

Physics 3
Quantum Mechanics and Solid State Physics for
Electric Engineers
Lecture Notes

A. Sólyom and P. Richter
Dept. Atomic Physics
Faculty of Natural Sciences
Budapest University of Technology and Economics - BME

February 3, 2014

Contents

Introduction	1
1 Quantum Mechanics	3
Quantum Mechanics	3
2 Experimental foundations	4
2.1 Black-body radiation.	5
2.2 Photoelectric effect	13
2.3 Compton effect	15
3 Stationary states	21
3.1 Stationary States	21
3.2 Wave-particle duality	25
3.3 Uncertainty relations	31
3.4 The wave function	35
3.5 The Schrödinger equation.	36
3.5.1 Free electron in 1 dimensional	40
3.5.2 One dimensional potential step	40
3.5.3 Potential box in 1 dimension	44
3.5.4 Potential box in 3 dimensions	50
3.5.5 Density of states	54
3.5.6 Linear harmonic oscillator.	56
3.5.7 One dimensional square potential well	58
3.6 Central potentials	60
3.7 The potential barrier, tunnel effect	61
4 Time dependent Schrödinger equation	66
4.1 Solutions of the time dependent Schrödinger equation	66
4.1.1 Free electron in 1D	69
4.1.2 Particle in a 1 dimensional potential box	69

4.2	Perturbation theory	70
4.3	Transition probabilities and selection rules	71
4.3.1	Selection rules	73
4.4	Radiative transitions	74
5	Formal quantum mechanics	76
5.1	Formal quantum mechanics. Operators	76
5.1.1	Operators	78
5.1.2	Operators in Quantum Mechanics. Angular momentum	82
5.2	Measurement in quantum mechanics	86
6	Central potential. The hydrogen atom.	91
6.1	Angular momentum.	91
6.2	The hydrogen atom.	95
6.3	Electron spin	104
6.3.1	Addition of angular momenta	107
6.4	Quantum mechanical analysis of the spectrum of the H atom. Spin-orbit coupling.	108
6.5	Spin-orbit coupling	111
6.6	The structure of atoms	113
6.7	He atom. Independent particle model. Pauli exclusion principle.	113
6.7.1	Independent particle model	114
6.7.2	Shielding potential	115
6.7.3	The Pauli exclusion principle	115
7	Electron structure of atoms.	120
7.1	The periodic table of elements.	120
7.2	Hund's rules.	122
7.3	Valence electrons	123
7.4	X-ray emission	124
8	Molecules	129
8.1	H_2^+ - The hydrogen molecule ion	129
8.2	Diatomic homonuclear molecules. Molecular orbitals. Chemical bond. . .	133
8.3	Heteronuclear molecules.	136
8.4	Polyatomic molecules.	137
8.5	Hydrocarbon molecules. Hybridization.	138
8.6	Rotation and vibration of molecules.	139
8.6.1	Rotation of diatomic molecules	139
8.6.2	Vibration of molecules	140
8.6.3	Vibration of polyatomic molecules	141

8.6.4	Franck-Condon principle	141
8.6.5	Scattering of light by molecules. Rayleigh and Raman scattering .	142
9	Statistical physics.	151
9.1	Statistical equilibrium.	151
9.2	Maxwell-Boltzmann distribution.	152
9.2.1	Application of the Maxwell-Boltzmann statistics to the ideal gas .	157
9.3	Quantum statistics.	159
9.4	Fermi-Dirac distribution.	161
9.5	Bose-Einstein distribution.	166
10	Interaction of light and matter.	170
10.1	Photon gas	170
10.2	Interaction of light and matter	172
10.3	Laser operation.	174
10.3.1	Optical amplification	174
10.3.2	Laser operation	175
10.3.3	Types of lasers	177
	Solid State Physics	179
11	Fundamentals	180
11.1	Categorization of Solids	180
11.2	Bonding in crystals	181
11.2.1	Covalent Crystals	182
11.2.2	Ionic Crystals	183
11.2.3	Hydrogen bond crystals	185
11.2.4	Molecular crystals	186
11.2.5	Metals	187
11.3	Crystal structures, unit cells and lattices.	187
11.4	Symmetries. Bravais lattices.	191
11.5	The Wigner-Seitz cell	196
11.6	Non-ideal crystals. Crystal defects	197
11.6.1	Point Defects	197
11.6.2	Line defects (<i>Dislocations</i>)	199
11.6.3	Planar defects	200
11.6.4	Bulk defects	201
11.6.5	Effect of defects on the properties of crystals	201

12 Determination of crystal structures by X-ray diffraction	203
12.1 Reciprocal lattice. Miller indices.	203
12.1.1 The reciprocal lattice.	203
12.1.2 Miller indices	206
12.2 Determination of crystal lattices by X- ray diffraction. Bragg and Laue formulas	211
12.2.1 Bragg diffraction formula	211
12.2.2 Laue equations	212
12.3 X-ray diffraction methods.	214
12.3.1 The Laue Method	214
12.3.2 The Rotating Crystal Method	214
12.3.3 The Debye-Scherrer Powder method	215
13 Theory of lattice vibrations	217
13.1 Monatomic linear chain, phonons	218
13.2 Diatomic linear chain. Optical and acoustical branches of the dispersion relation	221
13.3 Three dimensional lattices	224
13.4 Specific heat of lattice vibrations	226
13.5 Debye model	227
13.6 Specific heat of metals	229
14 Electrical properties	230
14.1 Conductors and insulators. Band theory of solids	230
14.2 The Drude model	234
14.3 The Sommerfeld model of metals	238
14.3.1 Specific heat of metals	242
14.3.2 Conductivity	244
14.4 Work function, thermionic emission and contact potential	246
14.4.1 Work function	246
14.4.2 Thermionic emission	246
14.4.3 Contact potential	248
15 Electrons in conductors	252
15.1 Quantum mechanics of electrons in periodic lattices. Adiabatic principle. Brillouin-zone. Bloch functions	252
15.1.1 The Adiabatic Principle	253
15.1.2 Hartree-Fock method	253
15.1.3 Bloch electrons	254
15.2 Crystal momentum of Bloch electrons. Dispersion relations	255
15.3 Kinematics of electrons. Effective mass	262

15.4	Width of the energy bands. Tight binding model.	268
15.5	Conduction of metals. Electrons and Holes	270
15.5.1	Effective mass of electrons and holes	273
16	Semiconductors	274
16.1	Homogeneous semiconductors	274
16.1.1	Intrinsic semiconductors	274
16.1.2	Extrinsic (doped) semiconductors	281
16.2	Semiconductor structures. The p-n junction. Applications	286
16.2.1	Inhomogeneous semiconductors. The (unbiased) p-n junction.	286
16.2.2	The biased p-n junction.	289
16.2.3	Transistors	291
16.3	Metal-semiconductor junctions	294
17	Superconductivity	307
17.1	Superconductivity	307
18	Optical properties	319
18.1	Optical properties. X-ray emission and absorption.	319
18.1.1	X-ray emission	319
18.1.2	X-ray absorption	320
18.2	Emission and absorption of visible light by solids. Luminescence and phosphorescence	323
18.2.1	Absorption of visible light	323
18.2.2	Luminescence and phosphorescence	325
19	Magnetism	327
19.1	Magnetic susceptibility	327
19.2	Types of magnetism	328
19.3	Magnetism of free atoms.	329
19.4	Diamagnetism	330
19.5	Pauli paramagnetism of metals	331
19.6	Paramagnetism of independent atomic moments	334
19.7	Ferromagnetism	336
19.8	Antiferromagnets	340
19.9	Ferrimagnetism	340
20	Dielectric properties of solids	343
20.1	Induced polarization	344
20.2	Orientation polarization	346
20.3	Solid Dielectrics	348

20.4 Application of the oscillator model	349
20.5 Non-linear effects	351
21 Appendices	353
Appendices	353
22 Quantum Mechanics	354
22.1 Spectrometers	354
22.2 The spread of a wave packet in time	356
22.3 Derivation of the Compton formula	359
22.4 Uncertainty relations for a wave packet	360
22.5 The linear harmonic oscillator - Analitical solution	363
22.6 The linear harmonic oscillator - Ladder operators	366
22.7 1 dimensional potential well	370
22.8 Derivation of Perturbation theory formulas	373
22.9 The operator of the angular momentum and its z component in spherical polar coordinates	376
22.10 Russel-Sounders (LS) and jj coupling of angular momenta. Effects on the electronic structure of atoms	377
22.11 Other type of hybridization: sp^2 and sp	379
22.12 Conjugated molecules	381
22.13 Calculating the maximum probability partition of the Maxwell-Boltzmann distribution	384
22.14 Superfluidity in helium 4.	388
23 Solid State Physics	390
23.1 The origin of van der Waals forces	390
23.2 Examps of Bravais lattices	390
23.3 X-ray diffraction methods Laue-, rotating crystal and Debye-Scherrer meth- ods.	392
23.3.1 The Laue Method	394
23.3.2 The Rotating Crystal Method	395
23.3.3 The Debye-Scherrer Powder method	395
23.4 Classical linear chain models of lattice vibrations	397
23.4.1 Single atomic linear chain	397
23.4.2 Diatomic linear chain.	400
23.4.3 3D Linear model of lattice vibrations	401
23.5 Mathematical note: From summation to integration	401
23.6 Derivation of the Bloch function	402

23.7 Kinetic energy of a Bloch electron	403
23.8 Tight-binding Bloch function	404
23.9 The explanation of the mass action law for semiconductors	408
23.10 Fabrication of Si based integrated circuits	409
23.11 Determination of $n_c(x)$ and $p_v(x)$ in a p-n structure	410
23.12 Temperature dependent resistivity of materials	411
23.13 The explanation of the color of gold	413
23.14 Derivation of the Larmor formula	414
23.15 Calculating the Pauli paramagnetic moment of metals	415
23.16 Derivation of the orientation polarization	416
23.17 Determination of the local electric field \mathbf{E}_{loc}	418
Index	420

Introduction

The *Master Course Physics 3 for Electrical Engineers* is an introductory lecture to the fundamental concepts of modern physics. Here we present the basis of the disciplines Quantum Mechanics and Solid State Physics, all in one semester. As both of these topics are very broad we had to restrict the material presented to those areas which have the greatest practical importance. The unconventional concepts of these disciplines provide the physical basis for up to date engineering. Therefore the method we are following does not require complicated and subtle mathematics. We rely on disciplines well known for electrical engineers: the differential and integral calculus. Although during the semester we introduce the basics of operator calculus, to understand that part only elementary algebra is required.

The material in this book is organized in three distinct parts: Quantum Mechanics, Solid State Physics and the Appendices.

The first part deals with (non-relativistic) Quantum Mechanics which is the base of all of modern quantum physics. The phenomena, unexplainable in the frame of classical physics (see Chapter 2), required a re-evaluation of our knowledge of the world. In the beginning of the 20th century this led to the development of quantum mechanics. In Chapter 3 we introduce the stationary *wave function* (or state function) of a microscopic particle (e.g. electron) and solve a handful of problems that help to understand the concepts. In Chapter 4 we discuss the problem of time dependent phenomena and introduce the time dependent Schrödinger equation. The most abstract chapter is Chapter 5 where the basics of the operator calculus and measurement theory is discussed. The next two chapters (Chapter 6 and Chapter 7) are devoted to the study of atoms with a central potential containing either a single or multiple electrons. This involves the quantum mechanics of the angular momentum, the hydrogen atom and elements in the periodic table. A discussion of formation of molecules follows (Chapter 8), which, leads to the understanding of chemical bonds. In Chapter 9 the basis of statistical physics are introduced and the distribution functions of classical and quantum statistical physics are compared. The final chapter in this part (Chapter 10) deals with the interaction between light and matter and the operation of lasers is also discussed.

The second part is about Solid State Physics. The first chapter of this part (Chapter 11) introduces the fundamental concepts. This mostly means crystal physics, although amorphous materials are also discussed briefly. Basic concepts like *crystal lattice* and *crystal symmetries*, primitive and other types of cells are introduced here. Chapter 12 presents experimental methods that are used for the determination of crystal structures. Constituent atoms and molecules in crystals are vibrating around their equilibrium positions. These vibrations are the theme of Chapter 13, in which both classical and quantum mechanical models are discussed. Chapter 14 introduces the concept of energy bands used in solid state physics. This is the chapter where questions about the electrical resistivity, the work function and contact potential of metals are discussed using semi-classical theories. This theme is examined from the viewpoint of quantum mechanics after introducing key concepts about movement of electrons in periodic structures in Chapter 15. While the previous chapters deal with electrons in conductors Chapter 16 is about homogeneous and inhomogeneous semiconductors and their applications. Superconductivity is also discussed. Chapter 17 is a short introduction to this topic with detailed examples for their practical application. Optical properties of solids are discussed next in Chapter 18. X-ray and visible light absorption and emission are the topics of this chapter. Chapter 19 discusses the magnetic properties of solids both from a phenomenological and microscopical point of view. Some of these can be explained using individual magnetic moments but quantum physics is required to explain for instance ferromagnetism. This part is closed with Chapter 20 which is about the dielectric properties of crystals.

Important 0.0.1. *To make it easier to recognize important statements, we mark them similar to this sentence¹.*

Example 0.1. *Problems with solutions are presented throughout the book marked similar to this.*

The Appendix gives the reader an opportunity to see the details of the theories presented and understand the formulas more deeply.

¹The PDF version of this document marks the important statements and the examples by putting them into colored boxes, however this feature is not available in the WEB version.

Chapter 1

Quantum Mechanics

Chapter 2

Experimental foundations

At the turn of the 20th century physics seemed to be a closed discipline¹. Everything seemed to fit perfectly. At the end of the 19th century all physical phenomena were described by one or more of the well known disciplines of Mechanics, Statistical Physics, Thermodynamics and Electrodynamics. This was the time when Maxwell's electromagnetic theory was considered the theory of the “ether”, the elastic solid medium whose mechanical waves are the electromagnetic waves including light. The time when physicists tried to trace back all problems to problems of classical mechanics. There were of course some marginal unsolved problems left but almost all physicists agreed that physics in the 20th century will be “the physics of the 6th decimal place” (Michelson 1903). But some of the best physicists saw that this was not the case.

A closer examination shows that in fact many unsolved mysteries remained in physics at that time. It is now clear some of these could not have been incorporated into a classical theory at all. The problems that were considered unsolved at the beginning of the 20th century among others included the following

the velocity of light in vacuum is invariant Why is it independent of the frame of reference used? What happens between the ether and the bodies that move through it?

periodic system of elements What are the principles behind the periodic table? Why are the chemical behavior of elements in the same column similar?

spectra of atoms and molecules Why do we have discrete spectral lines? What is the reason behind the simple rules that govern the spectrum? Why do the splitting

¹For instance the Munich physics professor Philipp von Jolly advised the young Max Planck against going into physics, saying, “in this field, almost everything has been already discovered, and all that remains is to fill a few holes.” Planck replied that he did not wish to discover new things, but only to understand the known fundamentals of the field, and so began his studies in 1874 at the University of Munich. In spite of this remark Planck's discovery of the *energy quantum* was the most important step towards Quantum Mechanics.

of the spectral lines in a magnetic field (the Zeeman effect) not follow the laws of the classical physics?

the problem of X-ray emission and absorption Why are X-rays emitted? How are they absorbed?

specific heat Why is the equipartition theorem true at high temperatures and why does it brake at low temperatures? Is it possible to have other statistical distribution functions than the Maxwell -Boltzmann function?

thermal radiation What formula describes the shape of the electromagnetic spectrum of an object at a given temperature? What is the physics behind it?

stability of the atoms According to electrodynamics an accelerating charge emits electromagnetic radiation and thereby loses energy. If we apply this principle to an electron in an atom then we find the electron should radiate all of its kinetic energy in about 10^{-8} sec, after that it should fall into the nucleus. But this is evidently not the case.

chemical bond Is chemistry based on physics? Can physics explain the chemical bond?

(external) photoelectric effect Why are the laws for an electron emission from a metal surface so complicated?

Compton effect Why does the frequency of the light change - when it is scattered by a free electron - the way it does?

radioactivity What causes the radioactive decay?

This list is not complete. Some of the problems (e.g. the invariance of the speed of light in vacuum) has lead to the development of the special then the general theory of relativity, others may only be explained using another new branch of physics: quantum mechanics. In this chapter we first discuss *some* of these phenomena that lead to the development of this new physics, some others will be addressed in later chapters.

2.1 Black-body radiation.

It was well known that if we heat an object to temperatures high enough it will emit visible light. The color of the light depends on the temperature of the material. The higher the temperature the bluer the color. We know that (visible) light is electromagnetic radiation with wavelengths in the 380 – 740 *nm* range. With suitable detectors we can verify that heated materials emit electromagnetic radiation not only in the visible part, but in every other part of the electromagnetic spectrum (infrared, ultraviolet).

This emission is characterized by the *emission coefficient* (also known as the black-body *irradiance* or *emissive power*) $\mathcal{E}(\nu, T)$, which gives the energy a body emits at a given frequency at a given temperature in unit time. It has the dimension J/sec . The emission can be more exactly described by its derivative according to surface area and solid angle, called *spectral radiance* $\varepsilon(\nu, T) = \frac{\partial \mathcal{E}(\nu, T)}{\partial \nu \partial T}$ which is the amount of energy emitted at a frequency ν per unit surface area per unit time per unit solid angle per unit frequency. It has the dimension $J \cdot m^{-2} \cdot sec^{-1} \cdot sr^{-1} \cdot Hz^{-1}$.

The color of a heated light emitting object is determined by the frequency dependence of $\varepsilon(\nu, T)$, (e.g. the frequency at which $\varepsilon(\nu, T)$ is maximum). As the temperature drops then the position of the maximum will shift toward the red then into the infra red range and the radio frequency range, but we can still detect electromagnetic radiation in the whole frequency range emitted by the material.

The frequency dependance of the intensity of electromagnetic waves on the wavelength frequency or energy over a specific portions of the electromagnetic spectrum is measured by *spectrometers*. See Appendix 22.1 for further information on spectrometers.

Materials not only emit but also absorb electromagnetic radiation. This absorption heats up the material. Therefore it is possible for an object to be in *thermal equilibrium* with electromagnetic radiation. The absorption of radiation can be characterized by the *absorption coefficient* a which tells us the ratio of the incoming radiation a body can absorb. The absorption coefficient is dimensionless, may depend on the frequency of the electromagnetic radiation and the temperature and must be between 0 and 1: $0 \leq a \leq 1$.²

Important 2.1.1. *Every material at all temperatures absorbs and above 0 K also emits electromagnetic radiation in the whole frequency range. Some of the thermal energy of a body is converted into this electromagnetic radiation therefore it is called thermal radiation.*

At the middle of the 19th century Kirchoff has found an interesting relationship between the absorption and emission of electromagnetic radiation (light) based on thermodynamics. Kirchoff deduced that the ratio $\frac{\varepsilon(\nu, T)}{a(\nu, T)}$ must be the same for every material:

$$\frac{\varepsilon^{(1)}(\nu, T)}{a^{(1)}(\nu, T)} = \frac{\varepsilon^{(2)}(\nu, T)}{a^{(2)}(\nu, T)} = \dots$$

otherwise a device could be constructed that transfers heat from a body of lower temperature to a body of higher temperature without any external energy input, hereby violating the 3rd law of thermodynamics.

²The part of the radiation that is not absorbed may be reflected back or may pass through the material. The amount of this part is $1 - a$.

