should expect to see if the electric current were pressed, not moved, toward one side of the conductor.* –Edwin Hall, 1879

We will next introduce the ***Hall effect***, which can be utilized to measure the carrier concentrations in conducting solids. This effect was discovered by Edwin Hall in 1879, and has become a very important phenomenon in solid-state physics.[27]

We start by considering a situation, where a current $\mathbf{j}$ is flowing in a conductor. When placed into an external magnetic field $\mathcal{B}$, the flow of particles associated with a total charge of q and the macroscopic current $\mathbf{j} = q\mathbf{v}$ will feel a Lorentz force given by

$$\mathbf{F} = q\mathcal{E} + q\mathbf{v}\times\mathcal{B} = \mathbf{j}\times\mathcal{B}\,, \tag{6.4.26}$$

where the latter equality arises from the fact that no external electric field $\mathcal{E}$ is present in the solid. In order to simplify the vector notation, we will next restrict our treatment to a magnetic field that only points in the z-direction $\mathcal{B}_z$. In this case, the relevant component for the current points along x-axis as is marked by j_x (see Fig. 6.6).

The direction of the induced force $\mathbf{F} = F_y$ now follows from application of the right-hand rule, and is seen to point in the y-direction. Using Hall's terminology, this force

[26]Recall that the effective mass is related to the curvature of the energy band via $(m^*)^{-1} \equiv \hbar^{-2}\,\partial^2 E(k)/\partial k^2$.
[27]The original paper from Hall is an educating

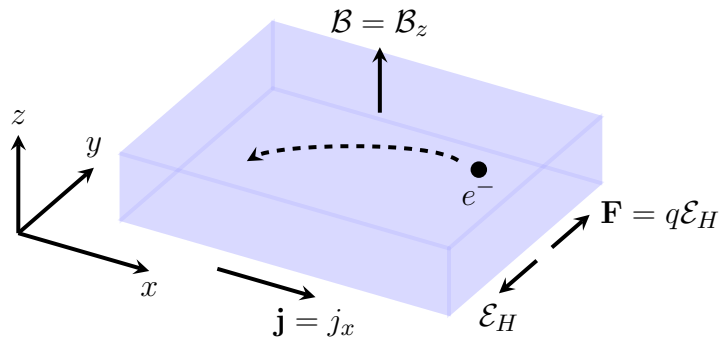

Fig. 6.6: Geometry for the Hall effect.

can be understood to **press** charges (but not move them) along y-direction.[28] Recalling that the electric field is defined through the above Lorentz force equation $\mathbf{F} = q\mathcal{E}$, this force could be equally well treated as an electric field which we will denote here as $\mathcal{E}_H$, also known as the Hall field.

We can now relate the Hall field $\mathcal{E}_H$ with the applied magnetic field $\mathcal{B}_z$ and the current j_x by requiring the forces to be equal:

$$q\mathcal{E}_H = j_x \mathcal{B}_z \,, \tag{6.4.27}$$

which allows to solve for the total charge q. The inverse of the total charge is known as the Hall coefficient R_H, and is found to be

$$R_H \equiv \frac{\mathcal{E}_H}{j_x \mathcal{B}_z} = \frac{1}{q} \,, \tag{6.4.28}$$

which for the case of negatively charged electrons turns out to be

$$R_H = \frac{1}{q} = -\frac{1}{ne} \,, \tag{6.4.29}$$

where n is the conduction electron concentration. For the case of holes giving rise to the current in the conductor, the Hall coefficient becomes

$$R_H = \frac{1}{q} = \frac{1}{pe} \,, \tag{6.4.30}$$

where p is the concentration of the holes. The last two equations are the main results of this section. From the historical point of view, the Hall effect and subsequent measurements provided the first experimental evidence that an electric current in metallic

[28] The author is aware that unfortunately there is some confusion in the literature regarding this point. As pointed out already by E. H. Hall himself in his ***seminal 1879 treatise***, and by J. C. Maxwell before him, the applied $\mathcal{B}$-field **does not move charges to the other side**. Rather, it gives rise to a force that presses the charges, which can be equally well interpreted to be an electric field, termed as the Hall field $\mathcal{E}_H$.

solids occurs mostly due to negatively-charged carriers, i.e. electrons (and not protons). This finding helped to form a proper understanding of the situation and has played an important role in the development of the field of solid-state physics.

We conclude by outlining the general situation of the current being cause both by electrons and holes. The derivation of the result is deemed outside the scope of our treatise, but results in a Hall coefficient given by

$$R_{\mathrm{H}} = \frac{p\mu_{\mathrm{h}}^2 - n\mu_{\mathrm{e}}^2}{e\left(p\mu_{\mathrm{h}} + n\mu_{\mathrm{e}}\right)^2}\,, \tag{6.4.31}$$

where μ_{h} and μ_{e} are the hole and electron mobilities, respectively. Looking at the above equation we see that measurement of the Hall coefficient R_H provides experimental information of the carrier concentrations, therefore acting as an important technique to characterize the properties of semiconductors.

7. DIELECTRIC MATERIALS

Dielectrics are a class of materials that has been so-far largely untouched during our course. In this Chapter, we will introduce the main features and properties of dielectric materials, paying special attention to the overall approach used in this course. Particularly, we aim to improve our understanding of the microscopic origins of the properties, and will also try to connect this picture with the concepts we have already learned during our previous courses.

First, we will briefly (re-)introduce the Maxwell's equations and the associated fields that can be used to understand most of the electronic and optical properties of dielectric materials. Emphasis will be placed on elucidating the connections between the microscopic description and the macroscopic description, which can be done by introducing so-called local-field effects. Second, we will revisit the well-known Lorentz model for an atom that can be used to understand the optical properties of dielectric materials. During this revisit, we can now firmly link the band structure theory with the Lorentz model (and associated approaches), providing us new tools to understand the occurring light–matter interaction.

At the end of the Chapter, we will discuss some common effects taking place in dielectrics, and how defects and/or impurities in dielectrics can be utilized in materials science.

7.1 Macroscopic Description

For completeness, we start by writing out the already familiar macroscopic Maxwell's equations. In order to keep the notation as clean as possible, we use here normal bold font for the vector field quantities.[1]

Here we start with the equation for the ***Lorentz force***, known to be

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}\,, \tag{7.1.1}$$

where $\mathbf{F}$ is the force felt by a particle with charge q, $\mathbf{E}$ is the electric field, $\mathbf{B}$ magnetic field and $\mathbf{v}$ the velocity of the particle. It is this equation which can be used to define the field quantities $\mathbf{E}$ and $\mathbf{B}$ as the sources of the movement of the particle.

Now that the fields have been properly defined, we can write the Maxwell's equations

[1] When we were discussing the energy states and associated energies of our systems, we often labelled the energies with E. As a consequence, notation $\mathcal{E}$ was used for the electric field in order to avoid confusion. Here our emphasis is on the field quantities and not on the energy states of the system. Subsequently, we resort to using the conventional notation for field.

in a macroscopic medium as

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho_{\mathrm{f}}, & \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, & \nabla \times \mathbf{H} &= \mathbf{J}_{\mathrm{f}} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned} \tag{7.1.2}$$

where ρ_{f} is the free electric charge density and $\mathbf{J}_{\mathrm{f}}$ is the free electric current density.[2] In the above equations, we have used the so-called ***auxiliary fields*** $\mathbf{D}$ and $\mathbf{H}$ defined through equations

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \qquad \mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}\,, \tag{7.1.3}$$

where $\mathbf{D}$ is the electric displacement, $\mathbf{P}$ material polarization, $\mathbf{H}$ the magnetizing field and $\mathbf{M}$ the magnetization. Final set of equations connect the macroscopic bound charges ρ_{b} and currents $\mathbf{J}_{\mathrm{b}}$ to the material polarization $\mathbf{P}$ and magnetization $\mathbf{M}$ via definitions

$$\rho_{\mathrm{b}} = -\nabla \cdot \mathbf{P}, \qquad \mathbf{J}_{\mathrm{b}} = \nabla \times \mathbf{M} + \frac{\partial \mathbf{P}}{\partial t}\,. \tag{7.1.4}$$

The above equations describe the light–matter interactions on the macroscopic scale.

The problem with the above equations is that they do not suit our current needs. Our motivation is to understand the macroscopic properties of solids from their microscopic properties, and the above equations do not tell us how the microscopic constituents of matter (mostly just the electrons) respond to applied fields. Therefore, we need to resort to the so-called microscopic Maxwell's equations in order to link the applied fields to the microscopic properties of solids.

7.2 Microscopic Polarization

We continue our treatment by linking the macroscopic polarization $\mathbf{P}$ to the microscopic dipole moment $\mathbf{p}$ via definition

$$\mathbf{P} \equiv \mathbf{p}/V = N\mathbf{p}\,, \tag{7.2.5}$$

where V is the unit volume of a small piece of solid and $N = 1/V$ the number density associated with the unit volume.

[2]The sources of the equations, ρ_{f} and $\mathbf{J}_{\mathrm{f}}$, are connected to each other. This is most easily seen by moving our treatment from the (space + time) coordinate system into the ***Lorentz-invariant*** relativistic ***spacetime*** system. Then we can simply note that using the Minkowski metric the quantities constitute a four-current given by $J^\alpha = \left(c\rho, j^1, j^2, j^3\right) = (c\rho, \mathbf{J})$. In other words, the electric charge ρ and the current $\mathbf{J}$ are really just different components of the same quantity! Enlightenment awaits us on this winding road, because similar connections can be made between the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$.

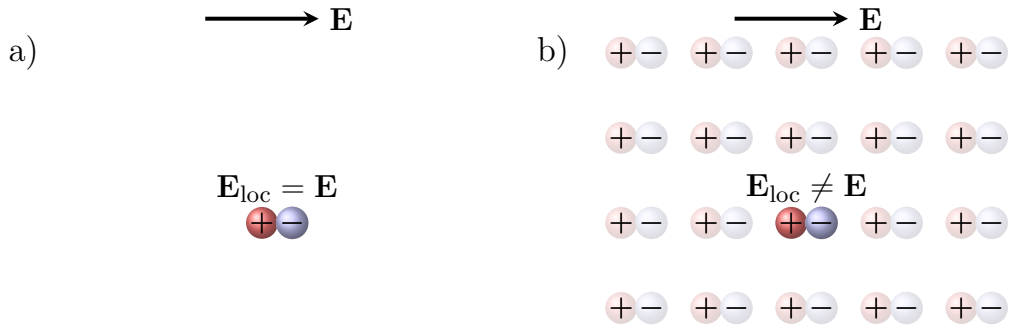

Fig. 7.1: a) For the case of a single atom/molecule and the associated dipole moment, the local field $\mathbf{E}_{\rm loc}$ equals the incident field $\mathbf{E}$. b) In a solid consisting of many atoms/molecules the local field may considerably differ from the incident/applied electric field ($\mathbf{E}_{\rm loc} \neq \mathbf{E}$.

The dipole moment $\mathbf{p}$ can be linked to the electric field by introducing a new quantity α, known as the polarizability of the microscopic piece of solid. The polarizability α is defined via

$$\mathbf{p} = \alpha\epsilon_0\mathbf{E}_{\rm loc}\,, \tag{7.2.6}$$

where the electric field $\mathbf{E}_{\rm loc}$ corresponds to the local field in the vicinity of the polarizability. It is important to note that the local field often considerably differs from applied/incident electric field ($\mathbf{E}_{\rm loc} \neq \mathbf{E}$) as has been schematically shown in Fig. 7.1. Therefore, the problem is to find a relation between the macroscopic field $\mathbf{E}$, associated with the macroscopic polarization through[3]

$$\mathbf{P} = \epsilon_0\chi\mathbf{E}\,, \tag{7.2.7}$$

where χ is the material susceptibility, and the local field $\mathbf{E}_{\rm loc}$. The microscopic field $\mathbf{E}_{\rm loc}$ is associated with the dipole moment $\mathbf{p}$ via the above Eq. (7.2.6). A solution to this problem will be shown in the next section.

7.2.1 Clausius–Mossotti Relation

Without going into the details of the lengthy and non-trivial derivation, the local field for an isotropic dielectric medium is found to be[4]

$$\mathbf{E}_{\rm loc} = \mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P}\,. \tag{7.2.8}$$

[3]Note that we have implicitly assumed frequency-independent material responses here. This choice is purely due to conceptual clarity. Generalization to frequency-dependent responses can be made by moving the time-dependent quantities from time domain description into the frequency domain description using standard ***Fourier analysis***.

[4]The derivation for the local field can be found for example from J. D. Jackson's *Classical Electrodynamics*: Section 4.

Inserting above into the defining Eq. (7.2.5) for the dipole moment, and recalling Eq. (7.2.6) we get

$$\mathbf{P} = N\mathbf{p} = N\epsilon_0\alpha\left(\mathbf{E} + \frac{1}{3\epsilon_0}\mathbf{P}\right). \tag{7.2.9}$$

We can now solve the above equation for polarization density resulting in

$$\mathbf{P} = \epsilon_0\frac{N\alpha}{1 - N\alpha/3}\mathbf{E} = \epsilon_0\chi\mathbf{E}\,, \tag{7.2.10}$$

where we have used Eq. (7.2.7) for the latter equality. Looking at the above, we see that the macroscopic material susceptibility χ can now be linked to the microscopic polarizabilities α of the dielectric via relation

$$\chi = \frac{N\alpha}{1 - \frac{1}{3}N\alpha}\,. \tag{7.2.11}$$

Because for most of materials (and for most wavelengths) the polarizabilities α are positive quantities, we see by looking at the above equation that the local-field corrected suceptibility χ is larger than the simplistic $\chi = N\alpha$ prediction.

The above result can be described also in terms of relative permittivities ϵ_r, or squares of refractive indices n^2.[5] This follows by using $\epsilon_r = 1 + \chi$ in order to write the above Eq. (7.2.11) as

$$\begin{aligned} \epsilon_r - 1 &= \frac{N\alpha}{1 - \frac{1}{3}N\alpha} \\ \implies \qquad (\epsilon_r - 1)\left(1 - \frac{1}{3}N\alpha\right) &= N\alpha \\ \implies \qquad \epsilon_r - \epsilon_r N\alpha/3 - 1 + N\alpha/3 &= N\alpha\,. \end{aligned} \tag{7.2.12}$$

We can now move the terms depending on the polarizability α to the right-hand side of the equation, giving us eventually[6]

$$\frac{\varepsilon_{\rm r} - 1}{\varepsilon_{\rm r} + 2} = \frac{N\alpha}{3}\,, \tag{7.2.13}$$

[5] Recall that at optical wavelengths, also in many other occasions, the refractive index of a medium is given by $n = \frac{c}{v} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \approx \sqrt{\frac{\epsilon}{\epsilon_0}} = \sqrt{\epsilon_r}$.

[6] Note here that we are using a notation where the polarizability α is given in the CGS-units (m^3), because of our decision of using the defining equation $\mathbf{p} = \alpha\epsilon_0\mathbf{E}_{\rm loc}$ instead of $\mathbf{p} = \alpha\mathbf{E}_{\rm loc}$.

which is known as the ***Clausius–Mossotti relation***. Replacing the relative permittivities ϵ_r with the squares of refractive indices we get

$$\frac{n^2 - 1}{n^2 + 2} = \frac{N\alpha}{3}\,, \tag{7.2.14}$$

which is known as the ***Lorentz-Lorenz equation***.

We end our treatment of local-field effects by briefly mentioning that such effects have recently resurfaced in the photonics research community. This has been attributed to the recent advances in nanofabrication and development of metamaterials, that have allowed researchers unforeseen possibilities to engineer the local fields present in the fabricated nanomaterials. It is particularly interesting to note that the above derivation of the local field Eq. (7.2.8) no longer necessarily holds for such materials.

7.3 Frequency-dependent Permittivity

First we discuss briefly the microscopic mechanisms behind the permittivity, which provide insights into the overall frequency dependence of the material responses. After this brief introduction, we will discuss more thoroughly the optical wavelength range, often associated with electronic transitions. The treatment for the infrared wavelength range, associated with excitation of lattice vibrations, is left as a self-study exercise for interested readers.[7]

Depending on the frequency range of interest for the light–matter interaction, the polarizability α can be seen to arise from different processes. At optical (and shorter) wavelengths, i.e. at high frequencies, the polarizability is mostly due to the oscillatory movement of electrons. At lower frequencies the oscillations of atoms and/or molecules and their vibrations start to dominate, while at even lower frequencies, excitation of lattice vibrations of crystalline solids becomes the most important mechanism. This frequency-dependence of the polarizability has been illustrated in Fig. 7.2. Particularly, it is worth to notice that the values of polarizabilities α vary quite a bit depending on the wavelength range. Furthermore, near the presence of material resonances, the frequency dependence becomes richer in its features.

[7]See e.g. Section 9.4.1 in *Solid State Physics – An Introduction* by Philip Hofmann.