If we now introduce a hypothetical object called a *black-body*³ which absorbs all electromagnetic radiation falling on it independently of frequency and temperature, i.e. for which $a \equiv 1$ then we need to deal with the frequency and temperature dependence of $\varepsilon(\nu, T)$ only:

$$\frac{\varepsilon^{(1)}(\nu, T)}{a^{(1)}(\nu, T)} = \frac{\varepsilon^{(2)}(\nu, T)}{a^{(2)}(\nu, T)} = \dots = \varepsilon(\nu, T) \quad (2.1.1)$$

Black-bodies do not exist in nature. But some materials with high absorption can be considered very close to this ideal black-body. Graphite, soot and lamp black⁴ have an absorption coefficient $a \geq 0.95$. NASA⁵ has developed a more modern ultra absorbent material coated with *carbon nanotubes* which absorbs 99.5 % of the incoming UV and visible light and 98 % of the longer wavelengths, with an average $a = 0.99$. Such coatings can be used for instance for stealth aeroplanes.

The classical model of a black-body is a cavity in a rigid opaque body with rough absorbing walls like the one in Fig 2.1. A small hole in the wall allows a small part of the radiation exit that allows measurement of the internal radiation. Such objects are fabricated by some companies and may be bought. They are certified in a given frequency range to behave like ideal black-bodies.

Because the absorption of a black-body is the largest absorption coefficient possible ($a = 1$), its emission coefficient is larger than for any other object. This means that at a given temperature the black-body emits the most intensive radiation. It follows that for instance if we compare the emission of various objects we find that at temperatures when the maximum of the emission is in the visible range then black-bodies are the brightest objects

It was found that

Important 2.1.2. for black-bodies in thermal equilibrium with the electromagnetic radiation *the product of the absolute temperature T and of the wavelength λ_{max} where the emission per unit wavelength has its maximum is constant (peak wavelength):*

$$\lambda_{max} \cdot T = 2.8977721(26) \cdot 10^{-3} \text{ K m} \quad (2.1.2)$$

This is Wien's displacement law.

³Sometimes written as *black body* or *blackbody*.

⁴Also known as *carbon black*, furnace black or thermal black is a form of amorphous carbon that has a high surface-area-to-volume ratio, although its surface-area-to-volume ratio is low compared to that of activated carbon. It is dissimilar to soot in its much higher surface-area-to-volume ratio and significantly lower (negligible and non-bioavailable) PAH (polycyclic aromatic hydrocarbon) content. Its most common use is a pigment and reinforcing phase in automobile tires.

⁵NASA is an acronym for the agency *National Aeronautics and Space Administration* of the United States government that is responsible for the nation's civilian space program and for aeronautics and aerospace research.

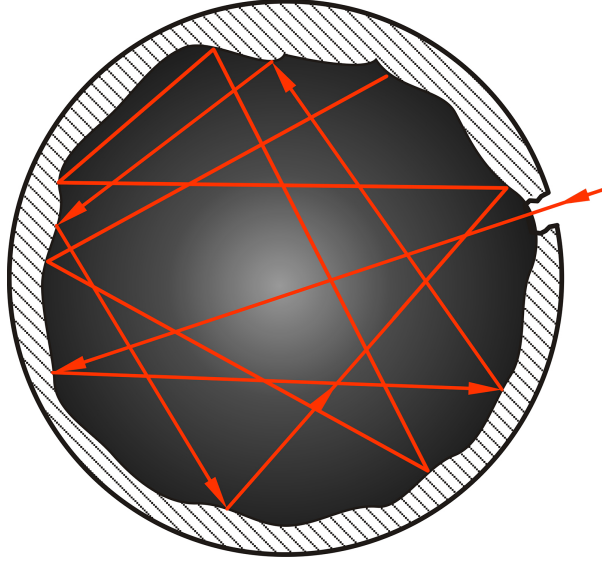


Figure 2.1: A model of a black-body is the hole of a cavity in a rigid opaque body. All incoming electromagnetic radiation is absorbed by the rough inner surface of the cavity in consecutive steps even when $a < 1$ for the material of the body as there is little chance for any reflected radiation to exit again through the hole.

Example 2.1. *The effective temperature of the Sun is 5778 K. What is the value of λ_{max} for the Sun?* **Solution**

$$\lambda_{max} = 2.90 \cdot 10^{-3} / 5778 = 5.02 \cdot 10^{-7} m = 502 \text{ nm}$$

This corresponds to the wavelength of green light near the peak sensitivity of the human eye.

Example 2.2. *According to theory, approximately a second after its formation the Universe was a near-ideal black-body in thermal equilibrium at a temperature above 10^{10} K. The temperature decreased as the Universe expanded and the matter and radiation in it cooled. The cosmic microwave Background radiation observed today is "the most perfect black-body ever measured in nature" as it has an anisotropy less than 1 part per 100,000. Now, some 15 billion years after the Big Bang the peak of the observed cosmic Background radiation is at 1.07 mm. What is the temperature of the cosmos?* **Solution**

$$T = 2.898 \cdot 10^{-3} / \lambda_{max} = 2.7 \text{ K}$$

The *Stefan-Boltzmann law* states that the total energy emitted ($\lambda \in [0, \infty]$) by a black-body per unit surface area is proportional to the 4th power of the absolute temperature:

$$\mathcal{P}_A = \sigma T^4 \quad (2.1.3)$$

where $\sigma = 5.670373(21) \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan–Boltzmann constant

Example 2.3. *A human body also radiates energy. Calculate the total energy needs for an adult to keep the body temperature constant. Because the mid- and far-infrared emissivity of skin and most clothing is near unity we may approximate the human body with a black-body. The average total skin area of an adult human being is about 2m^2 , and in an ambient temperature of 20°C the temperature of the bare skin is about 33°C , while under the clothing it is about 28°C .*

Solution From Wien’s law (equations (2.1.2)) the peak wavelength of the thermal radiation of a naked human body is about $9.5\text{ }\mu\text{m}$ ⁶. To calculate the energy needed to keep the temperature of the body constant can be obtained from the Stefan-Boltzmann law (2.1.3). The radiated power is the difference between the power absorbed from the environment (which is also considered a black-body) and the one emitted by the body:

$$\begin{aligned}\mathcal{P}_{\text{body}} &= \mathcal{P}_{\text{absorption}} - \mathcal{P}_{\text{emission}} = 4\pi \sigma (T_{\text{environ}}^4 - T_{\text{body}}^4) A \\ &= -95.10 \text{ W}\end{aligned}$$

The total energy requirement for a whole day therefore is

$$\mathcal{E} = -\mathcal{P}_{\text{body}} \cdot 24 \cdot 3600 = 8.216 \text{ MJ} = 1965 \text{ kcal}$$

Example 2.4. *Let us model the Earth with a perfect spherical black-body without an atmosphere! Determine the effective or average surface temperature if the solar constant I_o , i.e. the amount of incoming solar electromagnetic radiation per unit area – that is incident on a plane perpendicular to the rays, at a distance of one astronomical unit (AU) (roughly the mean distance from the Sun to the Earth) – was 1361 kW/m^2 ! **Solution** In the stationary state the “model ‘Earth’” absorbs the same amount of energy from the Sun as it emits. The Earth-Sun distance is so large that the rays of sunshine are almost parallel when they reach us. Half of the Earth surface is illuminated all the time by the Sun. The total energy absorbed by the Earth as a black-body, therefore equals to the solar constant multiplied by the cross section of the Earth perpendicular to the Earth-Sun direction⁷ and by the duration Δt*

$$\mathcal{E}_{\text{tot,absorbed}} = R^2 \pi \cdot I_o \cdot \Delta t$$

⁶Therefore thermal imaging devices are tuned to be most sensitive in the 7–14 micron range. But the human body emits at much larger wavelengths too. New imaging devices used in some border stations or airports use wavelengths in the 1 cm–1 mm (terrahertz) range. These are most suited to detect people smuggled in trucks.

⁷The sunlight I is perpendicular to the surface only at the point nearest to the Sun. Let us take a cross section of the sunlight with an area of A at this point. At a θ angle to the direction of the Sun this part of the sunlight hits a larger area $A' = A \cdot \cos\theta$, but only the component perpendicular to the surface is absorbed, which is $I' = I_o / \cos\theta$. The total absorbed radiation flux therefore $P = A \cdot \cos\theta \cdot I / \cos\theta = IA$ is the same at every point of the illuminated surface with a perpendicular surface area of A .

If the surface temperature is T then the total radiated energy from the Earth according to the Stefan-Boltzmann law is

$$\mathcal{E}_{tot,rad} = 4 \pi R^2 \sigma T^4 \cdot \Delta t$$

In a stationary state these two energies must be equal:

$$\mathcal{E}_{tot,absorbed} = \mathcal{E}_{tot,rad}$$

from which

$$T = \sqrt[4]{\frac{I_o}{4 \sigma}} = 278.3 \text{ K} = 5.3^\circ \text{C}$$

The real effective temperature of the Earth is higher, because of the atmosphere.

Example 2.5. *The albedo or reflection coefficient of the Earth is 0.3. This means that 30% of the solar radiation that hits the planet gets scattered back into space without absorption.*

- a) In the previous example what would be the temperature if the absorption coefficient of the Earth was $a = 0.7$ instead of 1?*
- b) In climate calculations it is sometimes assumed that regardless to reflection the Earth still emits like a black-body (this contradicts Kirchoff's law). What would the temperature be with this assumption?*

Solution

a)

If $a = 0.7$ then the absorbed energy is $I_a = I_o a$, a times as much as above, and according to Kirchoff's law the emission must be lower by the same factor, i.e. $E'_{tot,rad} = a E_{tot,rad}$, therefore the temperature is the same as was in the previous example, namely 5.3°C .

b)

In this case

$$\mathcal{E}_{tot,absorbed} = 0.3 R^2 \pi \cdot I_o \cdot \Delta t \mathcal{E}_{tot,rad} = 4 \pi R^2 \sigma T^4 \cdot \Delta t$$

and the temperature

$$T = \sqrt[4]{\frac{0.7 I_o}{4 \sigma}} = 254.58 \text{ K} = -18.58^\circ \text{C}$$

If a black-body can be fabricated (See Fig. 2.1), then $\varepsilon(\nu, T)$ can be measured. The resulting spectrum of such a measurement on materials, which at least in a limited frequency (wavelength) range, absorb almost all of the electromagnetic radiation look like

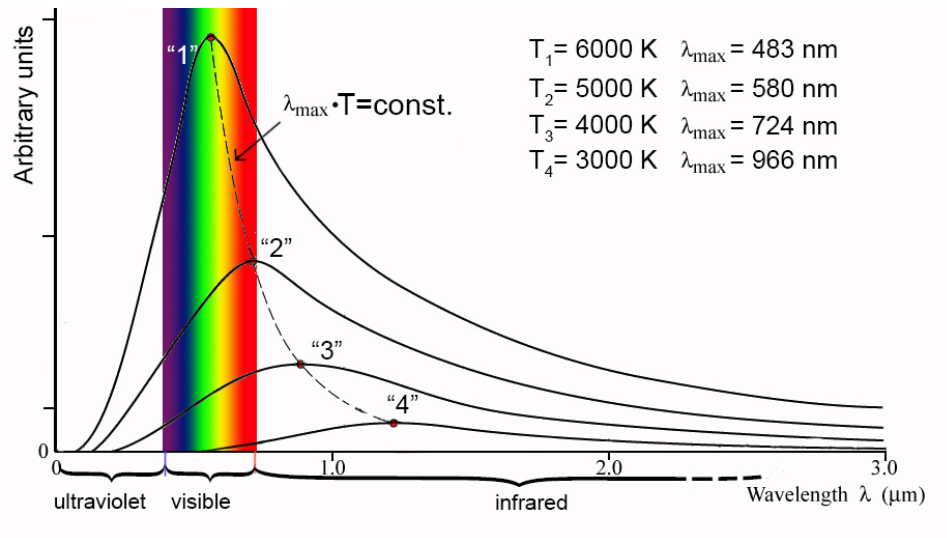


Figure 2.2: Calculated black-body emission curves and λ_{max} values.

the curves shown in Fig 2.2. Knowing the spectrum of a black-body in equilibrium with the electromagnetic radiation makes it possible to determine its temperature. Although real objects just approximate black-bodies the knowledge of the black-body spectrum and the spectra of the objects still gives a fairly good opportunity to remotely measure their temperatures even when the condition of thermal equilibrium does not hold. An example that the laws of black-body radiation can be applied with a good approximation is shown in Fig. 2.3 where the emission curve of our Sun, measured both above and below the atmosphere is shown⁸ together with the calculated spectrum of a black-body of temperature 5250 K. In Fig. 2.3 the calculated black-body spectrum reflects our present knowledge. Previous attempts could describe the radiation only at part of the spectrum. The two most famous are: the formula of Lord Rayleigh and Sir James Jeans, based on the equipartition theorem and *Wien's approximation*. The *Rayleigh-Jeans* law describes only the long wavelength (i.e. low frequency) part of the spectrum:

$$\varepsilon(\lambda, T) = \frac{2ckT}{\lambda^4} \quad (2.1.4a)$$

or with the frequency

$$\varepsilon(\nu, T) = \frac{2kT\nu^2}{c^2} \quad (2.1.4b)$$

⁸At sea level 3% of the radiation is ultraviolet, 44 % is in the visible range and 53 % is in the infrared. The gaps in the spectrum measured at sea level (red) are caused by *greenhouse gases* like water vapor and carbon dioxide. Water vapor has a larger absorption than CO_2 .

– and tends to infinity as the wavelength decreases (or the frequency increases). This is called the “*ultraviolet catastrophe*”. For short wavelengths (i.e. high frequencies) *Wien’s approximation* combines Wien’s displacement law with the Stefan-Boltzmann law. It contains two empirical constants⁹ C_1 and C_2 :

$$\varepsilon(\lambda, T) = \frac{C_1}{\lambda^5} e^{-\frac{C_2}{\lambda T}} \quad (2.1.5a)$$

$$\varepsilon(\nu, T) = \frac{C_1 \nu^3}{c^4} e^{-\frac{C_2 \nu}{c T}} \quad (2.1.5b)$$

– which tends to infinity as the wavelength increases (or the frequency decreases). This is called the “*infrared catastrophe*”.

Max Planck found an interpolation formula, which used empirical constants, in late 1900 and published it in 1901. Later he discovered the derivation of the same formula for a black-body modeled as a cavity whose walls were in thermal equilibrium with the electromagnetic radiation inside the cavity.

$$\varepsilon(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/(\lambda k_B T)} - 1} \varepsilon(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(k_B T)} - 1} \quad (2.1.6a)$$

But this derivation required that the energy of the material of the wall of the cavity and of the electromagnetic field may only change in discrete values of $\varepsilon = h\nu$, where

$$h \approx 6.62 \cdot 10^{-34} \text{ J} \quad (2.1.7)$$

is the *Planck constant*¹⁰, which we now call the *quantum of energy*. This was a significant leap for physics because every physicist (even Planck himself) were sure that electromagnetic energy is a continuous quantity. He tried for some months to get rid of this quantum in his derivation but without success.

In Fig. 2.4 we visually compare the three laws for a black-body of temperature 0.008 K.

Important 2.1.3. *The concept of the existence of an energy quantum cannot be incorporated in classical physics. This was the first and most important step of the development of quantum mechanics.*

⁹Comparing Wien’s with the correct Planck’s formula in the limits we can determine these: $C_1 \equiv 2hc^2$ and $C_2 \equiv \frac{hc}{k_B}$

¹⁰The Planck constant h was introduced first in 1899.

2.2 Photoelectric effect

In 1887 Heinrich Hertz observed that that electrodes illuminated with ultraviolet light create electric sparks easily. It was established later that metal surfaces illuminated by light of suitable frequencies emit electrons. This is called the *photoelectric effect*. The experimental setup is shown in Fig. 2.5. Three electrodes (cathode (C), grid (G) and anode (A)) are sealed inside a *vacuum tube*¹¹. When light illuminates the cathode electrons are emitted from it. These electrons are then accelerated toward the anode by the U voltage producing a current which is measured by an ammeter. The grid electrode is used to measure the kinetic energy of the electrons emitted. When the voltage U_G between the grid electrode and the cathode is such that the i current disappears then the kinetic energy of the electrons is $e \cdot U_g$. It seems easy to explain this behavior by (classical) electrodynamics but the predictions of such a classical model disagree with the measured characteristics of the emitted electron current, which are summarized in Fig. 2.6.

1. Electron emission occurs less than $1 \mu\text{sec}$ after the illumination starts independent of the light intensity. Classical theory predicts a light intensity dependent emission time.
2. increasing the light intensity the current increases linearly, therefore all emitted electrons have the same kinetic energy independent of the light intensity. Classical theory predicts a non-linear relationship.
3. No current is observed if the frequency of the light is less than a threshold frequency, which depends on the metal of the cathode. Classical theory predicts that emission is independent of the frequency.
4. The kinetic energy of the electrons depends only on the frequency of the light but independent of the light intensity. Classical theory predicts that kinetic energy depends on the light intensity.

It was Albert Einstein who explained the measured characteristics of the photoelectric effect¹². He used Planck's idea of the energy quantum and assumed that the energy of

¹¹A.k.a. *electron tube* (in North America), *thermionic valve*, tube, or valve. It is a device controlling electric current through a vacuum in a sealed container for switching and amplification of electrical signals. Before transistors and integrated circuits were developed these vacuum tubes were used in all electronic equipments. In those devices the cathode is heated to a high temperature and produces electrons by the *thermionic effect* (see Section 14.4). The electrons are accelerated by the voltage between the anode and cathode. One or more grids could also be placed inside the tube which may control the current. In the photoelectric experiment the cathode needs not to be heated.

¹²Contrary to popular belief Einstein received the Nobel prize for the explanation of the photoelectric effect and not for his theory of relativity.

the light is also quantized, namely that the light consists of “particles” he called *photons* that have an energy of $h\nu$. The intensity of the light is the number of photons in it, while the energy is the photon energy multiplied by the photon number. Therefore when we illuminate a surface with light of frequency ν we bombard it with photons of energy $h\nu$. An electron is emitted from the material only when an incoming photon, that has enough energy to overcome the electron binding energy W , called the *work function*¹³.

$$h\nu = \frac{1}{2} m_e v^2 + W \quad (2.2.1)$$

where W is the work function, v is the velocity and, $m_e = 9.1 \cdot 10^{-31} \text{ kg}$ is the mass of the electron.

Important 2.2.1. *In quantum mechanics the energy is usually measured in electronvolts (eV). 1 eV is the energy an electron obtains if accelerated through 1 V. Because $\Delta\mathcal{E} = eU$, where U is the potential difference measured in volts and e is the elementary charge, which is $\approx 1.60 \cdot 10^{-19} \text{ C}$:*

$$1\text{eV} \approx 1.6 \cdot 10^{-19} \text{ J}$$

*Units of milli-, kilo-, mega-, giga-, tera- or peta- electronvolts (meV, keV, MeV, GeV, TeV and PeV respectively) are also used in practice*¹⁴.

The probability of multiple “collisions” is negligible at normal light intensities, therefore if the frequency is lower than a threshold frequency (i.e. the photon energy is smaller than the work function) no electrons are emitted. As the emission of an electron occurs immediately after the arrival of a photon with a frequency above the threshold the intensity of the light does not matter. Furthermore increasing the light intensity only increases the number of photon collisions and not the kinetic energy of the electrons, therefore the current, which depends on the number and velocity of the electrons, will be proportional with the intensity, as the velocity (kinetic energy) of the electrons depends only on the frequency.

Important 2.2.2. *The photoelectric effect can only be explained by assuming the existence of photons: the discrete energy quanta of electromagnetic radiation. These photons behave like particles. They may collide with electrons for instance. Light (or electromagnetic radiation of any frequency) is emitted and absorbed in quanta. Therefore electromagnetic radiation must be a corpuscular phenomena. But diffraction and interference experiments can only be explained by assuming that light is a wave. This is the so called particle-wave duality of the electromagnetic radiation.*

¹³See also section 14.4.

¹⁴The velocity of an electron with a kinetic energy of 1 eV calculated using classical Newtonian (non-relativistic) mechanics is $v = \sqrt{\frac{2 \cdot e}{m_e}} = 1.5 \cdot 10^6 \text{ m/s} = 0.015 \cdot c$, where $c \approx 3 \cdot 10^8$ is the velocity of light in vacuum. Using the correct relativistic formula, which ensures that the velocity of the electron never reaches c no matter the kinetic energy of the electron yields the lower value of $v = 5.93 \cdot 10^5 \text{ m/s}$

Example 2.6. Determine the work function of potassium in electronvolts knowing that when illuminated by a light with a wavelength of $\lambda = 560\text{nm}$ it emits electrons with a velocity of 190 km/s ! **Solution** From equation (2.2.1)

$$W = h\nu - \frac{1}{2}m_e v^2 = h\frac{c}{\lambda} - \frac{1}{2}m_e v^2 = 3.38 \cdot 10^{-19}\text{ J} = 2.11\text{ eV}$$

Example 2.7. Determine the maximum speed of a photoelectron emitted from a chromium surface when illuminated with light of a wavelength of 180 nm , from knowing that at a wavelength of 150 nm the maximum photoelectron energy is 3.92 eV ? How large is the work function? ($m_e = 9.1 \cdot 10^{-31}\text{ kg}$) **Solution** Let $\lambda_1 = 1.8 \cdot 10^{-7}\text{ m}$ and $\lambda_2 = 1.5 \cdot 10^{-7}\text{ m}$ and the maximum photoelectron kinetic energy at λ_2 $\mathcal{E}_{kin}(\lambda_2) = 3.92\text{ eV}$. From equation (2.2.1) and using $\nu = c/\lambda$ the work function can be determined:

$$W = h\frac{c}{\lambda_2} - \mathcal{E}_{kin}(\lambda_2) = 6.96 \cdot 10^{-19}\text{ J} = 4.35\text{ eV}$$

Therefore the maximum velocity at λ_1 :

$$v(\lambda_1) = \sqrt{\frac{2}{m_e} \left(h\frac{c}{\lambda_1} - W \right)} = 945,970\text{ m/s}$$

2.3 Compton effect

By the early 20th century, researchers found that when X-rays of a known wavelength interact with electrons, the X-rays are scattered through an angle θ and emerge at a *different wavelength* related to θ . Although classical electromagnetism predicted that the wavelength of scattered rays should be equal to the initial wavelength, multiple experiments found that the wavelength of the scattered rays was longer (corresponding to lower energy) than the initial wavelength. That is, regardless of light intensity inelastic scattering *always* occurs when the frequency of the light is high enough (so that photon energies are in the range corresponding to the electron rest mass: $m_e c^2 = 511\text{ keV}$) and the scattering angle is not zero.

According to classical electrodynamics the incident harmonic electromagnetic wave accelerates the charged particle which, in turn, then emits an electromagnetic radiation of the same frequency as of the incident wave. As long as the velocity of the particle is much smaller than the speed of light in vacuum the magnetic component of the electromagnetic wave does not affect the motion of the particle. The resulting, scattered wave therefore will have the same frequency as the original one¹⁵.

¹⁵Classically, the electric field in light of sufficient intensity may accelerate a charged particle to relativistic speeds, which will cause radiation-pressure recoil and an associated Doppler shift of the scattered light, but the effect would become arbitrarily small at sufficiently low light intensities regardless of wavelength.

Light scattering on free electrons (or on other charged particles) can be elastic or inelastic. Elastic scattering in which neither the particle kinetic energy, nor the frequency of the light changes is called *Thomson scattering*. Inelastic scattering in which both the energy and momentum of the electron (or any charged particle) and the frequency of the light changes – the frequency of the light always decreases – is called *Compton scattering*, and the frequency shift of the light is the *Compton effect*. Thomson scattering is the low energy limit of Compton scattering.

Assuming light consists of photons which can collide with electrons we can easily explain the observed behavior, by applying the (relativistic) energy and momentum conservation laws¹⁶. Details of the derivation are in Appendix 22.3. The result is

$$\lambda' - \lambda = \frac{h}{m_e c} (1 - \cos \theta), \quad (2.3.1)$$

where λ is the initial wavelength, λ' is the wavelength after scattering, h is the Planck constant, m_e is the *electron rest mass*, c is the speed of light, and θ is the scattering angle. The quantity

$$\frac{h}{m_e c} = 2.43 \cdot 10^{-12} \text{ m} \quad (2.3.2)$$

is known as the *Compton wavelength* of the electron. The amount $\Delta\lambda = \lambda' - \lambda$ the wavelength changes by is called the *Compton shift*. It is between zero (for $\theta = 0^\circ$) and twice the Compton wavelength of the electron (for $\theta = 180^\circ$).

Important 2.3.1. *The Compton effect is another phenomenon that can only be explained by assuming the existence of photons.*

Example 2.8. *Calculate the scattering angle and the energy transferred to the electron compared to the energy of the incoming photon in a Compton effect, if at wavelength $\lambda = 0.01 \text{ nm}$ $\Delta\lambda = 0.0024 \text{ nm}$.* **Solution** From (2.3.2) the Compton angle is

$$\cos \theta = 1 - \frac{m_e c \Delta\lambda}{h} = 0.989 \quad \Rightarrow \quad \theta = 8.445^\circ$$

The energy transferred to the electron is

$$\mathcal{E}_e = h c \left(\frac{1}{\lambda} - \frac{1}{\lambda'} \right) = h c \left(\frac{1}{\lambda} - \frac{1}{\lambda + \Delta\lambda} \right) = 3.84 \cdot 10^{-15} \text{ J} = 24 \text{ keV}$$

The energy of the incoming photon according to the theory of relativity is

$$E_{ph} = h \nu = h c / \lambda = 1.99 \cdot 10^{-14} \text{ J, i.e. } \frac{\mathcal{E}_e}{E_{ph}} = 0.19.$$

¹⁶Because photons are characterized by their frequency the photon after the scattering is not the same particle as the photon before the scattering.

Example 2.9. *What will be the momentum of the Compton electron if for $\lambda = 0.005 \text{ nm}$ the photon scattering angle is 90° ?* **Solution** If the Compton angle is 90° then $\cos \theta = 0$ and

$$\lambda' = \lambda + \frac{h}{m_e c} = 7.426 \cdot 10^{-12} \text{ m} = 0.007426 \text{ nm}$$

Because of the momentum conservation the total momentum of the electron after the collision equals to the total momentum difference between the incoming and outgoing photons. The photon momentum and energy is connected by the formula $p_{\text{photon}} = \mathcal{E}_{\text{photon}}/c = h\nu/c$. Therefore

$$\Delta p_e = \frac{h\nu}{c} - \frac{h\nu'}{c} = \frac{h}{\lambda} - \frac{h}{\lambda'} = 4.33 \cdot 10^{-23} \text{ kg m s}^{-1}$$

Solar Radiation Spectrum

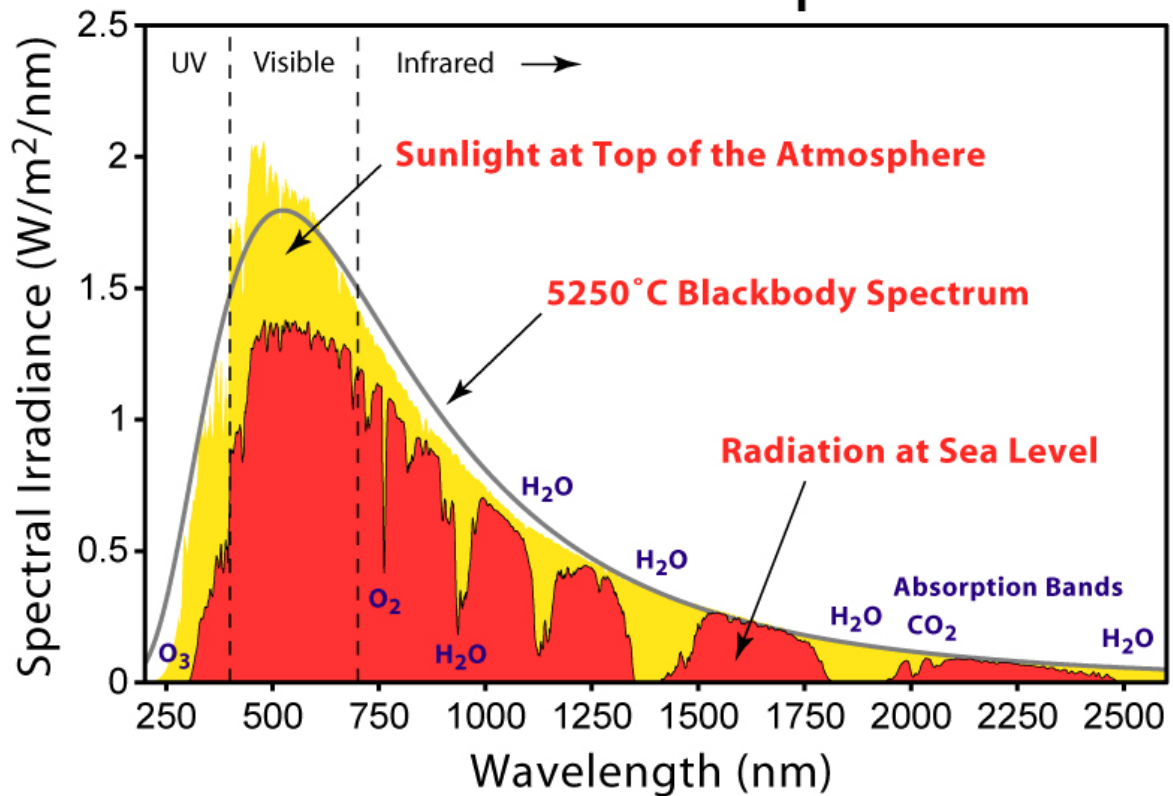


Figure 2.3: Emission spectrum of the Sun. When measured above the atmosphere (yellow) the spectrum resembles that of a black-body whose temperature is 5250 K. At sea (ground) level the spectrum (red) is distorted because gases in the atmosphere absorb radiation at some wavelengths.

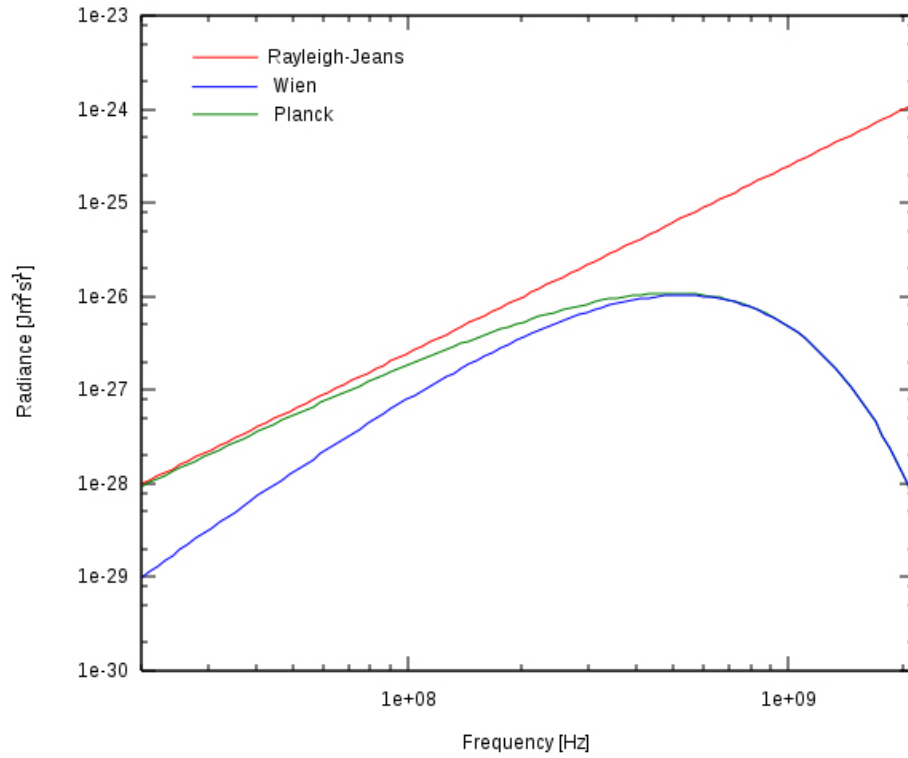


Figure 2.4: Comparison of the Rayleigh-Jeans, Wien and Planck formulas for a black body of temperature 8 mK.

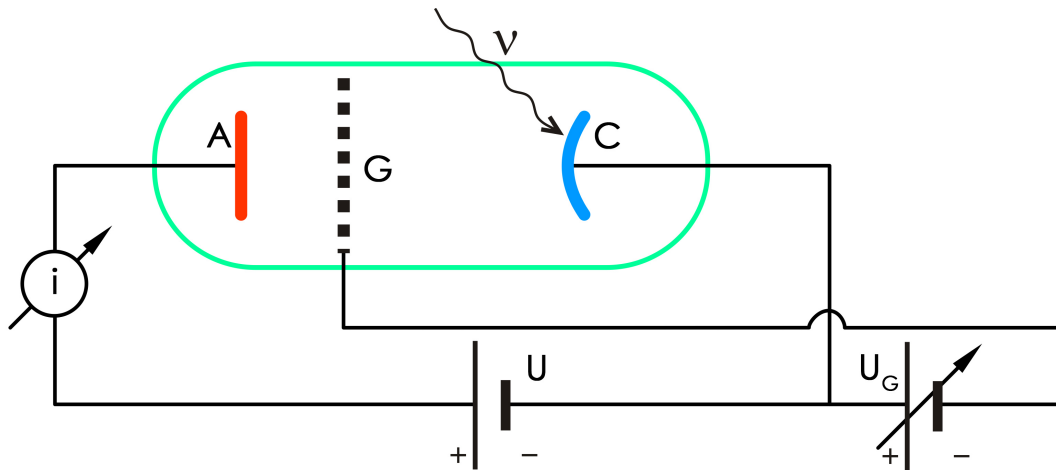


Figure 2.5: Experimental setup of the measurement of the photoelectric effect. For description see the main text.

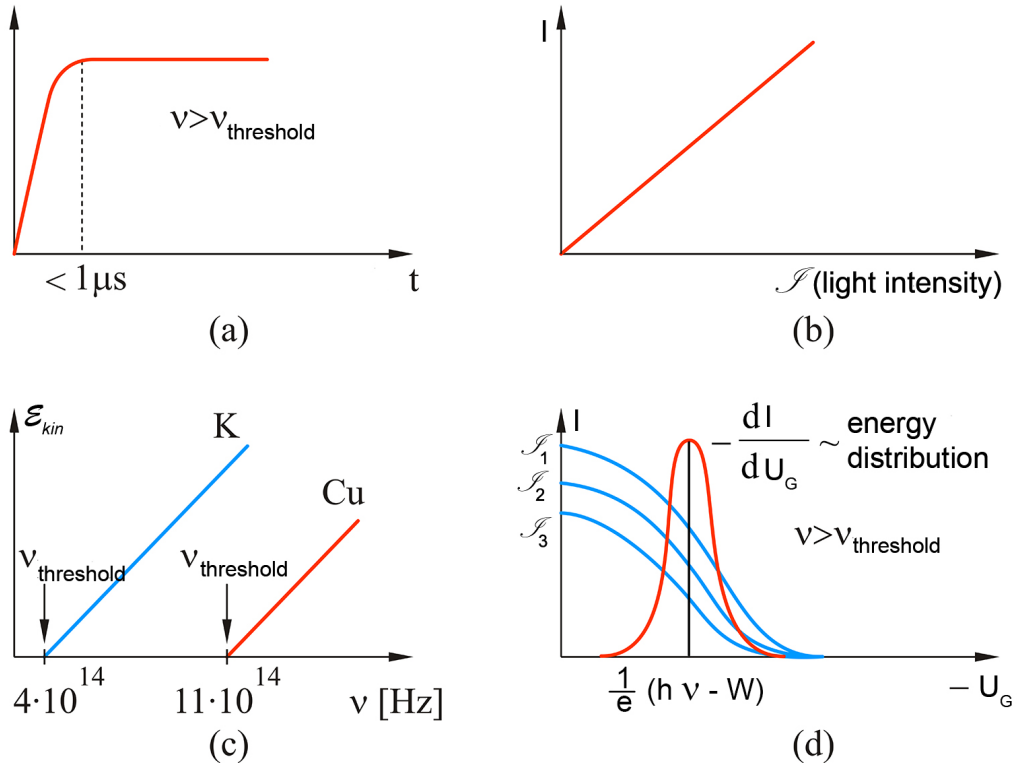


Figure 2.6: The photoelectric effect. a) current vs emission time, b) current vs light intensity, c) kinetic energy vs frequency curve for K and Cu, d) The kinetic energy of the electrons ($\frac{\mathcal{E}_{\text{kin}}}{e} = U_G$) is independent of the light intensity. It only depends on the frequency and a material specific W constant.

Chapter 3

Stationary states

3.1 Stationary States

One of the phenomena, mentioned above, which could not be explained by classical physics, is the stability of atoms, molecules and even the stability of the atomic nucleus.

According to classical electromagnetic theory¹ there can be no stable equilibrium in a system of charged particles if no other interaction is present. As we now know all mechanical interactions between macroscopic bodies are in fact electrostatic interactions of the outermost atoms of the interacting objects, so this means that the system of charged particles must be in continuous motion, otherwise no stable macroscopic or microscopic bodies could exist. On the other hand if moving charged particles are confined into a finite volume of space they cannot move with constant velocity, therefore must accelerate. But, according to the classical theory of electromagnetism, all accelerating charges emit electromagnetic radiation, which means energy loss. Therefore neither the atoms, nor the molecules could be stable. The calculation of the energy loss rate of the single electron in a hydrogen atom for instance yields that the time during which all kinetic energy of the electron is lost is less than 10^{-8} sec. As the universe is older than that clearly something must be wrong with this picture.

The discrepancy between the theoretical predictions and the experiments can be resolved if we assume, that there exist *discreet stationary states* of a system of charged particles in which they do not emit electromagnetic radiation contrary to the prediction of classical (macroscopic) electromagnetic theory. This was exactly what Bohr proposed to explain the observed line spectra of atoms.

Important 3.1.1. *The existence of stationary electronic states does not mean that the laws of electrodynamics are invalid for electrons. Neither do they mean that there exists a special kind of “quantum mechanical interaction” between particles. There is no*

¹which itself was only discovered in the second half of the 19th century.

such thing. There are only four fundamental interactions (sometimes called fundamental forces), namely electromagnetism, strong interaction (or strong nuclear force), weak interaction (or weak nuclear force) and gravitation.