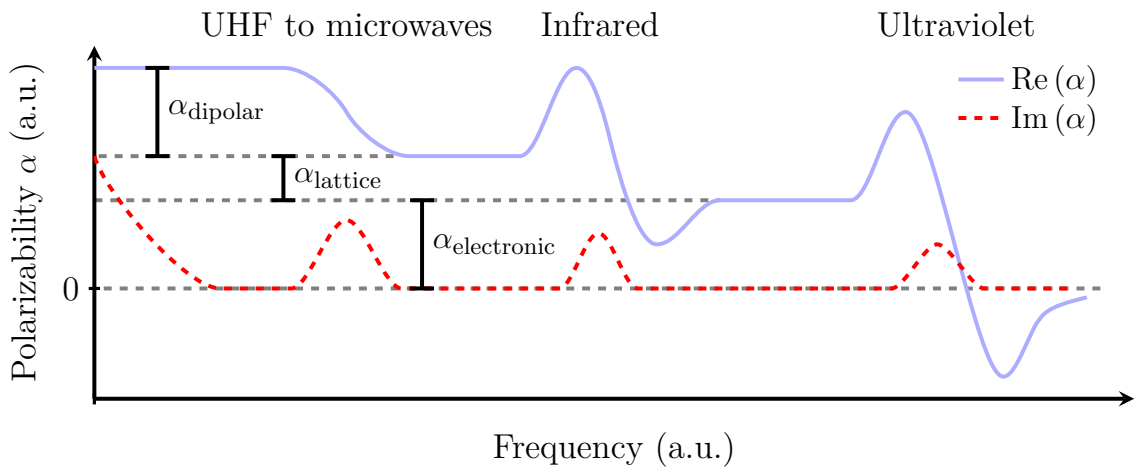

Fig. 7.2: Frequency dependence of the polarizability α over a wide range of frequencies. The changes in the real part of the α (blue curve) are associated with material resonances taking place in the ultraviolet, infrared and microwave frequencies. The imaginary part of α (dashed red curve) is linked to these resonances.

7.3.1 Harmonic Oscillator Model

We now briefly recall our first encounter with the Lorentz model for an atom and an electron, outlined in the course on Optics.[8] During those good old times we started our treatment by considering a damped harmonic oscillator described a following differential equation

$$m\ddot{x} + m\gamma\dot{x} + kx = qE_0 \mathrm{e}^{-\mathrm{i}\omega t}\,, \tag{7.3.15}$$

where m is the mass of the oscillating electron, γ is the damping term, k the spring constant and q charge of the electron. For simplicity, the applied field was taken to be time-harmonic wave of form $E(\omega) = E_0 \mathrm{e}^{-\mathrm{i}\omega t}$.

The solutions were found to be given by

$$x(\omega) = \frac{q/m}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega} E(\omega)\,, \tag{7.3.16}$$

which can be combined with the equation for polarization in order to estimate the relative permittivity of the system

$$P(\omega) = Nqx(\omega) = \frac{q^2 N/m}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega} E(\omega)\,. \tag{7.3.17}$$

Recalling Eq. (7.2.7) relating the fields P and E via the material susceptibility $\chi = \epsilon_r - 1$

[8] Please see Section 5.3 in the Optics -course ***Booklet***.

we see that the permittivity $\epsilon = \epsilon_0 \epsilon_r$ becomes

$$\epsilon(\omega) = \epsilon_0 + \frac{q^2 N}{m\left(\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega\right)} \,. \tag{7.3.18}$$

Equivalently, the above equation can be written in terms of the square of the refractive indices as

$$n^2(\omega) = 1 + \frac{Nq^2}{\epsilon_0 m} \frac{1}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega} \,. \tag{7.3.19}$$

The interesting points in the result are that we see the refractive index n of the solid to depend on the frequency ω. Furthermore, the refractive index n depends on the resonance frequency ω_0 associated with the harmonic oscillator.

Table 7.1: Relative permittivities of materials in the electrostatic approximation.[a]

Material	ϵ_r
Air[b]	1.0006
Glass	$5 - 10$
Diamond	5.7
Si	11.7
NaCl	6.1
Ethanol (liquid)	25.8
Water (liquid)	81.1
Si	11.7
GaAs	13.13
Ge	15.8
SiC	10.2

[a] Permittivity of material is $\epsilon = \epsilon_r \epsilon_0$.
[b] at 283 K and 101.3 kPa.

The above result is readily generalized to L different uncoupled harmonic oscillators, giving us a very interesting relation[9]

$$n^2(\omega) = 1 + \frac{Nq^2}{\epsilon_0 m} \sum_{j=1}^{L} \frac{M_j'}{\omega_{0,j}^2 - \omega^2 - \mathrm{i}\gamma_j\omega} \,, \tag{7.3.20}$$

where j labels the transitions and M_j' is known as the oscillator strength associated with the j^{th} transition. This equation is known as the ***Helmholtz–Ketteler formula***, and provides us an excellent starting point to improve our understanding of the microscopic origins of the frequency-dependent refractive indices of solids. In the next section, we will connect the above classical formula with the quantum-mechanically correct band structure theory in order to understand how solids absorb light.[10]

7.3.2 Electronic Transitions

We start our treatment by a few general remarks. As we saw in our earlier Chapter on semiconductors, photons associated with energy $E_{\text{phot}} = \hbar\omega$ can be absorbed by

[9] Please see Claus F. Klingshirn's great book *Semiconductor optics* for a generalization for N **coupled** oscillators. Unsurprisingly, such a situation (the refractive index n of a semiconducting material) exhibits also spatially dispersive behavior. In other words, the responses are associated with dispersion relations very similar to what we have seen during this course. Such is the nature of things.

[10] Because the imaginary and real parts of the permittivity ϵ are related to each other via ***Kramers–Kronig relations***, we see that the transitions taking place in the solid entirely dictate its refractive index profile.

the material if the photon energy exceeds the band gap energy E_g ($\hbar\omega > E_g$).[11] This approach is therefore perfectly valid for dielectrics and semiconductors. In addition, this overall idea is conceptually the correct approach in order to treat absorption of metals.

The main point to understand in the absorption process is that the band structure of the solid entirely dictates the allowed transitions and therefore the whole process of absorption. Because each absorption process needs to conserve the total energy and momentum of the system, absorption can take place only at very specific energy (and momentum) values. Next we seek to find a simple equation that would comply with the above idea. We know that absorption of light can be described by using a complex-valued permittivity ϵ, where the imaginary part of the permittivity $\mathrm{Im}(\epsilon) \equiv \epsilon_{\mathrm{im}}$ corresponds to absorption.[12] The conservation of energy during the absorption process then leads us to write the imaginary part of the permittivity to be proportional to a sum over all possible energy differences in the first Brillouin zone[13]

$$\epsilon_{\mathrm{im}}(\omega) \propto \sum_{\mathbf{k}} M^2 \delta\left(E_C(\mathbf{k}) - E_V(\mathbf{k}) - \hbar\omega\right) , \tag{7.3.21}$$

where M is an element describing the transition probability,[14], δ is the Dirac delta function while $E_C(\mathbf{k})$ and $E_V(\mathbf{k})$ are the dispersion relations of the conduction band and valence band, respectively. Looking at the above equation, we see that ϵ_{im} can become non-zero only when the condition $E_C(\mathbf{k}) - E_V(\mathbf{k}) = \hbar\omega$ is fulfilled.

In addition, we note that the conservation of the total momentum of the system restricts the number of allowed transitions. The momentum can be conserved only when the valence and conduction bands exhibit similar curvatures, because only in that case the effective masses of the electron states, and subsequently their momentum, can be matched.[15] This occurs because the momentum of the photon is negligible compared to the momenta of electron states.

[11] We're implicitly restricting our treatment to linear processes here. Nonlinear processes, such as ***Raman scattering*** or ***Brillouin scattering*** could also take place transferring part of the photon energy into the solid. We leave such processes and their description to our courses on nonlinear optics.

[12] While the real part of the permittivity $\mathrm{Re}(\epsilon) \equiv \epsilon_{\mathrm{re}}$ corresponds to scattering of light.

[13] This equation could be also derived using the ***Fermi's golden rule***.

[14] This matrix element M is where we hide the conservation of momentum, and other selection rules for the process.

[15] More precisely it is the total crystal momentum that is conserved. This means that the crystalline lattice can also absorb momentum by an amount of integer multiple of reciprocal lattice vector. Processes where this happens are called ***umklapp scattering*** processes, which are important e.g. when properly explaining the thermalization of solids.

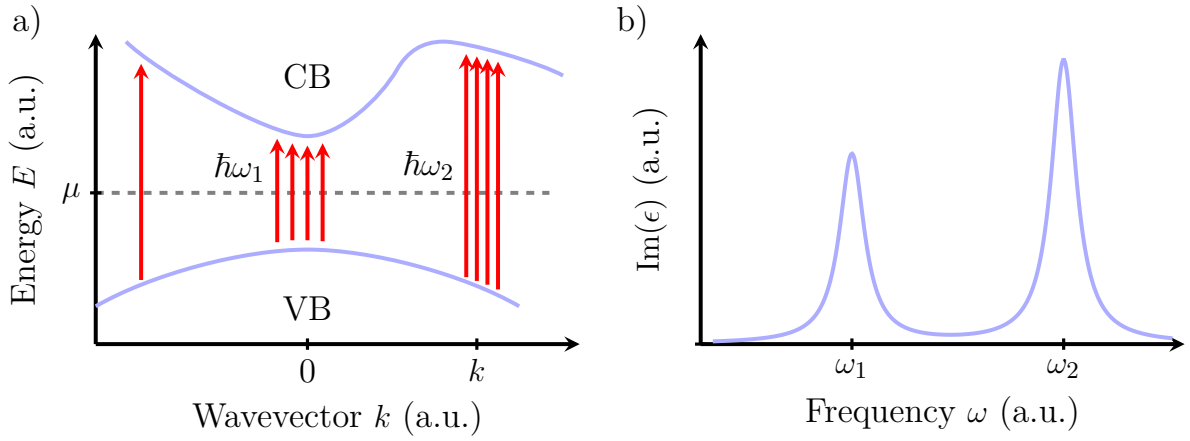

Fig. 7.3: a) Strong absorption takes place in the solid, when the energy and momentum are both conserved during a transition from an occupied electron state at the valence band (VB) up to an unoccupied electron state at the conduction band (CB). Transition probabilities are high near frequencies ω_1 and ω_2 because electron state momenta match near these points due to similar band curvatures. b) Corresponding frequency dependence of the imaginary part of the material permittivity $\mathrm{Im}(\epsilon)$.

This overall idea behind the absorption process and the role of the band structure in the process has been illustrated in Fig. 7.3a. Noticeable absorption can take place only at frequencies ω_1 and ω_2, for which both the energy and momentum are both simultaneously conserved Fig. 7.3b. The latter is conserved either near locations where the bands are parallel over a large range of $\mathbf{k}$ values, or when the dispersion relations flatten out.[16]

Finally, we note that the Eq. (7.3.21) provides us a clear understanding of why the Helmholtz–Ketteler formula works so well in describing refractive indices of solids. The electronic transitions in solids can only occur at very specific frequencies, and their contributions to the permittivities and refractive indices can be taken into account by appropriate summation over allowed transitions.

7.4 Impurities in Dielectrics

In this final section we aim to connect our discussion and understanding of dielectrics with our earlier notions on lattice impurities and defects (Chapter III). Although often the presence of impurities and defects degrades the behavior of the fabricated device, there are also many occasions where their presence is essential. Here we wish to provide

[16] We may write a few words of these flattened out dispersion relations and their role in absorption processes in our Chapter on finite materials. Some low-dimensional solids may exhibit so-called ***van Hove singularities***, that are related to flat dispersion bands.

a tangible example of an application of the latter case. We also wish to provide an example that you may have already heard of before, but which we can now understand more thoroughly using the concepts recently learned during this course.

7.4.1 Yttrium Aluminium Garnet (YAG)

A crystal known as the yttrium aluminium garnet (***YAG***, $Y_3Al_2\left(AlO_4\right)_3$) is a synthetic ***garnet*** crystal of generic form $A_3B_2\left(CO_4\right)_3$. YAG exhibits a fairly large band gap of 6.52 eV being therefore a dielectric material. What makes YAG a very relevant dielectric technologically, is that YAG crystal can be turned into a laser gain medium by doping it with impurities. Particularly important dopants are rare earth ions, which can replace yttrium ions without strongly modifying the lattice structure of the crystal. Historically by far the most important dopant has been trivalent neodymium (Nd^{3+}), turning the crystal into a neodymium-doped yttrium aluminum garnet (Nd:YAG) laser gain medium.

Because YAG crystal exhibits a relatively large band gap, no absorption of visible light can take place. However, the Nd^{3+}-ions used to dope the YAG crystal give rise to defect states and bands that in the doped Nd:YAG crystal can be excited using visible light (see Fig. 7.4). Particularly, a transition from the ground state ($^4I_{9/2}$) into the $^4F_{5/2}$-state is associated with 1.53 eV (808 nm) photon energies becomes possible. The excited Nd^{3+}-ions will then quickly relax non-radiatively (e.g. via phonon scattering processes) into a fluorescent $^4F_{3/2}$-state. The fluorescence lifetime of this state is quite long (230 μs). This makes it possible to populate the fluorescence state efficiently by pumping the system optically, and furthermore it becomes even possible to achieve ***population inversion*** in the state, which is a key ingredient when making a laser.

The doping concentration is often kept at moderate values ($\sim$1 at.%),[17] corresponding to number densities of the order of 10^{20} cm^{-3}. This doping concentration may sound as a high value when compared against normal doping concentration levels used in semiconductors. However, the doping concentration dictates how much energy the laser gain medium can store at any given instance of time, and it is this energy which defines the maximal pulses energies that the laser can provide. Therefore, we see that the Nd^{3+}-dopant concentration is an important parameter of a Nd:YAG laser crystal when

[17] Atomic percentage is defined as the fraction of yttrium (Y^{3+}) ions which have been replaced with Nd^{3+} ions.

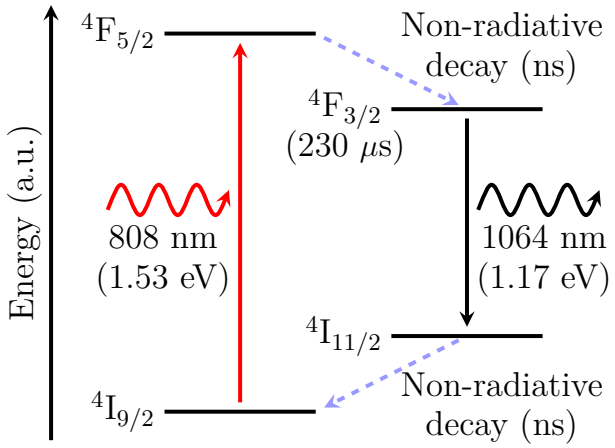

Fig. 7.4: Energy diagram for Nd^{3+}-ions inside a Nd:YAG crystal. Note the long fluorescence lifetime (230 μs) of the $^4F_{3/2}$-state and the fast non-radiative decay rates for the $^4F_{5/2} - {}^4F_{3/2}$ and $^4I_{11/2} - {}^4I_{9/2}$ transitions. This allows to achieve population inversion in the $^4F_{3/2}$-state by pumping the system at 808 nm.

one is making high-power pulsed lasers.[18]

We conclude this section by summarizing what we have just learned. We have seen that impurities in dielectrics are not only a nuisance that degrades the properties, such as the thermal conductivity, or electrical resistivity of an ideal crystalline dielectric materials. Quite contrary, many devices actually are based on impure dielectrics. Here, the laser gain media are a classical, and technologically also quite relevant, example of an application.

[18] We refer a novice reader towards our courses on laser physics for more details. More advanced readers may be interested to read more about ***thin-disk lasers***.

8. FINITE SOLIDS

All good things must come to an end, and so are we now almost at the end of our introductory course on solid-state physics, a branch of physics that is omnipresent in our lives. First we emphasize here, that this course is really just an introduction on the subject. Here we have really only briefly skimmed through the most very basic concepts and notions. But like good wine, also this material ages well, and the reader is warmly recommended to browse through the material a second time soon after finishing this first read. We all learn via repetition.

In the spirit of the above, this final Chapter focusing on finite solids, aims to also provide the reader a chance to re-encounter some of the basic concepts we have learned earlier. In particular, we will consider properties of **finite materials** in order to elucidate what changes in our treatment take place, when we relax our strict requirements regarding the periodicity of the problem.

First, we will introduce the key concepts that are important in finite materials. Second, we will go through different classes of finite materials. We will proceed by introducing one-dimensional (1D) materials, such as carbon nanotubes, that can be infinitely long in one dimensions, but finite in the other two spatial dimensions. After that we will discuss two-dimensional (2D) materials, such as graphene or molybdenum disulfide (MoS_2). Finally, we will consider properties of nanoparticles, which are finite along all their dimensions.

8.1 Quantum Confinement

We have seen already several times during our course, that all states of solids, such as the energy or momentum eigenstates, are ultimately discretized. The reason, why we have on many occasions treated our system as a quasi-continuum of states has arisen from the fact that a vast amount of states ($\sim 10^{22}$) exist already on a quite modest sounding material volume of 1 cm^3. However, if small enough solids are considered,[1] the assumption of quasi-continuum of state starts to fail leading to deviations in the properties of such finite materials. This effect is known as the ***quantum confinement***.