All of the “forces” we encounter in quantum mechanics are electromagnetic forces. What must be modified is the classical concept of the electron being a classical charged mass point.

So the fact electrons in stationary states in an atom do not emit electromagnetic radiation means that – contrary to classical physics – they *do not accelerate*, i.e. they are not orbiting the atom in a classical way.

But when charged particles change their state as a result of an interaction with their environment then during the transition from one stationary state to an other one they either emit or absorb energy. Energy may be absorbed for instance in a collision with a photon of suitable frequency, and emitted in the form of a photon of the same frequency.

This picture can explain for instance the absorption and emission spectra of atoms. We will denote stationary (bound) atomic or electronic state “A” by $|A\rangle$. The energy of the electron in state $|A\rangle$ will be $\mathcal{E}(|A\rangle)$. As we will see the energy of electrons in stationary states in atoms can have only discrete values $\mathcal{E}(|A\rangle), \mathcal{E}(|B\rangle)$, etc. An atom originally in the discrete stationary state $|A\rangle$ may absorb a photon and change its state to another, higher energy discrete state $|B\rangle$ if and only if the following criterion is met:

$$\mathcal{E}(|B\rangle) - \mathcal{E}(|A\rangle) = \Delta\mathcal{E}(|A\rangle \rightarrow |B\rangle) = h\nu$$

In the opposite process the atom goes from the higher energy state to the lower energy one with an emission of a photon with same ν frequency:

$$\Delta\mathcal{E}(|B\rangle \rightarrow |A\rangle) = -h\nu$$

The discrete nature of atomic energies can be observed by experiments in which no photons are involved at all.

The most famous such experiment was performed by *James Franck* and *Gustav Luis Hertz* in 1914². The schematic of the *Franck-Hertz experiment* can be seen in Fig. 3.1.

What would we expect from this experiment *if we assume* that atomic stationary states have discrete (quantized) energy values (called *energy levels*)? For simplicity let us assume mercury atoms have two possible stationary states with discrete energies (energy levels) with an $\Delta\mathcal{E}$ energy difference between them.

Until the kinetic energy of the emitted and subsequently accelerated electrons is smaller than $\Delta\mathcal{E}$ no interaction is possible with the Hg atoms, because the atoms can

²They were awarded the Nobel Prize in 1925 for this work.

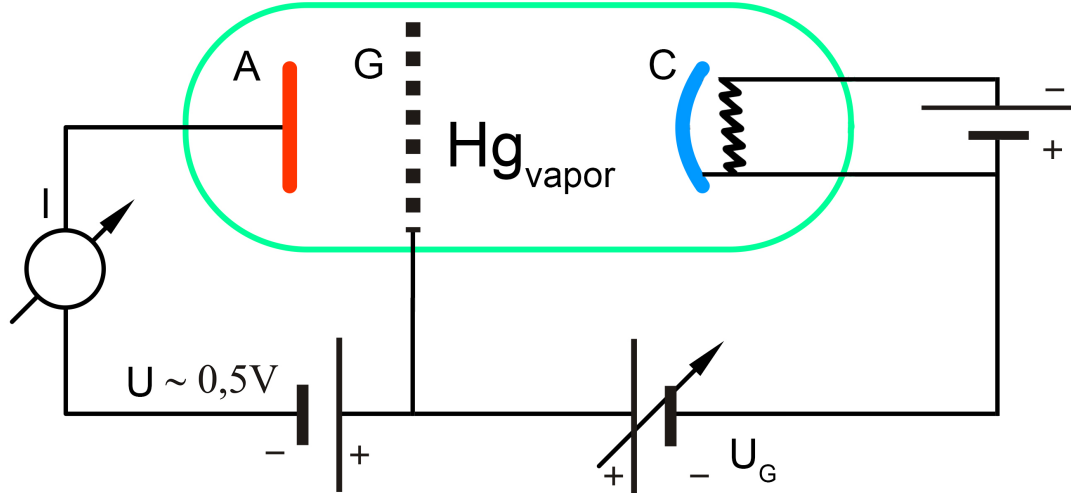


Figure 3.1: Experimental setup of the Franck-Hertz experiment. Electrons are generated by a heated cathode in a glass vacuum tube filled with low pressure mercury gas. They are accelerated by the variable voltage between the cathode and the grid. The anode is held at a slightly negative potential relative to the grid, so that only the electrons with suitable kinetic energy may reach it. The accelerating voltage is varied and the I current is measured.

only absorb $\Delta\mathcal{E}$ energy not less and not more. Therefore only elastic electron–atom collisions are possible, which do not change the kinetic energy of the electrons, only randomize the direction of their velocity. This means that the current will increase steadily after the acceleration voltage is higher than the grid voltage 0.5 V.

When the kinetic energy of the electrons reaches $\Delta\mathcal{E}$ the current should drop almost to zero, because now inelastic scattering leading to kinetic energy loss may also occur³.

A further increase of the acceleration voltage will lead to the increase of the current again, because electrons can only lose $\Delta\mathcal{E}$ of their kinetic energy in a collision and not more. Until the remaining kinetic energy is smaller than $\Delta\mathcal{E}$ no further inelastic collision may occur. However when the average electron kinetic energy reaches $2\Delta\mathcal{E}$ the current drops again, as electrons now have enough kinetic energy to participate in *two* consecutive inelastic collisions⁴. And this periodic increase and decrease of the current

³The current will not drop exactly to zero because of two reasons. First there will be electrons that travel from the cathode to the grid without any collisions and, second, some percentage of the electrons will still collide elastically with the atoms, as not all electrons will have the same amount of kinetic energy.

⁴The minimum current will be higher than in the previous case, because the probability of two

will repeat at $3\Delta\mathcal{E}$, $4\Delta\mathcal{E}$, etc, i.e. at every multiple of $\Delta\mathcal{E}$.

In Fig. 3.2 part of the results of the original experiment, which clearly displays the expected behavior is shown. From the figure we can determine that for mercury

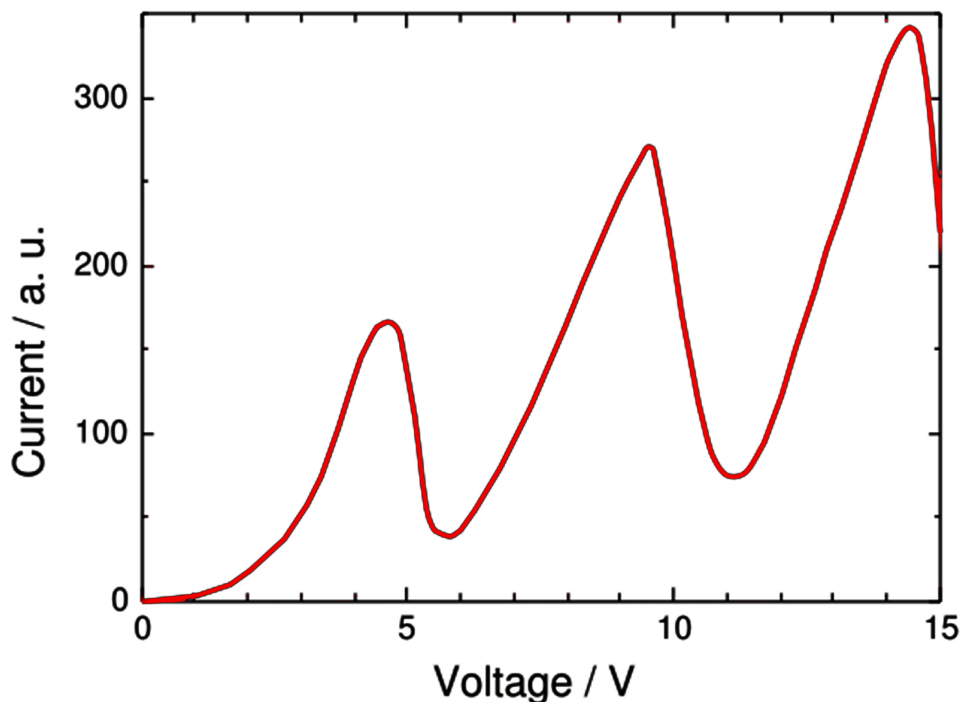


Figure 3.2: Accelerating voltage vs. anode current. The current is displayed in arbitrary units (a.u.)

$$\Delta\mathcal{E} = 4.9 \text{ eV}$$

This periodic current may be observed up to at least 100 volts. The same experiment with neon gas gives $\Delta\mathcal{E}_{\text{neon}} = 19 \text{ eV}$.

Important 3.1.2. *The Franck-Hertz experiment proves that stationary atomic energy values are discrete. Clearly the quantized nature of some physical quantities is a basic law of nature.*

Example 3.1. *The ground state and the first excited state (the stationary states with the smallest and the next lowest energy) in a hydrogen atom have an energy of $\mathcal{E}_0 = -13.6 \text{ eV}$ and $\mathcal{E}_1 = -3.4 \text{ eV}$ respectively relative to the energy of the free electron. What*

consecutive inelastic collisions is smaller than the probability of a single one.

is the frequency of the photon that, when absorbed, can excite the electron from the ground state to the first excited state? What will be the frequency of a photon emitted during the $\mathcal{E}_1 \rightarrow \mathcal{E}_0$ transition? **Solution** For a photon to be absorbed the photon energy must equal to the energy difference of the two states in question:

$$\begin{aligned} h\nu = \mathcal{E}_1 - \mathcal{E}_0 = 10.2 \text{ eV} = 1.634 \cdot 10^{-18} \text{ J} &\Rightarrow \\ \nu = 2.47 \cdot 10^{15} \text{ Hz} \end{aligned}$$

The frequency of the photon emitted in the reverse transition must be the same as that of the absorbed photon.

3.2 Wave-particle duality

In previous sections we encountered the dual nature the electromagnetic waves. In some cases (e.g. diffraction and interference) they behave like classical waves, in other cases (e.g. Compton effect, photoelectric effect) like classical particles (photons). We also saw that stationary states of physical systems may (and usually do) have discrete energy values. A valid question therefore whether particles (e.g. electrons) may exhibit wave-like behavior too.

This possibility first appeared in an 1924 paper of the French physicist and Nobel laureate Louis de Broglie. According to the *de Broglie hypotheses* any moving particle or object had an associated wave, with a wavelength determined by its momentum p :

$$\lambda = \frac{h}{p} \quad (3.2.1a)$$

With the introduction of the *wave number*

$$k = \frac{2\pi}{\lambda} \quad (3.2.1b)$$

this formula can be written as

$$p = \frac{h}{\lambda} = \hbar k \quad (3.2.1c)$$

where

$$\hbar \equiv \frac{h}{2\pi}$$

is the *reduced Planck constant*⁵. We can define the frequency of this wave by the energy \mathcal{E} as we did for photons:

$$\mathcal{E} = h\nu = \hbar\omega \left(= \frac{p^2}{2m_e} = \frac{\hbar^2 k^2}{2m_e} \right) \quad (3.2.1d)$$

⁵Pronounced “h-bar”. It is also known as *Dirac’s constant*.

If the de Broglie hypothesis is true, then diffraction and interference patterns should be observable in experiments involving only electrons and other particles. Indeed *electron diffraction* experiments performed on thin metal foils clearly show these patterns. In Figure 3.3 for instance we can compare X-ray and electron diffraction measurements performed on the same aluminum foil. Both the electromagnetic X-rays and low energy electrons have wavelengths of the same magnitude.

Example 3.2. *Determine the wavelength of an electron that is accelerated through a voltage U . What magnitude of voltage must be used to have a wavelength comparable to atomic distances around 0.05-10 nm in solids?* **Solution** The kinetic energy of an electron of momentum p is $\mathcal{E}_{kin} = \frac{p^2}{2m_e}$. If the electron is accelerated through a U voltage $\mathcal{E}_{kin} = eU$ (e is the elementary charge). The corresponding momentum is

$$p = \sqrt{2m_e e U}$$

The de Broglie wavelength

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m_e e U}}$$

Therefore the accelerating voltage for λ is

$$U = \frac{h^2}{2m_e e \lambda^2}$$

For wavelengths 0.05 nm and 10 nm the required voltages are:

$$U(0.05 \text{ nm}) = 601.7 \text{ V} \text{ and } U(10 \text{ nm}) = 0.015 \text{ V}$$

The Double-slit experiment with light and with electrons

The famous double-slit experiment illustrates best the difference between wave-like and particle-like behavior. The original experiment used light, but the double-slit experiment has been replicated with electrons (in 1961), atoms, and even entire molecules (in 1999). The principle of the original experiment is as follows: a light source producing *coherent* plane waves⁶ illuminates a thin plate pierced by two parallel slits. The light passing thorough the slits then hits a screen. Because the light is (or more exactly as we have seen before: may behave like) a wave we expect an interference pattern on the screen⁷. See Fig. 3.4.

⁶Light is said to be *coherent* if it can be split into two or more parts whose relative phases, when united again after traveling paths of different lengths does not change so the interference patterns produced

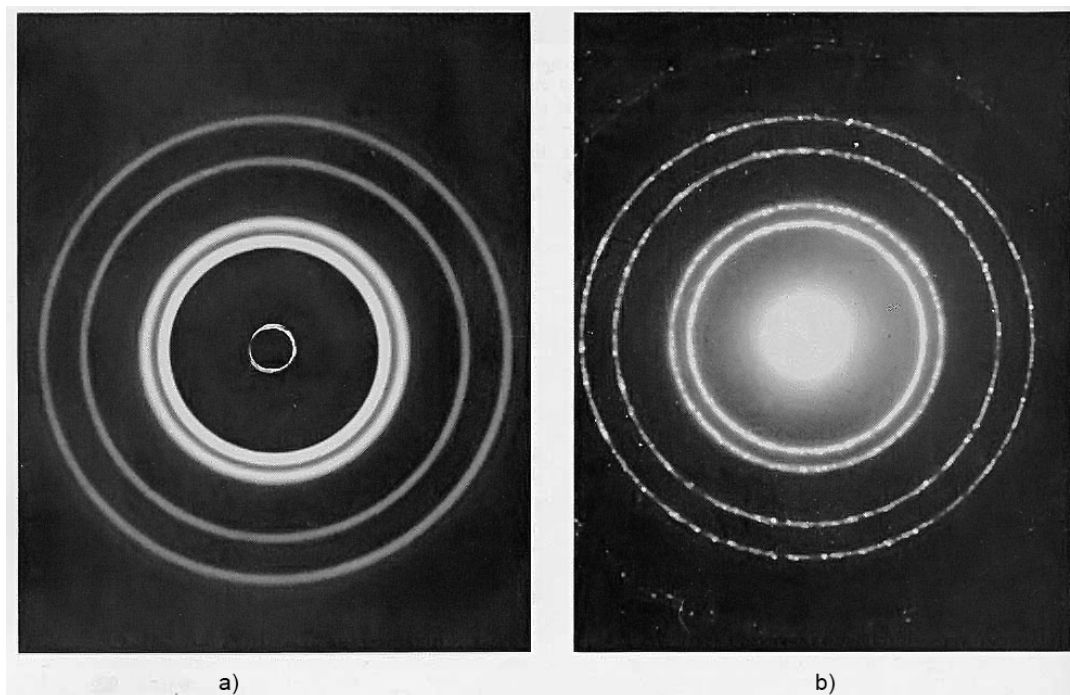


Figure 3.3: Comparison of X-ray and electron diffraction experiments. a) X-ray diffraction on a thin aluminum foil. b) an electron beam directed through the same foil gave this diffraction picture.

However when one of the slit is covered and light may only travel through the other one the interference vanishes⁸ (Fig. 3.5).

These experiments conform to the expected behavior from a wave. If we put a photodetector (photomultiplier, CCD or CMOS sensor, etc) near one of the slits we either detect a photon or nothing at all again since the energy in electromagnetic waves are in (particle like) photons. This makes it possible to determine which one of the slits the actual photon went through. Unexpectedly however, if we do this, the interference pattern vanishes and the screen shows the sum of two overlapping intensity peaks, corresponding to the two slits (See Fig. 3.6.) Therefore it may seem logical that the wave and particle properties of the light (or any particle whatsoever) are *complementary*: both cannot be observed at the same time.

are stationary in time. Such light may be produced e.g. a monochromatic laser as a light source.

⁷The same interference pattern forms regardless of the light intensity only the required measurement times vary. In 1909 this was proved in an experiment where such low light intensities were used that only a single photon was present in the device at one time. In this case the interference pattern is built-up the same way as is shown in Fig. 3.7 for electrons.

⁸You can still observe a *diffraction* pattern on c).

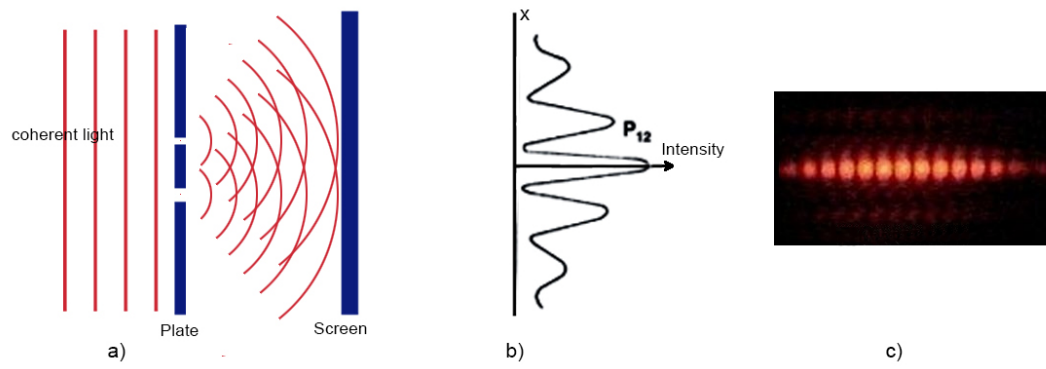


Figure 3.4: The double slit experiment 1. a) schematics with both slits open, b) intensity measured on the screen, c) observable interference pattern

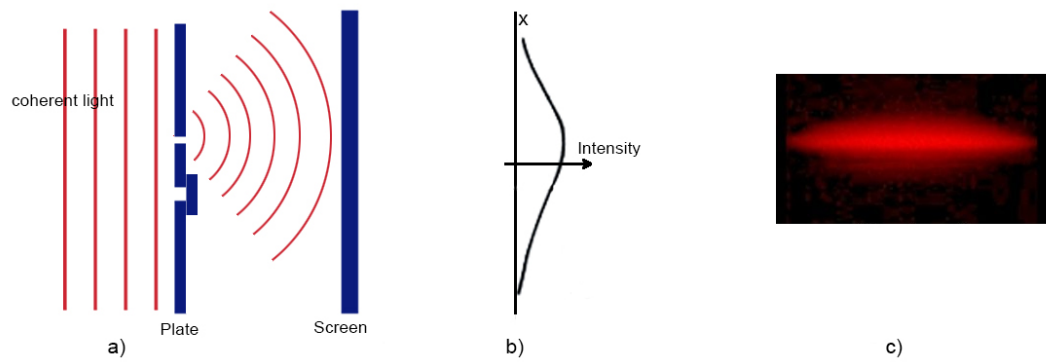


Figure 3.5: The double slit experiment 2. a) schematics with one slit covered, b) intensity measured on the screen, c) observable interference pattern

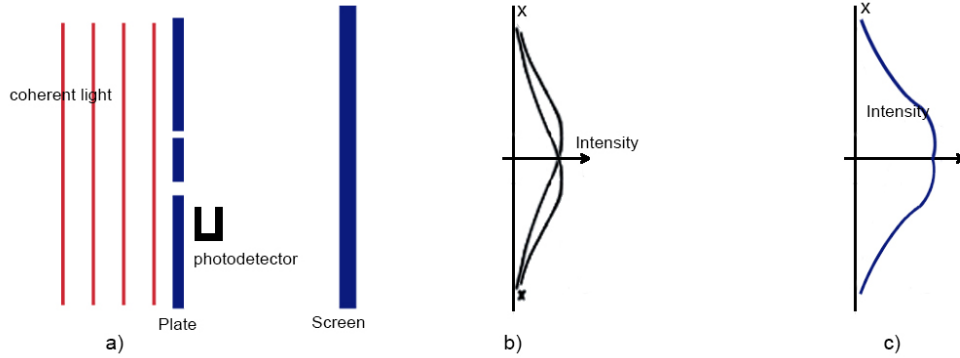


Figure 3.6: The double slit experiment 3. a) setup with a photon detector, b) intensities of the two independent slits, c) the intensity on the screen is the *sum* of the intensities of the two slits

Important 3.2.1. *The wave and particle like behavior of either particles and waves are both present simultaneously always. Which behavior manifests itself depends on the properties of the object (e.g. photon or electron) and on the parameters of the experiment.*

The measured intensity of a wave is the square of the absolute value of the instantaneous complex amplitude $A(x, t) = A_0 e^{i(\omega t - k r)}$. The difference between the interference pattern in Fig. 3.4 and the intensity curve in Fig. 3.6 is that in the first case, when the wave-like property of the photons is dominant, the resulting intensity is the square of the sum of the instantaneous amplitudes:

$$\mathcal{I}(x, t) = (\mathcal{A}_1(x, t) + \mathcal{A}_2(x, t))^2 = \mathcal{A}_1^2 + \mathcal{A}_2^2 + 2\mathcal{A}_1 \cdot \mathcal{A}_2$$

The third term in this expression is the *interference term*. It may be positive or negative depending on the phase difference between \mathcal{A}_1 and \mathcal{A}_2 .