Quantum confinement can be clearly illustrated by considering a thin metallic film on top of a semiconducting material or an insulator (see Fig. 8.1a). In this case, it is smart to separate the single-electron wave function $\Psi(\mathbf{r})$ associated with the electrons of the

[1]Rule of thumb is that the material dimensions are of the order of the de Broglie wavelength of the electron wave function.

metal into different parts as given by

$$\Psi(\mathbf{r}) = \Psi(z)\Psi(x,y)\,, \tag{8.1.1}$$

where $\Psi(z)$ is the part of the wave function along the z-direction and $\Psi(x,y)$ is the part of the wave function along the x- and y-directions. The coordinate system for our problem is shown in Fig. 8.1a. For conceptual clarity, we assume the electrons to behave as free electrons along the xy-plane. In this case, the (kinetic) energy of the wave function turns out to be

$$E_{xy} = \frac{\hbar^2|\mathbf{k}_{xy}|^2}{2m_e}\,, \tag{8.1.2}$$

where $\mathbf{k}_{xy}$ is in-plane wavevector component pointing along the xy-plane and m_e is the electron mass.

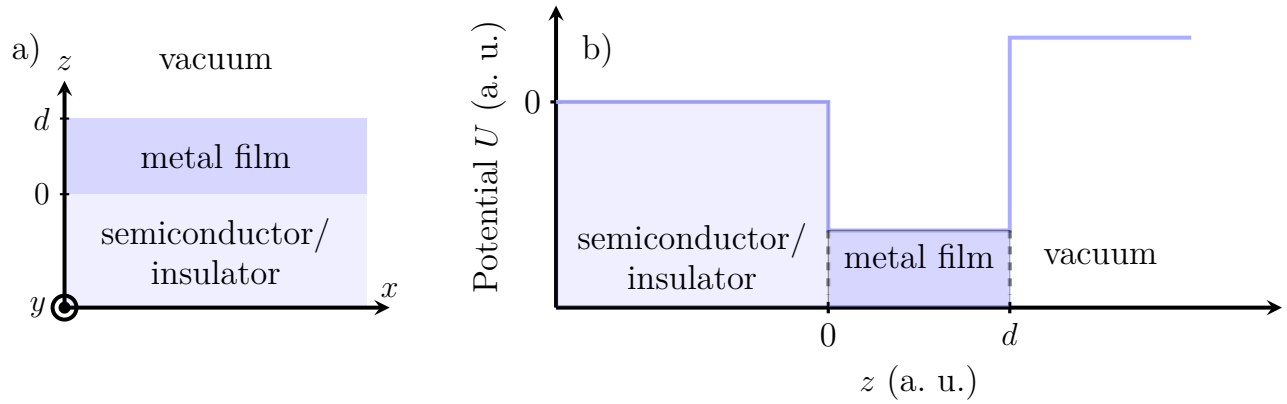

Fig. 8.1: a) Side-view schematic showing a thin metallic film place on top of a semiconducting/dielectric substrate. Note the coordinates we use for the situation. b) Potential U felt by the electron inside the metal film as a function of distance z. Note the resemblance of the situation of having a free particle in a (square) potential well.

The interesting direction is however the z-direction, where the system is confined in its dimensions. The problem in this direction is an exemplary case of having a free electron in an infinitely deep potential well. We know already that in this case the wave function turns out to be of form

$$\Psi(z) = A\mathrm{e}^{\mathrm{i}k_z z} + B\mathrm{e}^{-\mathrm{i}k_z z}\,, \tag{8.1.3}$$

where k_z is the wavevector component and A and B are constants determined by the boundary conditions.

The boundary conditions tell us, that the wave function $\Psi(z)$ should vanish at points $z = 0$ and $z = d$. This results directly in quantization of the wavevector component k_z

as given by[2]

$$k_z d = n\pi, \qquad n = 1, 2, 3, \ldots \tag{8.1.4}$$

which can be combined with $k_z = 2\pi/\lambda$ resulting in condition $2d = n\lambda$. This condition can be interpreted geometrically such that the electron wavelength λ must be an integer multiple of a round-trip length $2d$ inside the potential.

We have now all the terms defined in order to write out an equation for the kinetic energy of the electron associated with the movement of the electron along z:

$$E_z = \frac{\hbar^2 k_z^2}{2m_e}\,. \tag{8.1.5}$$

The total energy of the electron can now be calculated by summing the above two energies E_{xy} and E_z, and because k_z is clearly discretized, so will be the total energy of the electron. Therefore we see that the energy eigenvalues for the electron behave very differently from the case where the electron is in a infinite solid.

8.1.1 Quantum Dots

Next, we can generalize our treatment to semiconductors, and their electronic properties. If the semiconductor in question is of the order of ≈ 100 nm in all its dimensions, we call it a ***quantum dot***. Using arguments similar to the above case, we know that the energy levels of the quantum dot are strongly discretized. Furthermore, the band gap energy E_g of the quantum dot is larger than the one associated with a bulk semiconductor. An approximate energy difference between the electron and a hole state is given by

$$E_{\text{tot}} = E_g + \frac{\hbar^2\pi^2}{2\mu r^2} - \frac{1.8e^2}{4\pi\epsilon_0\epsilon r}\,, \tag{8.1.6}$$

where r is the radius of the quantum dot and μ is the reduced mass of the electron–hole pair. In the calculation of the reduced mass

$$\mu = \frac{1}{\frac{1}{m_e^*} + \frac{1}{m_h^*}} = \frac{m_e^* m_h^*}{m_e^* + m_h^*}\,, \tag{8.1.7}$$

the effective masses of the electrons m_e^* and holes m_h^* should again be used. This is because despite the small size of the quantum dot, the semiconductor material does still

[2]A finite well would give us a very similar quantization condition: $2k_z d + \Phi_i + \Phi_v = 2\pi n, \quad n = 1, 2, 3, \ldots$ where Φ_i and Φ_v would be the phase changes associated with reflections at the metal–semiconductor and metal–vacuum interfaces, respectively.

have its internal periodic lattice structure, which gives rise to electronic bands and thus the effective masses.

It is particularly the second term in the Eq. (8.1.6) which looks interesting. This energy term increases when the radius r decreases and is therefore associated with the quantum confinement. The third term is a standard Coulombic interaction term, which is modified by the permittivity ϵ of the semiconductor. Looking at the above equation, we see that absorption, and subsequent fluorescence, spectra of quantum dots can be modified simply by controlling the size of the quantum dot.

8.2 1D Materials (Carbon Nanotubes)

We focus next on 1D materials, which are solids that are finite in two dimensions but infinite in the third one.[3] Perhaps the most important examples of 1D materials are ***carbon nanotubes*** (CNTs), which are one ***allotrope of carbon***. What makes CNTs interesting to us is that they have found a wide variety of applications. Because of their mechanical rigidity, CNTs can be used to reinforce materials, such as (high-end) ice hockey sticks. In addition, depending on the fine structure of the CNTs, they can be either as metallic or semiconducting, potentially allowing their use in miniaturization of modern electronic components. And finally, because of their small size and quantum confinement, CNTs exhibit only a small number of electronic states. As a consequence, absorption, arising from transitions between these states, can saturate relatively quickly, making them an interesting material to use in opto-electronic applications.[4]

Next, we will look more closely into the electronic properties of CNTs deeming the mechanical and thermal properties to be beyond our current scope. Once the electronic band structure of CNTs is understood, we can use this information, for example, to understand how CNTs absorb light.[5]

8.2.1 Van Hove Singularities

Van Hove singularities can take place in the response of a solid, when the density of states (DOS) exhibits extraordinary behavior near the critical points of the first

[3] Strictly speaking we do not require the 1D materials to be really infinitely long along one dimension, just very long compared to the other two directions.

[4] ***Saturable absorption*** is a nonlinear optical process that can be utilized to make ***ultrashort pulse lasers***, particularly ***mode-locked lasers***.

[5] Or extending the treatment a bit into different semiconductor devices, we could also quickly start to understand how CNTs could be utilized to make smallscale ***field-effect transistors***.