When the particle-like property of the photons⁹ is dominant the resulting intensity is the sum of the intensities from the slits:

$$\mathcal{I}(x, t) = \mathcal{I}_1(x, t) + \mathcal{I}_2(x, t) = \mathcal{A}_1^2(x, t) + \mathcal{A}_2^2(x, t)$$

The difference being the missing interference term in the second case.

At first thought it may seem reasonable that the interference pattern is a result of some kind of interaction between the particles. However as experiments show the interference pattern can be detected even in cases when the particle flux is so low that only a single particle is present in the system at any time. In this case the individual particles arrive at the screen at seemingly random positions, but this randomness is not

⁹there are no fractional photons because photons cannot be split

uniform. More particles arrive around positions where in the fully formed interference pattern the maxima are located and almost no particle hits the screen at the minimum positions. Fig. 3.7 shows what happens in such an experiment. The exact position

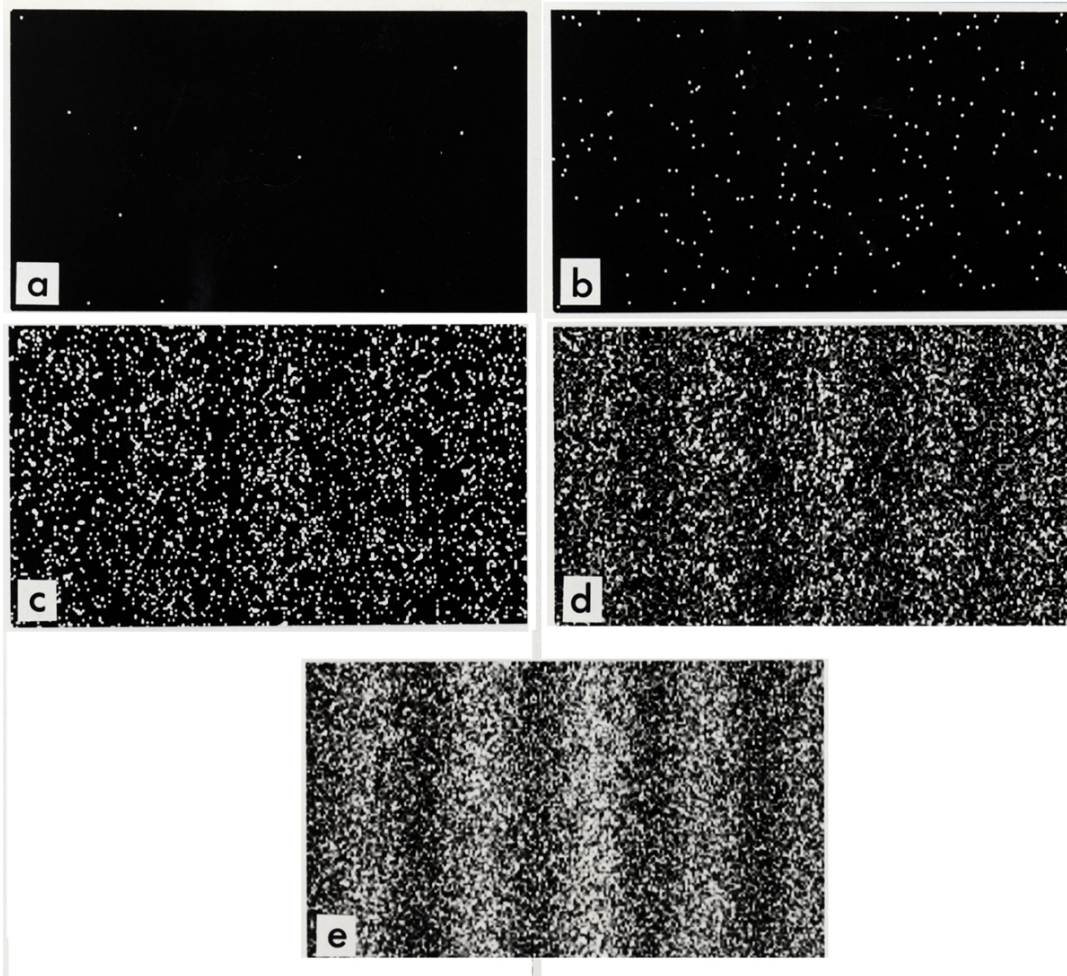


Figure 3.7: Simulation of a double-slit experiment with only a single electron present in the system at any given time.

where the next electron hits the screen cannot be calculated. The only thing that *can* be calculated is the *probability* of an electron hitting a given point¹⁰. These probabilities are not random they are higher at the interference maxima and lower at the interference minima. The more electrons arrive at the screen the easier to recognize the interference pattern.

¹⁰This probability is not the sum of the individual probabilities, it is square of the sum of the *probability amplitudes*, which – as we will see later – are connected of the wave function.

3.3 Uncertainty relations

The conclusion of the previous section, that physical objects show both the wave-like (characterized by λ) and particle-like (characterized by x) behavior at all times, means that certain concepts of classical physics must be reconsidered in quantum mechanics. There is no such thing as a particle at an exact spatial position or a particle with an exactly determined momentum for instance.

If we had a particle with an exact momentum its state should be described by a wave function which has an exact $\lambda = h/p$ wavelength. But such functions, e.g.

$$\sin(\omega t - \mathbf{k}\mathbf{r}) \quad \text{where} \\ |\mathbf{k}| = \frac{2\pi}{\lambda} = \frac{|\mathbf{p}|}{\hbar}$$

have the same amplitude in the whole infinite space which is clearly impossible if we want to ascribe any physical meaning to the wave function.

On the other hand electrons are point-like particles with no known substructure¹¹.

The nearest concept of a localized physical object (e.g. electron) therefore is a *wave packet*, which – according to Fourier analysis – is a superposition of an infinite number of plain waves. Neither exact position nor exact momentum can be attributed to a wave packet however. But there exist relations between the *uncertainty* of the momentum and position. As it turns out these are physical quantities whose uncertainties are not independent and there is a non-zero minimum value of their product¹²:

$$\begin{aligned} \Delta x \cdot \Delta p_x &\geq \frac{\hbar}{2} \\ \Delta y \cdot \Delta p_y &\geq \frac{\hbar}{2} \\ \Delta z \cdot \Delta p_z &\geq \frac{\hbar}{2} \end{aligned} \tag{3.3.1}$$

$$\tag{3.3.2}$$

These are the *Heisenberg uncertainty relations*¹³. The pair of quantities which are related by an uncertainty relation are called *canonical conjugates* or *conjugate variables*. A zero uncertainty of either of the conjugate variables would require an infinite uncertainty of

¹¹The 'diameter' of an electron given as $2.8179 \cdot 10^{-15}$ m is calculated from the assumption that it has a homogeneous charge density with an electrostatic self-energy that is equal to its mass-energy of about 511 keV. This however has nothing to do with the fundamental structure of the electron and highly inaccurate. Observation of a single electron in a Penning trap (a device that uses a strong homogeneous magnetic and an inhomogeneous electric field to confine charged particles) shows the upper limit of the electron radius is 10^{-22} meters.

¹²See Appendix 22.4 for a detailed derivation of this minimum value for a Gaussian wave packet

¹³No such non-zero minimum uncertainty exists for the product of uncertainties of the individual

the other member of the pair. Neither the zero nor the infinite uncertainty states are physical states.

The $\frac{\hbar}{2}$ numerical constant on the right hand side of the (3.3.1) relation is the mathematically correct value, see Appendix 22.4¹⁴.

Originally the uncertainty principle was considered as an example of the *observer effect*, which notes that measurements of certain systems cannot be made without affecting the systems. According to this explanation the interference pattern in the double-slit experiments when the slit the electron passed through is identified was destroyed, because to observe an electron in the vicinity of the slit requires an external interaction which affects the system. But this is not the real reason for the uncertainty relations. It has become clear since then that the uncertainty is simply due to the *matter wave* nature of all objects.

Important 3.3.1. *The uncertainty relations reflect the fact that a particle can never have an exact value of either of its conjugate variables at any time. There are always some inherent uncertainty in the values of conjugate variables in any physical state and this is independent of whether the particle interacts with other objects or not.*

Position and momentum are conjugate variables therefore it is not possible for an electron to have either an exact position or an exact momentum.

This means that such classical concepts as a particle at rest or electron trajectories do not exist. This violation of the classical concepts however can only be observed for very small objects: particles, atoms, molecules, etc, as it will be clear from the next example.

Example 3.3. *What is the momentum and velocity uncertainty for a) a dust particle of diameter $500\ \mu$ and mass of about $5.4 \cdot 10^{-4}\text{ mg}$, b) an ammunition bullet with a size of about $7 \times 40\text{ mm}$ and mass 5.2 g , c) a $75\text{ kg } 1.8\text{ m} \times 40\text{ cm} \times 20\text{ cm}$ object if all of them are seemingly at rest. Solution If these objects are at rest then the position uncertainty equals to their size. Therefore*

$$\Delta p = \frac{\hbar}{2 \cdot \text{size}}, \quad \Delta v = \frac{\hbar}{2 \cdot m \cdot \text{size}}$$

components of the same components of \mathbf{r} or \mathbf{p} or of the different components of \mathbf{r} and \mathbf{p} , for instance:

$$\begin{aligned} \Delta x \cdot \Delta y &\geq 0 \\ \Delta x \cdot \Delta p_y &\geq 0, \end{aligned}$$

but there are physical quantities (e.g. the angular momentum) between whose components an uncertainty relation does exist.

¹⁴Popular scientific books may use $\frac{h}{2}$ or even with h instead of the correct $\frac{\hbar}{2}$. These are less rigorous estimations for the minimum uncertainty and are all larger than $\frac{\hbar}{2}$

a)	$\Delta x = 5 \cdot 10^{-4} m,$	$\Delta p = 1.06 \cdot 10^{-31} kg \frac{m}{s}$
		$\Delta v = 1.95 \cdot 10^{-25} \frac{m}{s}$
b)	$\Delta x_1 = 7 \cdot 10^{-3} m,$	$\Delta p_1 = 7.54 \cdot 10^{-33} kg \frac{m}{s}$
		$\Delta v_1 = 1.45 \cdot 10^{-30} \frac{m}{s}$
	$\Delta x_2 = 40 \cdot 10^{-4} m,$	$\Delta p_2 = 1.31 \cdot 10^{-33} kg \frac{m}{s}$
		$\Delta v_2 = 2.54 \cdot 10^{-31} \frac{m}{s}$
c)	$\Delta x = 1.8 m,$	$\Delta p = 2.93 \cdot 10^{-35} kg \frac{m}{s}$
		$\Delta v = 3.91 \cdot 10^{-37} \frac{m}{s}$
	$\Delta x_1 = 0.4 m,$	$\Delta p_1 = 1.32 \cdot 10^{-34} kg \frac{m}{s}$
		$\Delta v_1 = 1.75 \cdot 10^{-36} \frac{m}{s}$
	$\Delta x_2 = 0.2 m,$	$\Delta p_2 = 2.63 \cdot 10^{-34} kg \frac{m}{s}$
		$\Delta v_2 = 3.52 \cdot 10^{-36} \frac{m}{s}$

As you can see the momentum and velocity uncertainties are too small to be measured. That is the reason why we may say these objects are at rest.

For macroscopic bodies the uncertainties are so small that no measurement is sensitive enough to show them, therefore in these cases there is no need to modify the classical mechanical concepts of position, velocity, trajectory, etc. But these concepts can also be applied to elementary particles like electrons, if the accuracy required is small. For instance when an electron moves between the plates of a plain capacitor classical mechanics can calculate its trajectory. In this case however the required Δx accuracy is low it is in the $0.1 - 1 mm$ range, which gives a minimum uncertainty of momentum to be $\approx 10^{-32} - 10^{-31} kg m/s$. This still gives a velocity uncertainty of $\approx 5.8 cm/s$ for an electron due to the low electron mass.

Quantum phenomena can be observed if the motion of the physical object we study is confined to a region of space.¹⁵ Since the electron mass $m_e = 9.1 \cdot 10^{-31} kg$ is very small $\Delta v \approx 5.9 \cdot 10^5 m/s = 1/9 \cdot 10^{-3} c$.

Mathematically the uncertainties are the *standard deviations*¹⁶ of the corresponding

¹⁵Take for example an electron in an atom. The size of the atom is about $\Delta x \approx 10^{-10} m$, $\Delta p = 5.3 \cdot 10^{-25} kg m/s$

¹⁶If we measure the value of some physical quantity (e.g. position) of a particle N times and the measured values are x_1, x_2, \dots, x_N then the average value of x is

$$\langle x \rangle = \frac{1}{N} \sum_{n=1}^N x_n$$

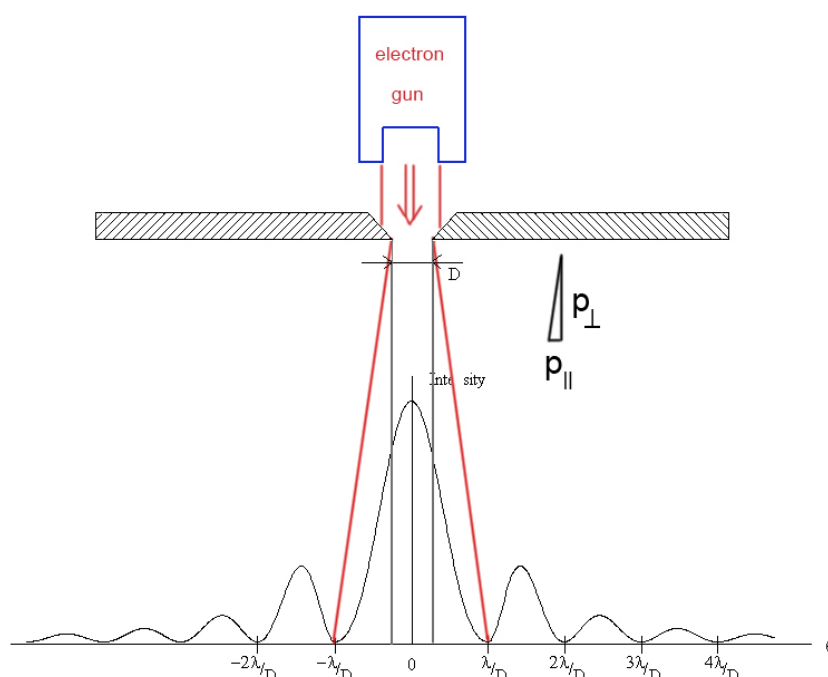
and the square of the standard deviation is

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \frac{1}{N} \sum_{n=1}^N (x_n - \langle x \rangle)^2 (= \langle x^2 \rangle - \langle x \rangle^2)$$

The standard deviation may be non zero even when the average is zero (e.g. velocity distribution in an ideal gas).

quantities. That is a Δp momentum uncertainty says nothing about the *value* of the momentum. The average momentum may be 0, while the uncertainty is never zero. In this case the maximum and minimum values of the momentum itself will lie between $-\Delta p$ and Δp .

Example 3.4. An electron gun emits electrons with a velocity of $v_{\perp} = 1 \text{ m/s}$ perpendicular¹⁷ to a thin metal plate which has a hole of diameter $D = 1 \text{ mm}$ (see figure). Determine the size minimum of the spot on a screen $l = 1 \text{ cm}$ behind the hole.



Solution The electrons arrive at the slit with a velocity and momentum perpendicular to the slit, so the component of their momentum parallel with the slit is $p_{\parallel} = 0$. The slit restricts the diameter of the electron beam to D , therefore right after the slit the position uncertainty of the electrons will be D . This means an uncertainty in the p_{\parallel} momentum of

$$\Delta p_{\parallel} \geq \frac{\hbar}{2D} = 5.27 \cdot 10^{-32} \text{ kg m/s}$$

and a velocity uncertainty of

$$\Delta v_{\parallel} \geq \Delta p_{\parallel} / m_e = 0.058 \text{ m/s}$$

¹⁷This is only an approximation, because an exactly 0 momentum component would require an infinitely large position uncertainty in the parallel direction.

The electrons need $\Delta t = l/v_{\perp}$ time to reach the screen, during which the maximum parallel distance they may travel is $\Delta d = v_{\parallel} \Delta t$. The minimal size of the spot on the screen is therefore:

$$d_{min} = 2 v_{\parallel} l/v_{\perp} + D = 2.2mm$$

Energy-time uncertainty relation

There exists a special kind of uncertainty relation between the uncertainty of the energy and the lifetime of a state:

$$\Delta \mathcal{E} \cdot \Delta t \geq \frac{\hbar}{2} \quad (3.3.3)$$

This relation is special first because, although formally corresponds to (3.3.1), contrary to \mathbf{r} and \mathbf{p} , \mathcal{E} and t are *not* conjugate variables, as t is simply a parameter. For instance if $\Delta \mathcal{E}$ is the uncertainty of energy then the average lifetime of a particle having this energy is $\Delta t = \frac{\hbar}{2 \Delta \mathcal{E}}$. Another example: when a decaying particle is created in a high energy particle collision $\Delta \mathcal{E}$ is its mass energy and Δt is the average lifetime of the particle before it decays.

For stationary electron states in an atom $\Delta \mathcal{E}$ can be exactly 0 which means an infinite lifetime of that state without external interactions.