Brillouin zone. Particularly, van Hove singularities can occur when the derivative of the DOS exhibits a discontinuity. Historically, such singularities were first studied by Léon van Hove in terms of the phononic DOS. For conceptual clarity, we will also use in our derivation the phononic DOS. However, the results can be directly generalized to the electronic DOS $g(E)$ of the solid. Next, we will apply our treatment to understand how van Hove singularities can affect absorption taking place in 1D materials, particularly in CNTs. In order to reinforce our trust on the approach introduced in Chapter IV, we start by recalling the definition for phononic DOS:

$$D(\omega, \mathbf{k}) \equiv \frac{\mathrm{d}N}{\mathrm{d}\omega}\,, \tag{8.2.8}$$

where N is the number of states and ω the frequency of the phonon. In our general case, the total number of states N should now be calculated by integrating over a volume that is no longer a simple sphere. Instead, the volume is a thin shell with the inside layer corresponding to frequency values of ω while the outside layer corresponds to frequency $\omega + \mathrm{d}\omega$ (see Fig. 8.2). In this case the total number of states is given by

$$N = \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} \mathrm{d}^3\mathbf{k}\,, \tag{8.2.9}$$

where L is again the length of the cubic volume of solid we consider. The calculation of this integral is our main problem here.

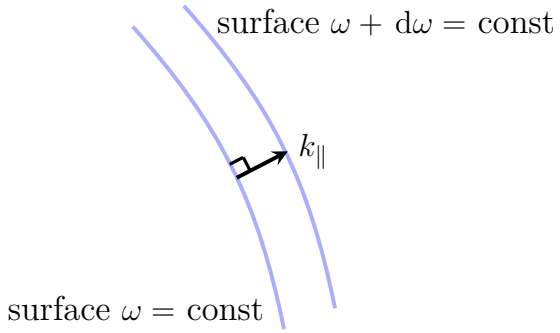

Fig. 8.2: The wavevector component $\mathrm{d}k_\parallel$ is the perpendicular distance between two constant frequency surfaces in the reciprocal space.

Calculation of the integral is made easier, if we separate the volume element $\mathrm{d}^3\mathbf{k}$ into components that are more symmetrical with respect to the volume of integration. Particularly, we will write the element in terms of an infinitesimal surface element $\mathrm{d}S_\omega$, which is parallel with the layers of the sheet, and an element that is perpendicular to this surface element. This element happens to be parallel with $\mathbf{k}$, due to which we label

this term as $k_{\parallel}$. The volume element is thus turned into

$$\mathrm{d}^3\mathbf{k} = \mathrm{d}S_\omega \, \mathrm{d}k_{\parallel} \, . \tag{8.2.10}$$

By noting that the group velocity v_g of our phonon state is generally written using a directional derivative along $\mathbf{k}$:

$$v_g \equiv |\nabla_{\mathbf{k}} \omega(\mathbf{k})| = \frac{\mathrm{d}\omega}{\mathrm{d}k_{\parallel}} \, , \tag{8.2.11}$$

we can calculate the derivative of Eq. (8.2.8). The result is

$$\begin{aligned} D(\omega, \mathbf{k}) \equiv \frac{\mathrm{d}N}{\mathrm{d}\omega} &= \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} \frac{\mathrm{d}^3\mathbf{k}}{\mathrm{d}\omega} \\ &= \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} \frac{\mathrm{d}S_\omega \, \mathrm{d}k_{\parallel}}{\mathrm{d}\omega} = \left(\frac{L}{2\pi}\right)^3 \int_{\text{shell}} \frac{\mathrm{d}S_\omega}{v_g} \, . \end{aligned} \tag{8.2.12}$$

The above equation is our main result for this section. It tells to us that the phononic DOS is inversely proportional to the amplitude of the group velocity v_g of the phonon state. This suggests that the DOS peaks at locations where the dispersion relation $\omega(\mathbf{k})$ is flat. Recalling Chapter IV, we notice that our solid may exhibit such dispersive behavior at the critical points of the first Brillouin zone. Such behavior has been illustrated in Fig. 8.3, where phononic DOS has been plotted for a simple Debye-like phononic dispersion, as well as the more general case exhibiting van Hove singularities.

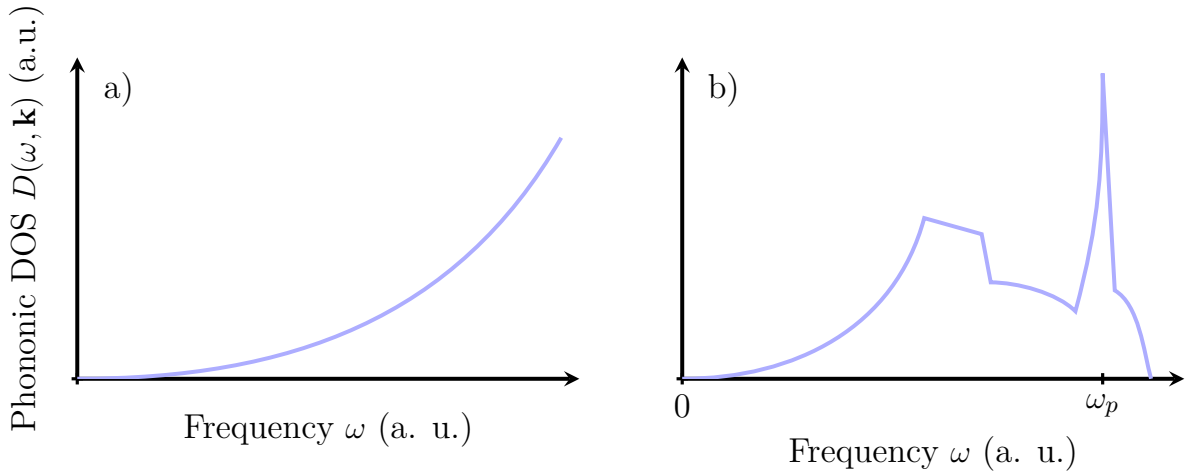

Fig. 8.3: The phononic DOS $\omega(k)$ for a) Debye-type phonons and b) for phonons residing in an actual crystalline structure. Discontinuities in the derivative of the DOS can take place especially near the critical points of the lattice, here marked with ω_p.

Lastly we point out, that the above derivation could be performed also for an electronic DOS $g(E)$, giving rise to similar van Hove singularities. Such singularities play a par-

ticularly important role in the electronic properties of CNTs, where they can strongly affect the optical absorption of carbon nanotubes. This process has been illustrated in Fig. 8.4, where absorption of a CNT is strongly enhanced due to presence of two van Hove singularities near the electron states corresponding to the transition.

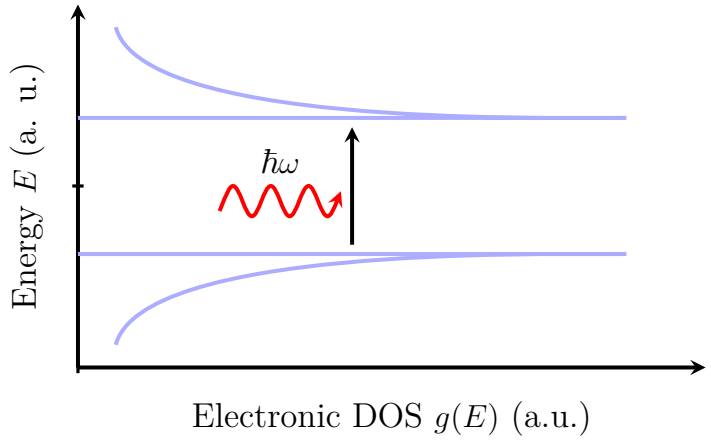

Fig. 8.4: The electronic DOS $g(E)$ of CNTs is strongly enhanced due to presence of van Hove singularities. The enhanced electronic DOS gives rise to strong absorption.

8.3 2D materials (Graphene)

Next we briefly describe how 2D materials can differ from bulk materials in terms of their properties. The topic is vastly beyond our current scope, but lat least we can provide the most basic concepts on the topic.

8.3.1 Dirac Cone of Graphene

Graphene was the first 2D material that was discovered in 2004 by playful fellows Andre Geim and Konstantin Novoselov.[6] The main reason why these discoverers of graphene were quickly awarded a Nobel price in physics is that graphene was the first real material that exhibited a so-called ***Dirac cone*** in its electronic band structure. It is precisely this fact which makes graphene (and other similar 2D materials) so special.

Let us first consider next in more detail the electronic band structure of graphene and the occurring Dirac cones. Leaving the detailed derivation for an interested reader, the dispersion relation $E(\mathbf{k})$ for our case of 2D hexagonal graphene lattice can be written

[6] A. Geim is also the recipient of the Ig nobel in physics in 2000 for levitating a frog in a magnet. He used a very power magnet, and wanted to demonstrate that every material responds to magnetic fields.

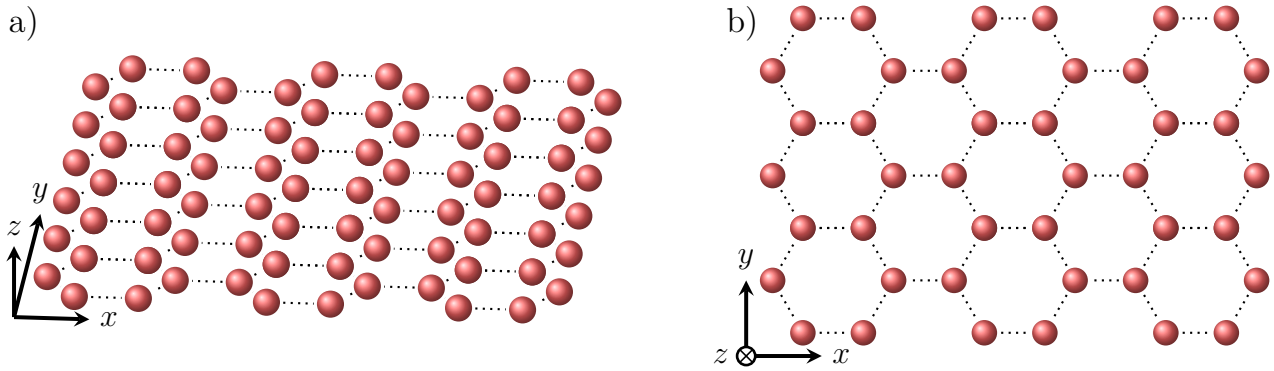

Fig. 8.5: a) Oblique and b) normal angle view of the honeycomb lattice of graphene. Note the hexagonal symmetry of the atomic lattice.

as

$$E\left(k_x, k_y\right) \propto \pm\sqrt{1+4\cos\left(\frac{\sqrt{3}ak_y}{2}\right)\cos\left(\frac{ak_x}{2}\right)+4\cos^2\left(\frac{ak_x}{2}\right)}\,, \tag{8.3.13}$$

where $a = \sqrt{3}a_0$ depends on the lattice constant a_0. The above dispersion relation $E(\mathbf{k})$ has been visualized in Fig. 8.6a and clearly exhibits the six-fold rotational symmetry as expected for a hexagonal lattice. Interestingly, we see that the upper (π) band and the lower (π^*) band seem to meet at the zero energy level (i.e. the Fermi level $E_F = 0$) at the critical points of the first Brillouin zone known as the **K** points, given by[7]

$$\mathbf{K} = \frac{4\pi}{3\sqrt{3}a_0}\mathbf{e}_x\,, \tag{8.3.14}$$

where $\mathbf{e}_x$ is the unit vector along x-axis. Because the bands seem to meet at this **K** point, it is called a ***Dirac point***. A zoomed up region of the dispersion relation near this Dirac point has been plotted in Fig. 8.6b.

We can now continue to discuss the importance of the Dirac cones that the dispersion relation of graphene exhibits near the **K** points of its first Brillouin zone. Recalling our earlier treatment of effective masses m^* for electrons, we see that Eq. (5.4.74) which earlier provided us our definition for the effective mass[8]

$$(m^*)^{-1} \equiv \frac{1}{\hbar^2}\frac{\partial^2 E(\mathbf{k})}{\partial k^2}\,. \tag{8.3.15}$$

[7]The other five **K** points in the first Brillouin zone are redundant can be found by rotating the given **K** point by an integer multiple of 60 degrees.

[8]In most general aproach, we should now generalize our effective mass to be compatible with a three-dimensional Brillouin zone. Consequently, the effective mass would become a 3×3 tensor m^*_{ij} making our notation clumsy. Instead, we will implicitly assume here that the movement of electrons is only looked along x- or y-directions.

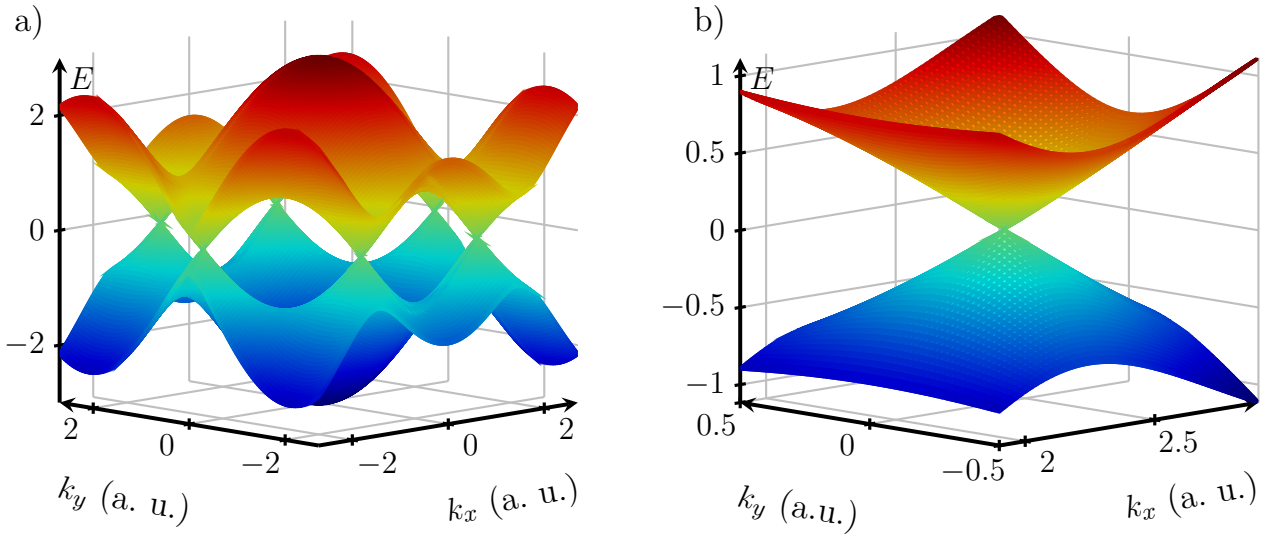

Fig. 8.6: a) Dispersion relation $E(\mathbf{k})$ for graphene ($a_0 = 1$). Note the six-fold rotational symmetry. b) Zoomed up region of $E(\mathbf{k})$ near the **K** point (Dirac point) of the first Brillouin zone showing the Dirac cone.

This definition would now result in diverging effective mass at the Dirac point! The above equation is therefore clearly not a satisfactory definition anymore.

Instead, by using an alternative definition for the effective mass m^* given by[9]

$$m^* \equiv \hbar^2 k \left(\frac{\partial E(\mathbf{k})}{\partial k} \right)^{-1} , \tag{8.3.16}$$

we can avoid the conundrum of diverging masses.[10] Expanding the dispersion relation of Eq. (8.3.13) near the **K** point we get a practically linear dispersion relation

$$E_{\pm}(\mathbf{k}) \approx \pm v_F \hbar |\mathbf{k}| + O\left(|\mathbf{k}|^2\right) , \tag{8.3.17}$$

where the higher-order terms $O\left(\mathbf{k}^2\right)$ are often neglected. This is a striking finding, because it substantially differs from the usual nonlinear $E(\mathbf{k}) = \hbar^2 |\mathbf{k}|^2 / 2m$ dependence. Instead, the above dispersion relation resembles dispersion relations usually associated with massless particles, such as photons.[11] This resemblance is also the reason why

[9] This is somewhat ad hoc way of defining the effective mass. Yet another possibility would be to define the effective mass as the experimentally measurable value for the effective mass, the so-called cyclotron mass.

[10] We would face problems when using the alternate definition for the rest mass for free electron states at $k = 0$ values, because both the energy and electron velocity both tend to zero. But otherwise, the definition complies with the physical requirements.

[11] Recall also that the acoustic, Debye-like, phonons exhibited similar linear dispersion relation.

electrons in graphene are sometime called as '*massless*'.[12] By inserting the above linear dispersion relation into the alternative definition for the rest mass we get

$$m^* = \frac{\hbar^2 k}{\pm v_F \hbar} = \pm \frac{\hbar k}{v_F}\,, \tag{8.3.18}$$

where the $(\pm)$-sign gives us the effective masses of electrons at the conduction band $(+)$ and electron holes at the valence band $(-)$.

Alternatively, we could recall the usual energy–momentum relation

$$p = \hbar k = mv = m^* v_F\,, \tag{8.3.19}$$

and replace the classical mass m and velocity v with their effective values to write the effective mass as

$$m^* = \frac{\hbar k}{v_F}\,. \tag{8.3.20}$$

Either way we calculate things, we see that the dispersion relation of graphene, and the associated effective masses of electrons differ considerably from those of ordinary materials. And because the effective mass m^* and the velocity of conduction electrons v_F affect the electrical conductivity of the solid, the conductivity of graphene is among the highest known to humankind.