3.4 The wave function

In classical mechanics the behavior of a particle is described by its position $\mathbf{r}(t)$ and momentum $\mathbf{p}(t)$. These are observable and measurable physical quantities which we will call simply *observables*. Similar observables are for instance the energy \mathcal{E} and the angular momentum L which may depend on \mathbf{r} and \mathbf{p} . Different particles can be identified by results of a set of observables received under special experimental conditions.

According to the Heisenberg uncertainty relations in quantum mechanics neither the position nor the momentum can be determined with an arbitrary precision without influencing the accuracy of the value of the other one. So in quantum mechanics the description must be based on something which takes into account the wave-like nature of the particle.

Important 3.4.1. *Because every physical object has wave-like properties a complex valued function, called the wave function (sometimes written as wavefunction or matter wave) $\psi(x, t)$ is introduced, that describes the state of the physical object.*

We will differentiate between the state itself and its description by a wave function. We will denote the state described by ψ with $|\psi\rangle$.

The wave function is inherently a complex function, therefore it cannot be directly measurable, because results of measurements must be real numbers. The absolute square of the wave function $|\psi(\mathbf{r}, t)|^2$ however is real and proportional to the probability of finding the particle in a $d^3\mathbf{r} \equiv \Delta x \cdot \Delta y \cdot \Delta z (\equiv \Delta V)$ volume around the position \mathbf{r} :

$$\mathcal{P}(\mathbf{r}, d^3\mathbf{r}) = C \cdot |\psi(\mathbf{r}, t)|^2 \cdot d^3\mathbf{r} \quad (3.4.1)$$

where the real number C is the proportionality constant.

The quantity $C \cdot |\psi(\mathbf{r}, t)|^2$ is called the probability density.

C then can be determined from the condition that the probability that the particle is present somewhere in the universe¹⁸ is 1:

$$\int_{\text{whole space}} |C \cdot \psi(\mathbf{r}, t)|^2 \cdot d^3\mathbf{r} = 1 \quad (3.4.2)$$

Using the wave function the double-slit experiment with electrons may be explained very similarly to the one with photons. Only the $\mathbf{E}(\mathbf{r}, t)$ field strength must be replaced by the wave function and the intensity, which for photons is proportional to $\mathbf{E}^2 = (\mathbf{E}_1 + \mathbf{E}_2)^2$, for electrons will be proportional to $|\psi(\mathbf{r}, t)|^2 = |\psi_1(\mathbf{r}, t) + \psi_2(\mathbf{r}, t)|^2$. Therefore the wave function may be called the *probability amplitude*.

3.5 The Schrödinger equation.

We look for the way of determining the wave function that describes the state of a physical object i.e. allows to make predictions on the outcome of measurements of observable physical quantities. Because we know from practice that classical mechanics is applicable to macroscopic objects, which themselves, as we also know, are built from microscopic constituents (atoms, molecules) for which quantum mechanics must be applied, we expect quantum mechanics to be a generalization of Newtonian mechanics. Classical mechanics is based on the Newton equations which themselves are based on experiments. Similarly the base equation of quantum mechanics first introduced by Erwin Schrödinger is based on experimental facts. The Schrödinger equation can be obtained by inductive reasoning,

¹⁸From this it follows that the determination of $\psi(\mathbf{r}, t)$ is not unique. First we may want to incorporate C into $\psi(\mathbf{r}, t)$ to get a *normalized wave function*. And second if we multiply $\psi(\mathbf{r}, t)$ with a complex number $\hat{A} = e^{iA}$, where A is a real number with absolute value of 1 it will not change the probabilities in (3.4.1) or (3.4.2). The set of all ψ wave functions which only differ in a such a multiplier are equivalent:

$$\text{If } \psi'(\mathbf{r}, t) = \hat{A} \cdot \psi(\mathbf{r}, t), \text{ where } |\hat{A}| = 1 \quad \Rightarrow \quad |\psi'(\mathbf{r}, t)|^2 = |\psi(\mathbf{r}, t)|^2$$

detailed below, starting from classical mechanical formulas and our knowledge about the wave-like nature of all particles, but this is not a deductive derivation.

In classical mechanics the velocity of a massive object changes only when external forces are acting on it. In many important cases these forces are *conservative*, i.e. they are the negative gradient of a potential energy:

$$\mathbf{F} \equiv -\text{grad } \mathcal{E}_{\text{pot}}$$

In one dimension

$$F = -\frac{d\mathcal{E}_{\text{pot}}}{dx}$$

In quantum mechanics forces usually do not enter our equations, but are represented by the potential energy.

Important 3.5.1. *A convention in quantum mechanics that the potential energy is called potential and usually denoted by V .*

Never confuse quantum mechanical “potentials” with the potentials of classical physics!

The total energy of a classical particle is the sum of its kinetic and potential energies

$$\mathcal{E}(\equiv \mathcal{E}_{\text{tot}}) = \mathcal{E}_{\text{kin}} + \mathcal{E}_{\text{pot}} = \frac{p^2}{2m} + V(x) \quad (3.5.1)$$

In quantum mechanics the state of the physical object is described by the wave function $\psi(x, t)$. In one dimension using complex exponentials a wave can be described by the formula

$$\psi(x, t) = A e^{-i(\omega t - k x)}$$

According to the de Broglie hypothesis

$$\begin{aligned} \hbar \omega &= \mathcal{E} \\ \hbar k &= p \quad \text{where } k = \frac{2\pi}{\lambda} \end{aligned}$$

this can be written as

$$\psi(x, t) = A e^{i(pr - \mathcal{E}t)/\hbar}$$

Its first order partial derivatives with respect to space and time are:

$$\begin{aligned} \frac{\partial \psi}{\partial x} &= \frac{i}{\hbar} p \psi \quad \text{and} \\ \frac{\partial \psi}{\partial t} &= -\frac{i}{\hbar} \mathcal{E} \psi \end{aligned} \quad (3.5.2)$$

The total energy formula (3.5.1) contains the square of the momentum, so let us calculate the second order partial derivative with respect to space:

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{p^2}{\hbar^2} \psi \quad (3.5.3)$$

Now multiply (3.5.1) with ψ and substitute the values for p^2 and \mathcal{E} from derivatives¹⁹ into (3.5.2) and (3.5.3)

$$\begin{aligned} \mathcal{E} \psi &= \frac{p^2}{2m} \psi + V(x) \psi \\ p^2 \psi &= -\hbar^2 \frac{\partial^2 \psi}{\partial x^2} \\ \mathcal{E} \psi &= i\hbar \frac{\partial \psi}{\partial t} \end{aligned}$$

which results in the equations

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V(x) \psi(x, t) = \mathcal{E} \psi(x, t) \quad (3.5.4)$$

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{E} \psi \quad (3.5.5)$$

(3.5.4) is called the *time independent* or *stationary* Schrödinger equation, because it does not depend on the time and it describes *stationary states*.

The two expressions for $\mathcal{E} \psi$ must be equal. From which we obtain the more general form of the basic equation of quantum mechanics, called the *time dependent Schrödinger equation*:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V(x) \psi(x, t) = i\hbar \frac{\partial \psi(x, t)}{\partial t} \quad (3.5.6)$$

For a classical wave, the wave function should be the solution of the wave equation instead, which in one dimension:

$$\frac{\partial^2 \psi(x, t)}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \psi(x, t)}{\partial t^2}$$

where c would be the velocity of the classical wave.

Important 3.5.2. *The Schrödinger equation is not a classical wave equation, because particles are not classical waves. It is a new concept based on the ideas of de Broglie and of the wave function being a quantity that describes the state of the physical object. It is not a real but a complex equation, therefore the solution (the wave function) will also be inherently complex.*

¹⁹We used the equality $-1/i = i$.

The solution for (3.5.5) is trivial:

$$\mathcal{T}(t) = e^{-i\frac{\mathcal{E}}{\hbar}t} \quad (3.5.7)$$

Therefore

$$\psi(x, t) = \varphi(x) \cdot e^{-i\frac{\mathcal{E}}{\hbar}t} \quad (3.5.8)$$

$\varphi(x)$ is the solution of the time independent Schrödinger equation (3.5.4). However its determination is far from simple. But there are some special idealized problems whose solution is relatively straightforward. These problems involve idealized potentials, which may be used as an approximation of real world potentials.

Both the time dependent and time independent Schrödinger equations are *linear*, which means that if $\psi_1(x, t)$ and $\psi_2(x, t)$ (or $\varphi_1(x)$ and $\varphi_2(x)$) are solutions of the corresponding equations then any linear combination of these

$$\begin{aligned} \psi(x, t) &= C_1 \psi_1(x, t) + C_2 \psi_2(x, t) \\ \varphi(x) &= C_1 \varphi_1(x) + C_2 \varphi_2(x), \end{aligned}$$

where C_1 and C_2 are arbitrary complex numbers, is also a solution of the corresponding equation

It follows that if we write the solution of the time dependent equation at $t = 0$ as a linear combination of the solutions of the corresponding time independent Schrödinger equation

$$\psi(x, 0) = \sum_n C_n \varphi_n(x),$$

then we can calculate it at any $t > 0$ time by

$$\psi(x, t) = \sum_n C_n \varphi_n(x) e^{-i\mathcal{E}_n/\hbar t} \quad (3.5.9)$$

This is one of the reasons why we will use the time independent Schrödinger equation most of the time. The main reason is, of course, that it describes the stationary states of a physical object.

Important 3.5.3. *As we will see the potential may put some restraints on the possible values of \mathcal{E} . In many cases our main goal is to determine the possible \mathcal{E} values. The physically possible \mathcal{E} values are called eigenvalues of the Schrödinger equation. The solutions of the stationary Schrödinger equation for these \mathcal{E} eigenvalues are called eigenfunctions.*

(The prefix *eigen-* is adopted from the German word for "self", because quantum mechanics were first developed by German physicists.)

In classical mechanics the complex solution of the one dimensional wave equation is written as $f(x, t) = e^{i(\omega t - kx)}$ while in quantum mechanics the usual convention changes the sign of the exponent, so the solution of the free electron problem is written as $\psi(x, t) = e^{i(kx - \omega t)}$. The corresponding solution of the time independent equation is $\varphi(x) = e^{ikx}$.

Important 3.5.4. *Although the solution of the time independent Schrödinger equation does not describe a wave in the classical sense of the word, as it does not depend on t , it is still usually called a wave. For instance $\varphi(x) = e^{ikx}$ with positive k is the “wave” that describes an electron moving in the positive x direction. The missing t dependence comes from the exponential factor $e^{-i\mathcal{E}/\hbar t}$.*

In the following sections we will study some simple problems, solve the stationary Schrödinger equation and find out how the potential and boundary conditions may restrict the possible values of the energy.

3.5.1 Free electron in 1 dimensional

For a free electron the solution of the stationary Schrödinger equation is straightforward. When $V = 0$:

$$-\frac{\hbar^2}{2m_e} \frac{d^2 \varphi}{dx^2} = \mathcal{E} \varphi, \quad \text{from which}$$

$$\varphi = C_{\pm} e^{\pm i k x}, \quad \text{where } \mathbf{k} = \frac{\sqrt{2m_e E}}{\hbar}$$

The plus/minus sign describes a particle moving in the positive/negative x direction. The value of the C_{\pm} complex constants is undetermined. This is not a physical wave function, because it describes an electron which is present everywhere in space and has no uncertainty in its momentum. But (an infinite number of) such waves may be used to construct *wave packets*, which may describe real electrons. The value of \mathcal{E} is not restricted to discrete values: *the energy spectrum is continuous*. The phase velocity of this wave

$$v_{ph} = \frac{\omega}{k} = \frac{\mathcal{E}}{\hbar k} = \frac{\hbar k}{2m_e} \quad (3.5.10)$$

depends on the value of k which means that the relative phases of waves with different wave numbers k in a wave packet change over time even in vacuum. This will lead to the spread of the wave packet over time (See Appendix 22.2).

3.5.2 One dimensional potential step

An electron is moving in the following potential (Fig. 3.8:

$$V(x) = \begin{cases} 0 & \text{when } x < 0 \\ V_0 & \text{when } x \geq 0 \end{cases}$$

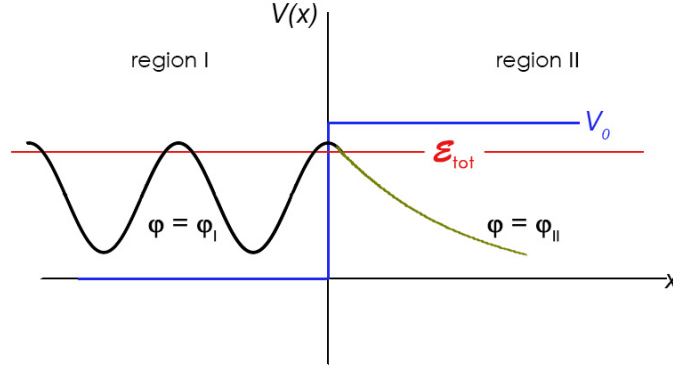


Figure 3.8: Electron in a one dimensional potential step. The wave function is displayed for $\mathcal{E}_{tot} < V_0$

Determine the wave function for an electron moving from the left to the right. The total electron energy \mathcal{E} can be either smaller or larger than V_0 . **Solution** Because the potential divide the space (the x axis in our case) to two distinct parts, both with a constant potential, the solution of the time independent Schrödinger equation in our case is best calculated by solving two equations: one for $x < 0$:

$$-\frac{\hbar^2}{2m_e} \frac{d^2 \varphi_1}{dx^2} = \mathcal{E} \varphi_1$$

and an one for $x \geq 0$:

$$-\frac{\hbar^2}{2m_e} \frac{d^2 \varphi_I}{dx^2} + V_0 \varphi_2 = \mathcal{E} \varphi_I, \quad \text{from which}$$

$$-\frac{\hbar^2}{2m_e} \frac{d^2 \varphi_{II}}{dx^2} = (\mathcal{E} - V_0) \varphi_{II},$$

and connect the two solutions obtained using suitable *boundary conditions*. Here m_e is the electron mass.

Both equations have the same structure:

$$\frac{d^2 \varphi}{dx^2} = -k^2 \varphi(x)$$

with

$$k^2 = \begin{cases} \frac{2m_e \mathcal{E}}{\hbar^2} & x < 0 \\ \frac{2m_e (\mathcal{E} - V_0)}{\hbar^2} & x \geq 0 \end{cases}$$

which have solutions in the form of:

$$\varphi(x) = \text{const} \cdot e^{\pm i k x}$$

The general solution of both equations is the linear combination of these, where we must distinguish between the values of k in the two regions. Instead of using subscripts we will denote k_I with k and k_{II} with q

$$\begin{aligned}\varphi_I(x) &= A \cdot e^{i k x} + B \cdot e^{-i k x} \\ \varphi_{II}(x) &= C \cdot e^{i q x} + D \cdot e^{-i q x}\end{aligned}$$

Here A and C are the amplitudes of the wave traveling in the positive x direction, while B and D are the corresponding amplitudes for waves moving in the opposite direction.

Because an electron cannot be divided to two “half-electrons”, the *wave function must be continuous*. This is the first *boundary condition*.

And because the resulting wave function must be the solution of a *single* equation containing a second derivative for the whole space, the second boundary condition is the *continuity of the first derivative of the wave function* at $x = 0$:

$$\begin{aligned}\varphi_I(0) &= \varphi_{II}(0) \\ \left. \frac{d\varphi_I}{dx} \right|_{x=0} &= \left. \frac{d\varphi_{II}}{dx} \right|_{x=0}\end{aligned}$$

Our electron originally arrives to the $x = 0$ boundary from the left traveling in the positive x direction. This means that the amplitude $A \neq 0$ for the wave traveling right in region 'I'. Part of the wave function may be reflected back from the potential step into region 'I' ($B \neq 0$) and part of it may enter the region of the higher potential ($C \neq 0$). But there will be no part traveling backwards there, therefore $D = 0$.

Substituting the wave functions into the boundary conditions we get the following equalities:

$$\begin{aligned}A + B &= C \\ i k (A - B) &= i q C\end{aligned}$$

i.e. 2 equations for the 3 unknowns. This means that we can set the value of one of the unknown parameters arbitrarily and determine the others depending on its value²⁰. In our case let us select the value $A = 1$. With this selection

$$B = \frac{k - q}{k + q} = \frac{\sqrt{E} - \sqrt{\mathcal{E} - V_0}}{\sqrt{E} + \sqrt{E - V_0}}$$

$$C = \frac{2k}{k + q} = \frac{2\sqrt{E}}{\sqrt{\mathcal{E}} + \sqrt{E - V_0}}$$

The wave function is then

$$\varphi(x) = \begin{cases} e^{ikx} + \frac{k-q}{k+q} e^{-ikx} & x < 0 \\ \frac{2k}{k+q} \cdot e^{iqx} & x \geq 0 \end{cases}$$

Up to this point we did not distinguish between the two cases when $\mathcal{E} > V_0$ and when $\mathcal{E} < V_0$.

When $\mathcal{E} > V_0$, i.e. the kinetic energy of the incoming particle is larger than the potential step both k and q are real. The part of the wave function which enters region 'II' is

$$\varphi_2(x) \equiv C \cdot e^{-iqx}$$

which is a wave traveling in the positive x direction with constant amplitude and constant $|\varphi_{II}|^2 = |C|^2$ *probability density*.

If the electron was a classical particle, whose movement is governed by the laws of classical mechanics it could never turn around, it would always move in the positive x direction. In quantum mechanics however there is always some possibility that the electron is reflected back from the boundary, because if $V_0 > 0$ then $B \neq 0$.

When $\mathcal{E} < V_0$, i.e. the kinetic energy of the incoming particle is larger than the potential step then k is still real but q is imaginary: $q = i \frac{\sqrt{2m_e(V_0 - E)}}{\hbar}$, this makes B and C complex. Part of the wave function still reaches into region 'II' but it is an exponentially decreasing function:

$$\varphi_{II}(x) = C \cdot e^{i \cdot i |q| x} = C \cdot e^{-|q| x}$$

In classical physics if the total energy of the particle is smaller than the potential energy in some region of space the particle is

²⁰Because the wave arriving from the left is infinite it can not be normalized.

always reflected back from the boundary and cannot enter the region where $\mathcal{E} < V_0$. This reflection always occurs in quantum mechanics too, but the wave function will not be 0 inside region 'II'. The probability density decreases exponentially from the boundary: $|\varphi_2(x)|^2 = |C|^2 e^{-|q|x}$. The *penetration depth* δ_P is the distance where the probability density falls to $1/e$ (about 37%) of its value at the boundary, i.e.

$$\delta_P = \frac{\hbar}{\sqrt{2m_e(V_0 - \mathcal{E})}}$$

I.e. the higher is the potential the smaller is the penetration depth.

We learned from this calculation that

Important 3.5.5. *The solution of the Schrödinger equation (the wave function) must be a finite valued, continuous and continuously differentiable function. This is true even for $V(x)$ potentials that have an abrupt, but finite jump.*

The potential we used in this example is of course an idealization. In realistic cases (see the next section) the potential raises from 0 to a constant value not abruptly but in some $\Delta x = s$ distance. In classical mechanics you may imagine a ramp connecting the ground with a raised platform. The potential energy of a classical object sliding or rolling on this ramp has a potential energy of this shape. In quantum mechanics this potential can be realized by a homogeneous electric field connecting two halves of space with different potential.

But do not forget that the particle is moving along the x axis and not along the potential curve!

3.5.3 Potential box in 1 dimension

Let us suppose that the potential is 0 inside a finite region of length L in 1 dimensional space while outside this region it is infinitely large. This potential arrangement is called a *potential box* (see Fig. 3.9). Put an electron into this box and determine the possible energy levels. **Solution According to the description we may select the following arrangement:**

$$V(x) = \begin{cases} \infty & x \leq 0 \\ 0 & 0 < x < L \\ \infty & x \geq L \end{cases}$$

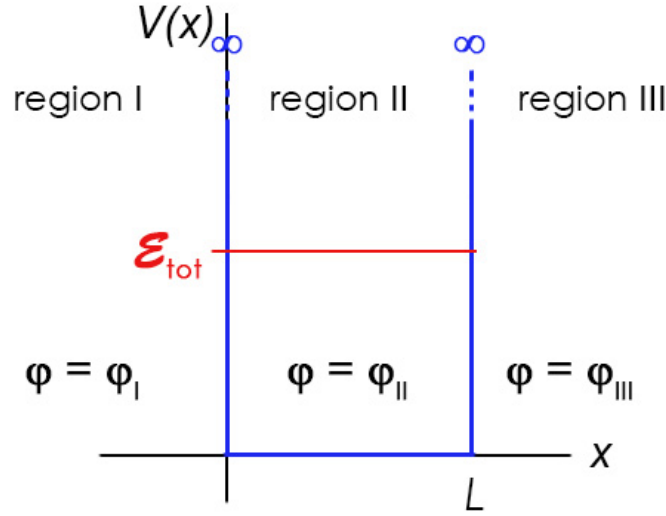


Figure 3.9: Electron in a 1 dimensional potential box. Outside region II the potential is infinite.

We present two solutions for this problem. The first one uses the exponential notation, the second one jumps straight to harmonic functions.

a)

Inside the box the electron can move freely, but it cannot leave the box. Because the potential jumps to infinity at the boundaries the penetration depth will be exactly 0. Therefore

$$\psi(x, t) = \varphi(x) e^{-i \frac{\varepsilon}{\hbar} t}, \text{ where}$$

$$\varphi(x) = \begin{cases} 0 & x \leq 0 \\ A e^{i k x} + B e^{-i k x} & 0 < x < L \\ 0 & x \geq L \end{cases}$$

In this case there are only two unknown constants A and B in the function and the boundary conditions at $x = 0$ and $x = L$ are enough. Because the wave function must be continuous these conditions are:

$$A + B = 0, \quad A e^{i k L} + B e^{-i k L} = 0$$

From these:

$$A (e^{i k L} - e^{-i k L}) = 2 i A \sin k L = 0 \quad \Rightarrow \quad \sin k L = 0$$

$$\begin{aligned}
k L &= n \pi & n &= 1, 2, 3... \\
k &= \frac{n \pi}{L} & n &= 1, 2, 3...
\end{aligned}$$

The wave numbers, the momentum and energy of the electron are not continuous variables but can only take discrete values. The value $n = 0$ would belong to a zero wave function $\psi(x, t) = 0$, which “describes” an empty box without any electron inside it.

b)

Because the solution of the Schrödinger equation is a linear combination of exponentials with imaginary power and harmonic functions can be expressed as sums or difference of these we may use a linear combination of sine or cosine functions instead of exponential functions. According to the boundary conditions the selected function must be 0 at both boundaries. This means that we may use a single sine function and that the integer multiple of the half wavelength must be equal to L with the condition that:

$$\begin{aligned}
\psi(x, t) &= \varphi(x) e^{-i \frac{\mathcal{E}}{\hbar} t} \\
\varphi(x) &= A \sin \frac{x}{\lambda} \\
n \frac{\lambda}{2} &= L & n &= 1, 2, 3... \\
\lambda_n &= \frac{2 L}{n} \\
k_n &= \frac{2 \pi}{\lambda} = \frac{\pi}{L} n & n &= 1, 2, 3...
\end{aligned}$$

Both ways we arrive to the same formulas (here we explicitly note that all quantities depends on the integer number n):

$$\varphi_n(x) = A_n \sin k_n x, \text{ where } k_n = \frac{n \pi}{L} \quad (3.5.11a)$$

$$p_n = \hbar k_n = \frac{\hbar \pi}{L} n \quad n = 1, 2, 3... \quad (3.5.11b)$$

$$\mathcal{E}_n = \frac{p_n^2}{2 m_e} = \frac{\pi^2 \hbar^2}{2 m_e L^2} n^2 \quad n = 1, 2, 3... \quad (3.5.11c)$$

$$\begin{aligned}
\text{or } \mathcal{E}_n &= n^2 \mathcal{E}_1, \quad \text{where} \\
\mathcal{E}_1 &= \frac{\pi^2 \hbar^2}{2 m_e L^2} = \frac{h^2}{8 m_e L^2} \quad (3.5.11d)
\end{aligned}$$

Notice that the energy expressed with k is of the same form as it

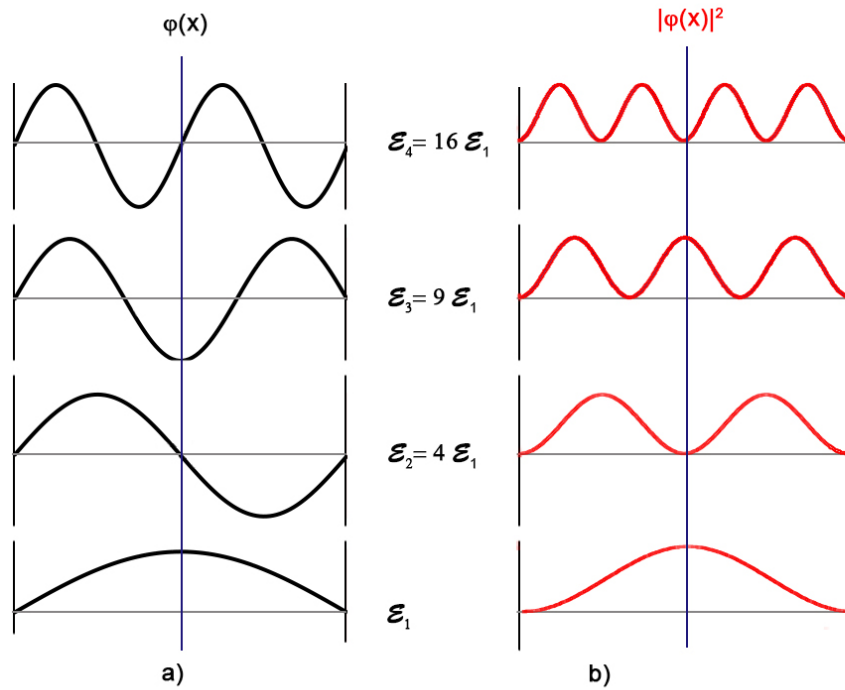


Figure 3.10: The wave functions (left) and the probability densities (right) for a 1 dimensional potential box.

would be for a free particle, the difference being that for a particle confined in a box the energy spectrum is discrete, only discrete energy levels are allowed.

The wave function and the probability density function for the 4 lowest lying state is shown in Fig 3.10.

Important 3.5.6. *The lowest lying energy state (here \mathcal{E}_1) is called the ground state, sometimes referred to as the zero point energy of the physical object.*

The value of the A constant can be determined from the *normalization of the wave function*. The probability that the electron is

found inside the box must be 1:

$$\int_{-\infty}^{\infty} |\varphi(x)|^2 dx = \int_0^L |\varphi(x)|^2 dx = 1$$

$$\int_0^L |A \sin \frac{n\pi}{L} x|^2 dx = |A|^2 \frac{L}{2} = 1$$

$$|A| = \sqrt{\frac{2}{L}}$$

$$\varphi(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi}{L} x$$

The square of the absolute value of the $\varphi(x)$ wave function in the integral is

$$|\varphi(x)|^2 \equiv \varphi^* \varphi$$

We will be interested in integrals containing *different* wave functions. For this purpose it is useful to know that if $\varphi_n(x)$ and $\varphi_m(x)$ are two eigenfunctions of the Schrödinger equation then

$$\int_{-\infty}^{\infty} \varphi_n(x)^* \varphi_m(x) dx = \delta_{n,m}$$

where $\delta_{n,m}$ is the **Kronecker's delta**²¹:

$$\delta_{n,m} = \begin{cases} 0, & \text{if } m \neq n \\ 1, & \text{if } m = n \end{cases} \quad (3.5.12)$$

The calculation is simple:

$$\int_{-\infty}^{\infty} \varphi_n(x)^* \varphi_m(x) dx =$$

$$\int_0^L \frac{2}{L} \left(\sin \frac{n\pi}{L} x \cdot \sin \frac{m\pi}{L} x \right) dx =$$

$$\frac{1}{L} \int_0^L \left(\cos \frac{(m-n)\pi x}{L} - \cos \frac{(m+n)\pi x}{L} \right) dx = 0$$

²¹Named after the 19th century German mathematician Leopold Kronecker who worked on number theory and algebra.

Integrals of different wave functions are so important in quantum physics, that they have a special notation

$$\langle \phi | \varphi \rangle := \int_{-\infty}^{\infty} \phi(x)^* \varphi(x) dx \quad (3.5.13)$$

which is called the *scalar product* of the wave functions $\phi(x)$ and $\varphi(x)$ ²². This result is generally true:

Important 3.5.7. *The scalar product of two different eigenfunctions of a given Schrödinger equation is always 0. We say that two different eigenfunctions are always “orthogonal” to each other. The scalar product of a wave function with itself must be positive and finite, because it is proportional to the probability of finding the particle somewhere in space. In other words the wave function must be square-integrable a.k.a. quadratically integrable.*

The ground state energy of a 1 dimensional potential box is in agreement with the Heisenberg uncertainty relation too if we take the constant on the right hand side to be h instead of the exact $\frac{\hbar}{2}$ value:

$$\Delta x \cdot \Delta p \geq h$$

In our case

$$\Delta x = L \quad \Rightarrow \quad \Delta p \geq \frac{h}{L}$$

If we now consider Δp as the change of the momentum of the particle when it collides with one of the walls, then

$$\begin{aligned} \Delta p = 2 \cdot p \quad \Rightarrow \quad \mathcal{E} &= \frac{p^2}{2m_e} = \frac{\Delta p^2}{8m_e} \\ \mathcal{E} &\geq \frac{h^2}{8m_e L^2} = E_1 \end{aligned}$$

Example 3.5. *What is the wavelength of the photon emitted by an electron transition from the 4th to the 3rd level in a 1 dimensional potential box of size 100 nm?* **Solution** From (3.5.11c) the energy difference between level 4 and 3 is

$$\begin{aligned} \Delta \mathcal{E} = \mathcal{E}_4 - \mathcal{E}_3 &= \frac{\hbar \pi^2}{2m_e L^2} (4^2 - 3^2) = \frac{7 \cdot \pi^2}{2 \cdot 9.1 \cdot 10^{-31} (10^{-7})^2} \\ &= 6.02 \cdot 10^{-24} \text{ J} \end{aligned}$$

²²This name reflects the similarity between the mathematical properties of these integrals and that of vectors in *linear algebra*, which we will discuss in some details in Chapter 5.

and the photon frequency is

$$\nu = \frac{\Delta \mathcal{E}}{h} = \frac{6.02 \cdot 10^{-24}}{6.62 \cdot 10^{-34}} = 9.09 \cdot 10^9 \text{ Hz}$$

The wavelength of the emitted photon then

$$\lambda = \frac{c}{\nu} = 3.3 \text{ cm}.$$

3.5.4 Potential box in 3 dimensions

A 3D potential box is a generalization of a 1 dimensional potential box. It describes the situation of a particle confined in space. It is a model e.g. for an electron in a metal, where it can move around freely, however it can not leave it. Let the sides of the box be equal to L_x , L_y , L_z and let us put the origin of the coordinate system into one of the corner of a box as in Fig 3.11 The potential:

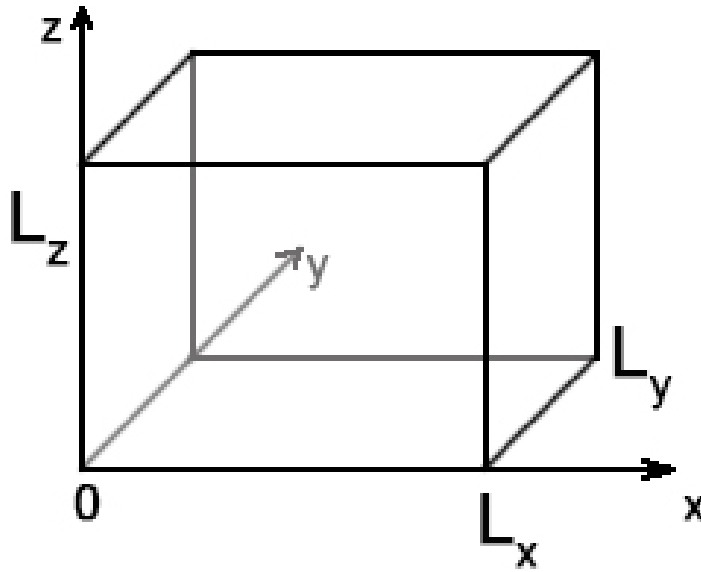


Figure 3.11: The coordinate system for the 3D potential box. The potential is zero inside and ∞ high outside of the box.

$$V(x, y, z) = \begin{cases} 0 & 0 < x < L_x, \\ & \text{when } 0 < y < L_y \text{ and} \\ & 0 < z < L_z \\ \infty & \text{otherwise} \end{cases}$$

The stationary Schrödinger equation in 3 dimension is:

$$-\frac{\hbar^2}{2m_e} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \varphi(x, y, z) + V(x, y, z) = \mathcal{E} \varphi(x, y, z)$$

In this case it is easy to prove by substitution that the wave function can be written as a product of three independent sine functions along the x,y and z axes, where all three functions must vanish (i.e. $\varphi = 0$) at the corresponding walls of the box:

$$\varphi(x, y, z) = A \cdot \sin k_x x \cdot \sin k_y y \cdot \sin k_z z$$

where the three *wave numbers* must be

$$k_x = \frac{2\pi}{\lambda_x} = \frac{\pi}{L_x} n_x \quad n_x = 1, 2, 3... \quad (3.5.14a)$$

$$k_y = \frac{2\pi}{\lambda_y} = \frac{\pi}{L_y} n_y \quad n_y = 1, 2, 3... \quad (3.5.14b)$$

$$k_z = \frac{2\pi}{\lambda_z} = \frac{\pi}{L_z} n_z \quad n_z = 1, 2, 3... \quad (3.5.14c)$$

The three wave numbers form the *wave vector* $\mathbf{k} \equiv (k_x, k_y, k_z)$. The total energy then becomes:

$$\mathcal{E}_{n_x, n_y, n_z} = \frac{\hbar^2 k^2}{2m_e} = \frac{\hbar^2 \pi^2}{2m_e} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right) \quad (3.5.15)$$

For a cubic box the three sides are of the same length L and

$$\mathcal{E}_{n_x, n_y, n_z} = \mathcal{E}_1 (n_x^2 + n_y^2 + n_z^2), \text{ where} \quad (3.5.16a)$$

$$\mathcal{E}_1 = \frac{\hbar^2 \pi^2}{2m_e L^2} = \frac{h^2}{8m_e L^2} \quad (3.5.16b)$$

The state of the physical object is characterized by the 3 *quantum numbers* n_x, n_y and n_z . Different combinations of n, m and l may give the same energy value. The simplest way to show this is by an example.

Example 3.6. Determine the first 3 energy levels in a cubic potential box whose size is $a = 10 \mu m$. **Solution** Substituting $L = L_x = L_y = L_z = 10 \mu m$ into (3.5.15) we get

$$\mathcal{E}_{n_x, n_y, n_z} = \frac{\hbar^2 \pi^2}{2 m_e L^2} (n_x^2 + n_y^2 + n_z^2) \quad n_x, n_y, n_z = 1, 2, 3, \dots$$

$$\mathcal{E}_{n_x, n_y, n_z} = 6,02 \cdot 10^{-26} (n_x^2 + n_y^2 + n_z^2) \quad n_x, n_y, n_z = 1, 2, 3, \dots$$

Because the result depends on the sum of the squares of the three numbers, the same energy values will result for all permutations of the same three numbers:

n	m	l	\mathcal{E} ($\mathcal{E}_1 := 1.81 \cdot 10^{-27} J$)	\mathcal{E} $\times 10^{-27} J$
1	1	1	$3 \cdot \mathcal{E}_1$	1.807
1	1	2	$6 \cdot \mathcal{E}_1$	3.61
1	2	1		
2	1	1		
2	2	1	$9 \cdot \mathcal{E}_1$	5.42
2	1	2		
1	2	2		
2	2	2	12	7.23
1	1	3	$11 \cdot \mathcal{E}_1$	6.63
1	3	1		
3	1	1		

Important 3.5.8. The set of states that have the same energy are called degenerate states. The number of different states with the same energy are called the degeneracy of the energy level. The scalar products of two different eigenfunctions are still 0 even when the two eigenfunctions belong to the same degenerate energy levels.

The result in the previous example can be presented in a way that emphasizes the degeneracies

factor	states	degeneracy
3	(1,1,1)	1
6	(2,1,1), (1,2,1), (1,1,2)	3
9	(2,2,1), (2,1,2), (1,2,2)	3
11	(3,1,1), (1,3,1), (1,1,13)	3
12	(2,2,2)	1
14	(1,2,3), (3,2,1), (2,3,1), (1,3,2), (2,1,3), (3,1,2)	6

The number of the energy levels for a potential box is unlimited, so this table could be continued indefinitely.

Example 3.7. *An electron is confined in a 3D potential box with sides $10\mu m$, $20\mu m$ and $30\mu m$. Give the energy and degeneracy of the 4 lowest lying states.* **Solution** The possible energy levels are

$$\mathcal{E}_{n_x, n_y, n_z} = \frac{\hbar^2 \pi^2}{2 m_e L^2} (n_x^2 + n_y^2 + n_z^2) \quad n_x, n_y, n_z = 1, 2, 3, \dots$$

n	m	l	$\mathcal{E}(\times 10^{-27} J)$
1	1	1	1.36
1	1	2	1.47
1	2	1	1.61
2	1	1	2.36
2	2	1	2.61
2	1	2	2.47
1	2	2	1.72
2	2	2	2.72
1	1	3	1.58
1	3	1	1.86
3	1	1	3.36
1	2	3	1.83
2	1	3	2.58
2	3	1	2.86
3	2	1	3.61
3	1	2	3.47
2	2	3	2.83
2	3	2	2.97
3	2	2	3.72
1	3	3	2.08
3	1	3	3.58
3	3	1	3.86
2	3	3	3.08
3	2	3	3.83
3	3	2	3.97
3	3	3	4.08

$$\begin{aligned} \mathcal{E}_{n_x, n_y, n_z} &= \frac{h^2}{8 m_e} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right) \\ &= 6.02 \cdot 10^{-28} \left(\frac{n_x^2}{1} + \frac{n_y^2}{4} + \frac{n_z^2}{9} \right) [J] \end{aligned}$$

The 4 lowest lying energy states can be determined by trying out different combinations of the numbers 1,2 and 3 and selecting the ones with the 4 smallest energy values.

From the table we can see that there are no degenerate states for this physical object and the indices for the 4 lowest lying levels sorted by energy in ascending order are: (1,1,1), (1,1,2), (1,1,3) and (1,2,1).