8.4 Metallic nanoparticles

The main motivation for this section is to elucidate to the reader, that many other types of excitations, often associated with respective quasiparticles, may take place in solids. Plasma excitations, which are associated with collective oscillations of conduction electrons near surfaces of conducting solids, are perhaps the simplest of such excitations that have not been yet discussed during this course.[13] Next, we will provide a simple picture to understand how plasma oscillations may arise when small metallic nanoparticles are excited optically.

---

[12] Relatedly, the linear dispersion relation resembles those of massless fermions when we look at the ***relativistic energy–momentum relation*** $E^2 = (pc)^2 + \left(m_0 c^2\right)^2$. Because of this connection, electrons in graphene are also sometimes called '*relativistic*' electrons. However, their Fermi velocities v_F are never relativistic (i. e. being close to the speed of light). This nomenclature is used only because of the functional form of the equations.

[13] We refer an interested reader to the book *Introduction to Solid State Physics* by Charles Kittelfor a more detailed description.

8.4.1 Localized Surface Plasmon Resonances

In the electrostatic approximation, it is possible to find a very simple equation for the polarizability α for a spherical nanoparticle of radius a. This equation is given by[14]

$$\alpha(\omega) = 4\pi a^3 \frac{\epsilon(\omega) - \epsilon_m}{\epsilon(\omega) + 2\epsilon_m}\,, \tag{8.4.21}$$

where ϵ_m is the relative permittivity of the medium surrounding the nanoparticle, and $\epsilon(\omega)$ is the relative permittivity of the nanoparticle.[15] What makes the above equation very interesting, is that if a condition known as the **Fröhlich condition**[16]

$$\mathrm{Re}[\epsilon(\omega)] = -2\epsilon_m\,, \tag{8.4.22}$$

is satisfied, the complex-valued polarizability α exhibits a resonance. In other words, the nanoparticle starts to strongly absorb light oscillating at frequencies ω fulfilling the above condition.

We can now estimate the frequencies at which common metals could exhibit resonant behavior. We use the Drude model for the complex-valued permittivity

$$\epsilon(\omega) = \epsilon_1 + \mathrm{i}\epsilon_2 = 1 - \frac{\omega_\mathrm{p}^2}{\omega^2 + \mathrm{i}\gamma\omega}\,, \tag{8.4.23}$$

where ω_p is the plasma frequency of the metal and γ is a dampening factor arising from electron scattering processes. The real part ϵ_1 and the imaginary part ϵ_2 can be solved from the above equation to be described by

$$\begin{aligned} \epsilon_1(\omega) &= 1 - \frac{\omega_\mathrm{p}^2\tau^2}{1+\omega^2\tau^2} \\ \epsilon_2(\omega) &= \frac{\omega_\mathrm{p}^2\tau}{\omega\left(1+\omega^2\tau^2\right)}\,, \end{aligned} \tag{8.4.24}$$

where the relaxation time is defined as the inverse of the damping factor $\tau \equiv 1/\gamma$.

For conceptual clarity, we assume the nanoparticle to be surrounded by air ($\epsilon_m = 1$). Inserting the above Eq. (8.4.24) into Eq. (8.4.22) we see that the polarizability exhibits

[14] We note the identical functional form with the Clausius–Mossotti relation of Eq. (7.2.13)

[15] We apologize for possible confusion, but all permittivities in this section are relative permittivities. But in order to keep the notation clean, we have omitted the use of sub-indices in order to explicitly clarify the nature of the quantities.

[16] We assume the imaginary part of the permittivity $\epsilon(\omega)$ to be small.

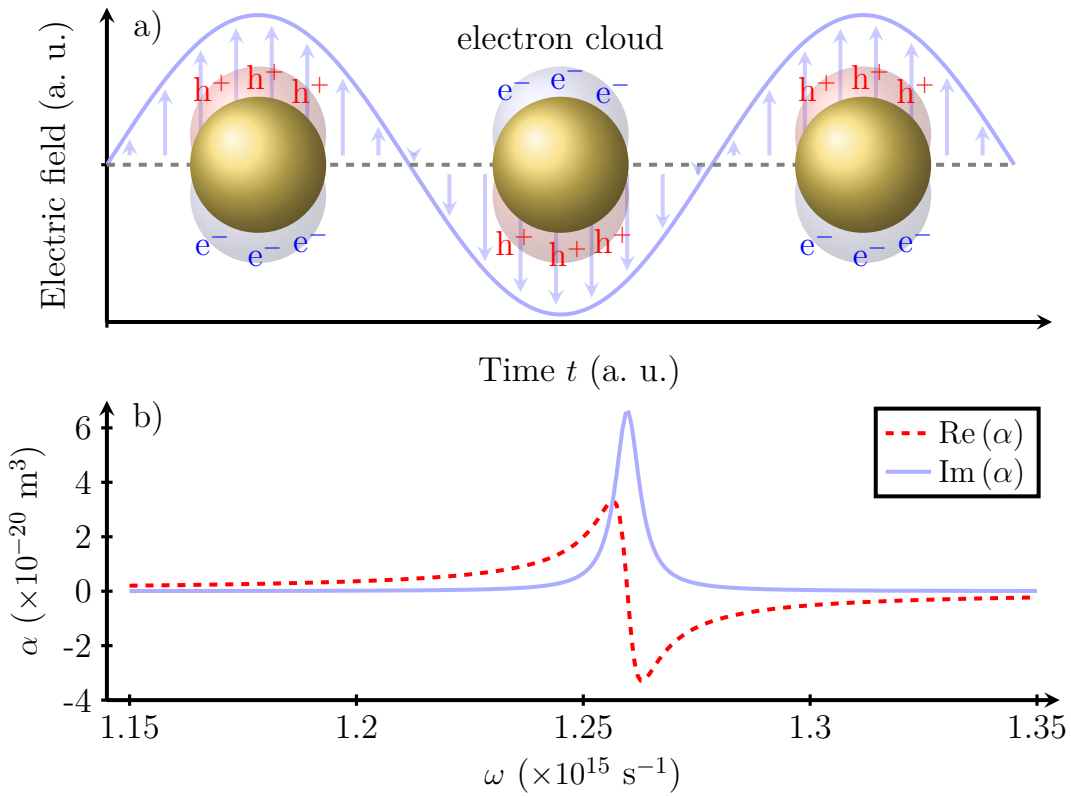

Fig. 8.7: a) Localized surface plasmon (LSP) is a collective oscillation of conduction electrons in a metallic nanoparticle. b) Polarizability α for a gold nanoparticle associated with a LSP resonance near frequency 1.26×10^{15} s^{-1} (238 nm). The imaginary part of the polarizability α (blue solid line) is associated with absorption of incident radiation, while the real part (red dashed line) is related to scattering of incident radiation.

a resonance near frequency

$$\omega_0 = \frac{\sqrt{\omega_p^2 - 3/\tau^2}}{\sqrt{3}} \approx \frac{\omega_\mathrm{p}}{\sqrt{3}}\,, \tag{8.4.25}$$

where ω_p was the plasma frequency of the metal.[17] For gold, the plasma frequency is $\omega_\mathrm{p} = 2.18 \times 10^{15}$ s^{-1}, giving rise to a resonance near the frequency $\omega_0 \approx 1.26 \times 10^{15}$ s^{-1} (238 nm). Because this resonance occurs due to both the plasma frequency of the metal and the finite size of the nanoparticle, it is called ***localized surface plasmon*** (Fig. 8.7a). The polarizability α for a gold nanoparticle has been plotted in Fig. 8.7b verifying the location of the resonance.

[17]The bulk plasma frequency ω_p is also associated with increased absorption of electromagnetic radiation. In this case, the absorbed energy is associated with conduction electron oscillations in the solid. These kinds of electron gas oscillations, that essentially take place at the surface of the metallic solid, are called surface ***plasmons***, which can be also interpreted as quasiparticles.

It is worth to note here that the above derivation is somewhat simplistic. For example, our derivation did not consider possible interband transitions taking place in metals. In addition, the size of the nanoparticle is known to experimentally affect the location of the LSP resonance. However, more detailed theoretical treatment of the origins behind the size-dependent red-shift of LSP resonances is beyond our current scope. We content ourselves by concluding that the above treatment provides already reasonable accuracy for wavelengths longer than the LSP resonance wavelengths and for nanoparticles exhibiting dimensions of only a few tens of nanometers.

To conclude, the above treatment is meant to provide some appreciation for the reader of the vastness of things and concepts that we have not even mentioned during this course! Solid-state physics is an enormous branch of modern physics, and we have only been able to briefly skim through the very basics. Hopefully for some readers, this has been only the beginning of an interesting journey!

INDEX