3.5.5 Density of states

To answer such questions as “How many energy levels are inside a $\Delta \mathcal{E}$ interval around a given \mathcal{E} ?” we have to enumerate all possible combinations of the 3 *quantum numbers* n_x, n_y and n_z to determine how many levels fall inside $\Delta \mathcal{E}$. This is tedious, but luckily there is an easier way.

From (3.5.14) we see that the difference between two consecutive k values

$$\Delta k_i = \frac{\pi}{L_i} \quad i = x, y \text{ and } z \quad (3.5.17)$$

is inversely proportional to the size of the box in that direction. It follows that if the size of the box is sufficiently large then consecutive k values are close to each other. Because the energy is dependent on k , in a sufficiently large box consecutive energy levels are also close to each other. Even for a cubic box of size $10 \mu m$ (see Problem 3.6) this energy difference is only about $10^{-27} J$ or about $10^{-8} eV$. To simplify our task we introduce a (*quasi*) *continuous* function $g(k)$ called the *density of states function* (per unit k) with the following definition:

$$\Delta \mathcal{N}(k, \Delta k) = g(k) \cdot \Delta k = \frac{d\mathcal{N}(k)}{dk} \cdot \Delta k$$

where $\mathcal{N}(k)$ is the number of all states with wave numbers smaller than or equal to k . Many physical quantities depend on \mathbf{k} . Knowing the density of state per unit k makes it possible to calculate the density of states for them as well. In our case this other physical quantity is the energy. The functional dependence of \mathcal{E} on \mathbf{k} is called the (*energy*) *dispersion relation*. For an electron in a 3 dimensional cubic potential box (see (3.5.16)):

$$\mathcal{E}(k) = \frac{\hbar^2 k^2}{2m_e} \quad \Rightarrow \quad k(\mathcal{E}) = \frac{\sqrt{2m_e \mathcal{E}}}{\hbar} \quad \text{and} \quad \frac{dk}{d\mathcal{E}} = \frac{\sqrt{2m_e}}{2\hbar\sqrt{\mathcal{E}}} \quad (3.5.18)$$

Let us determine density of state per unit k $g(k)$ and the density of state per unit energy $g(\mathcal{E})$ for a cubic box!

Imagine a Cartesian coordinate system where the values of n, m and l are measured on the three axes. Unless k is small the number of the possible states with wave vectors less than or equal to k is very large (e.g 1 million or larger). We want to count the states for which k is at most

$$k^2 \leq \frac{\pi^2}{L^2} (n_x^2 + n_y^2 + n_z^2)$$

(see (3.5.14)). If the 3 numbers could take on any values not just integers, this would occupy a sphere with a radius of

$$\xi := \sqrt{n_x^2 + n_y^2 + n_z^2} = \sqrt{\frac{k L}{\pi}}$$

Because the 3 quantum numbers are all positive we must only consider the part of this sphere inside the first octant of space. The volume of this eighth sphere is the (possibly fractional) number of the unit volumes inside it and every (n_x, n_y, n_z) triad determines such a unit volume, therefore the volume is equal to the total number of states $\mathcal{N}(k)$ with wave vectors whose length is shorter than or equal to k . Part of some of these volumes will be outside the sphere and this introduces some error in the calculations. However if the three numbers are large enough the error we make can be negligibly small compared to the volume itself. Therefore

$$\mathcal{N}(k) = \frac{1}{8} \cdot \frac{4\pi\xi^3}{3} = \frac{1}{8} \cdot \frac{4\pi}{3} \left(\frac{kL}{\pi} \right)^3$$

The factor $V = L^3$ is the volume of the potential box. The density of states per unit k therefore is

$$g(k) = \frac{d\mathcal{N}}{dk} = \frac{V}{2\pi^2} k^2 \quad (3.5.19)$$

From which the *density of states per unit energy* (or simply *density of states*) is

$$g(\mathcal{E}) \equiv \frac{d\mathcal{N}(\mathcal{E})}{d\mathcal{E}} = \frac{d\mathcal{N}(k(\mathcal{E}))}{dk} \cdot \frac{dk}{d\mathcal{E}} = g(k(\mathcal{E})) \frac{dk}{d\mathcal{E}}$$

From (3.5.18)²³

$$\begin{aligned} g(\mathcal{E}) &= \frac{V}{2\pi^2} k^2 \frac{\sqrt{2m_e}}{2\hbar\sqrt{\mathcal{E}}} = \frac{V}{2\pi^2} \frac{2m_e\mathcal{E}}{\hbar} \frac{\sqrt{2m_e}}{\hbar\sqrt{\mathcal{E}}} \\ &= \frac{4\pi V\sqrt{2m^3}}{h^3} \sqrt{\mathcal{E}} \end{aligned} \quad (3.5.20)$$

²³Remember that $\hbar \equiv h/2\pi$!

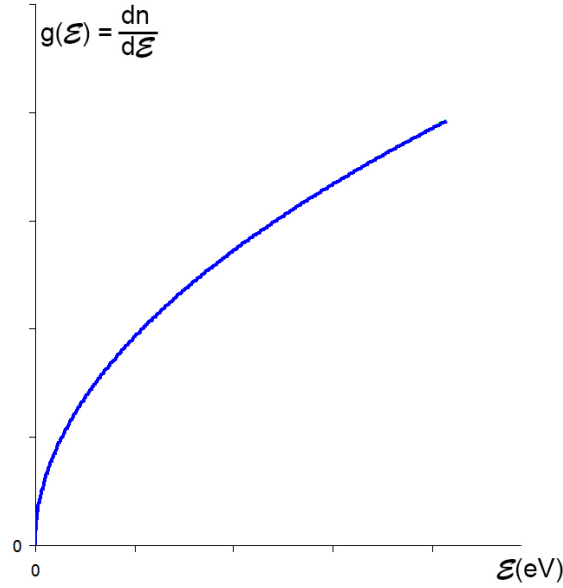


Figure 3.12: Density of state function in a 3 dimensional potential box.

3.5.6 Linear harmonic oscillator.

In classical mechanics the simplest model of a linear harmonic oscillator was a mass point on a spring. The potential energy of such a physical object is quadratic in the excursion of the point from the origin:

$$\mathcal{E}_{pot} = \frac{D}{2} x^2 = \frac{1}{2} m \omega^2 x^2, \quad (3.5.21a)$$

where D is the spring constant and

$$\omega = \sqrt{\frac{D}{m}} \quad (3.5.21b)$$

is the proper frequency of the harmonic oscillator. A similar potential describes vibration of atoms or molecules. The one dimensional Schrödinger equation for the harmonic potential

$$-\frac{\hbar^2}{2m} \frac{d^2 \varphi}{dx^2} + \frac{1}{2} m \omega^2 x^2 \varphi = \mathcal{E} \varphi \quad (3.5.22)$$

The solutions are in Appendices 22.5 and 22.6. We find again the possible energy values discrete:

$$\mathcal{E}_n = \left(n + \frac{1}{2}\right) \hbar \omega \quad n = 0, 1, 2, \dots \quad (3.5.23)$$

In this case the energy levels are equidistant, and the ΔE difference of consecutive levels is

$$E_{n+1} - E_n = \hbar \omega = h \nu \quad (3.5.24)$$

That is the energy of the linear harmonic oscillator may only change by an integer multiple of the energy quantum $h \nu$.

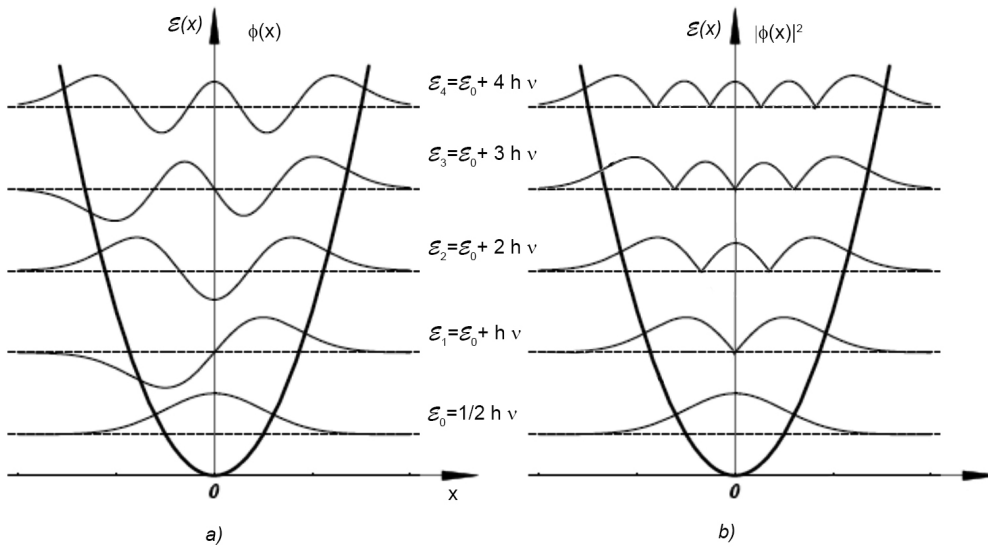


Figure 3.13: Wave functions a) and probability densities b) for a linear harmonic oscillator. Note that both overreach the limits of the classical motion and decay exponentially fast in the walls.

In Fig. 3.13 both the wave function and the probability densities are shown. You can see that even indices belong to even, odd indices to odd wave functions.

The fact that the energy quantum for the linear harmonic oscillator is the same as for photons is not a coincidence. The reason behind it is that both the vibration of the atoms in a solid (see Chapter 13) and of the electromagnetic field itself can be modeled by independent linear harmonic oscillators.

The n constant is the number of energy quanta in the oscillator. The zero point energy of this oscillator is

$$E_0 = \frac{1}{2} \hbar \omega = \frac{1}{2} h \nu$$

3.5.7 One dimensional square potential well

The potential shown in Fig. 3.14 may be used for example as a crude approximation of the quadratic potential of the linear harmonic oscillator or the Coulomb potential:

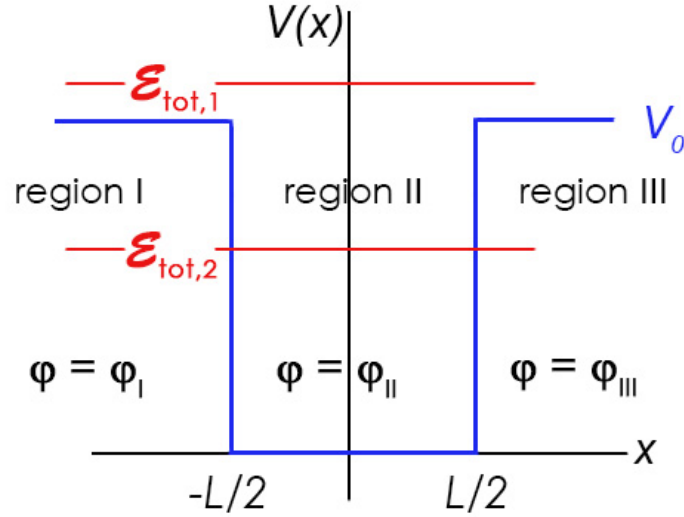


Figure 3.14: 1 dimensional potential well

$$V(x) = \begin{cases} V_0, & |x| \geq \frac{L}{2} \\ 0, & |x| < \frac{L}{2} \end{cases} \quad (3.5.25)$$

Again we can solve the Schrödinger equation stepwise and use boundary conditions to connect the pieces. But despite its simplicity this is a potential for which no analytical solution exists. We may either use graphical or numerical methods to determine the energy values. The reason we are talking about this problem at all is twofold. First we wanted to show that sometimes the solution to even simple quantum mechanical problems may be complicated and second, that it allows us to draw some conclusions about the quantum mechanical energy spectra of the physical objects. The whole calculation is in Appendix 22.7 here we only show the results.

As previously we must distinguish between two cases when the total energy of the particle is larger than V_0 or smaller than V_0 .

When the total energy of the particle is larger than V_0 then the particle is can move freely (not bound) and there are no constraints for the possible energy values of these *unbound* states.

Important 3.5.9. *The energy spectrum of unbound states is continuous i.e. not quantized.*

This result is also valid for instance for electrons, atoms and molecules when the total energy is larger than the potential energy anywhere in the whole space. Usually the zero of the energy scale is selected so that it marks the boundary between bound and unbound states²⁴.

In our case let us modify the potential using this convention:

$$V(x) = \begin{cases} 0, & |x| \geq \frac{L}{2} \\ -V_0, & |x| < \frac{L}{2} \end{cases} \quad (3.5.26)$$

When $\mathcal{E} < 0$, then in classical physics the particle is confined to the inside of the potential well, however the wave function of a quantum particle extend into the walls as it is shown in Fig. 3.15. The number of the possible energy

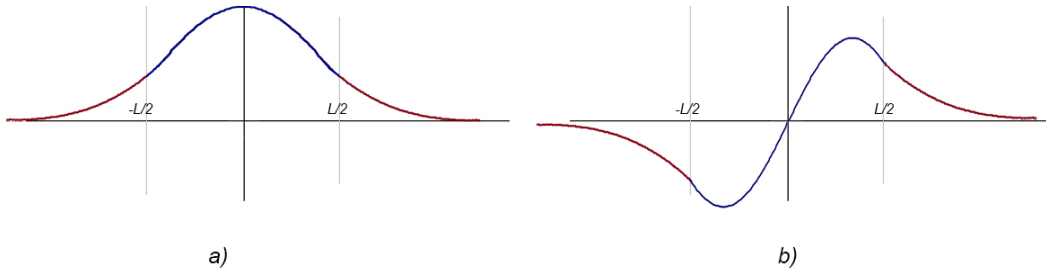


Figure 3.15: Possible wave functions in a potential well: a) even function, b) odd function. The blue line is the wave function inside the well (φ_{II} in Fig. 3.14), the red line is the part in the walls (φ_I and φ_{III})

levels for this potential well is *finite*.

²⁴For example the zero energy of a Coulomb potential of an atom is set at an infinite distance from the atom.

3.6 Central potentials

Real potentials have no jumps in their values. According to classical physics the potential between two oppositely charged ions is the attractive Coulomb potential. However if this was the only force acting on them then they would collide and stick tightly together instead of staying at a distance as in solids. We will see the physical reasons later on why these ions are held apart. Here we only show a potential which can be used to describe this behavior²⁵.

The potential which an ion is exposed to looks like the one in Fig. 3.16.

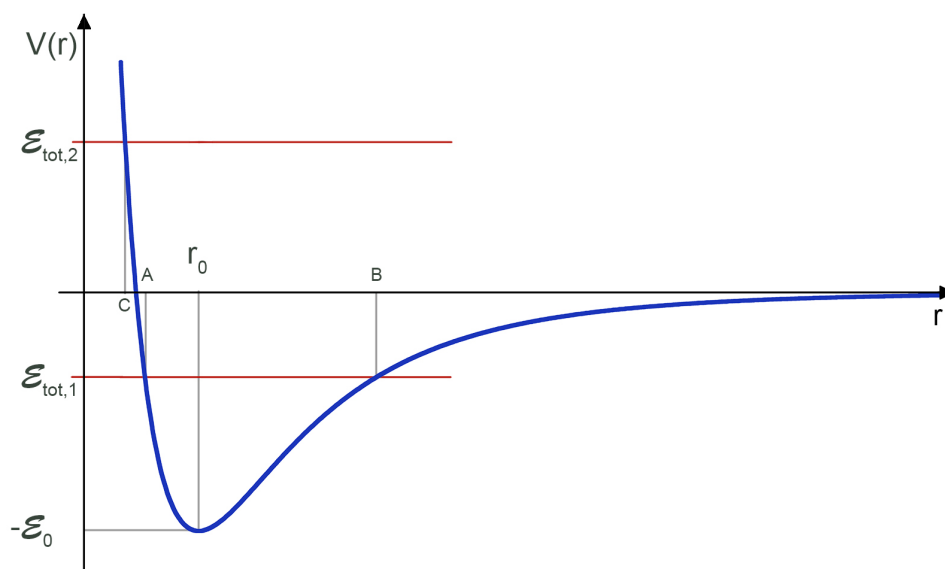


Figure 3.16: Schematic potential of a central force. The points labeled "A", "C" are the minimum possible distances between the ions and "B" is the maximum distance in a bound state. These are the classical turning points, r_0 is the equilibrium position.

In classical physics when the total energy $\mathcal{E}_{tot} < 0$ the movement of the particle is confined between points "A" and "B", while for $\mathcal{E}_{tot} > 0$ the particle can move between the point "C" and infinity. The kinetic energy of the particle $\mathcal{E}_{kin}(x) = \mathcal{E}_{tot} - V(x)$ becomes 0 in points "A", "B" and "C". These are the classical turning points for the particle. Therefore the particle is bound for $\mathcal{E}_{tot,1} < 0$ and its movement is oscillatory (generally it is not harmonic oscillation though). and unbound for $\mathcal{E}_{tot,2} > 0$. The actual value of both $\mathcal{E}_{tot,1}$ and $\mathcal{E}_{tot,2}$ is unrestricted: the energy spectrum is continuous in both cases.

²⁵The potential displayed in Fig. 3.16 is called the *Lenard-Jones (6-12) potential* and it has the formula: $V(x) = \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right]$.

In *quantum mechanics* the energy spectrum of the unbound state ($\mathcal{E}_{tot} > 0$) is continuous too, but the bound states may only have quantized energy levels, i.e. $\mathcal{E}_{tot} < 0$ has a discrete spectrum. The wave function will extend into the potential walls to regions where a classical particle may not go and vanish exponentially there. Near the equilibrium position r_0 the potential may be approximated by a quadratic one and the vibrations of the ions are approximately harmonic. This is one of the reasons why studying the linear harmonic oscillator is important.

If the particle is in a bound state with energy $\mathcal{E}_{tot} < 0$ then the *binding energy* is $\mathcal{E}_b = -\mathcal{E}_{tot}$. If this is a potential between only two ions, then this is also the *dissociation energy*.

3.7 The potential barrier, tunnel effect

As we saw in the previous examples the wave function may extend into regions where a classical particle could not go. This part of the wave function decreases exponentially. But what happens if the width of the region where the potential is larger than the total energy of the particle is so thin that the wave function is not negligibly small at the other side of this *potential barrier*? Can the particle cross the potential barrier? We know a classical particle cannot.

This question is not academic and the answer is a definite “yes”. The effect is called *quantum tunneling*, because a classical object with insufficient kinetic energy to climb a hill (i.e. a potential barrier) could only go through it in a tunnel.

Tunneling describes - among others - the emission of electrons from a conductor under the influence of an external electric field (*field emission*, see below). There are devices, for instance the *Scanning Tunneling Microscope* (STM) and its cousin the *Atomic Force Microscope* (ATM or AFM), whose operation is based on this effect, have about 1000 times the resolution of the optical microscope. But tunneling is also a source of current leakage in very-large-scale integration (VLSI) electronics and results in the substantial power drain and heating effects that plague high-speed and mobile technology. It influences how small computer chips can be made.

Let us model the situation with a single incident electron traveling in the positive x direction arriving at a square potential “wall” or “barrier” of width a like the one in Fig. 3.17. The stationary wave functions of the incident, reflected and transmitted electron (or wave) are a linear combination of complex exponential functions (or sine and cosine waves), while inside the wall ($\mathcal{E}_{tot} < V_0$) a linear combination of real exponentials. The wave functions in the three regions are:

$$\varphi_I = A e^{i k x} + B e^{-i k x} \quad k = \frac{\sqrt{2 m \mathcal{E}}}{\hbar} \quad (3.7.1a)$$

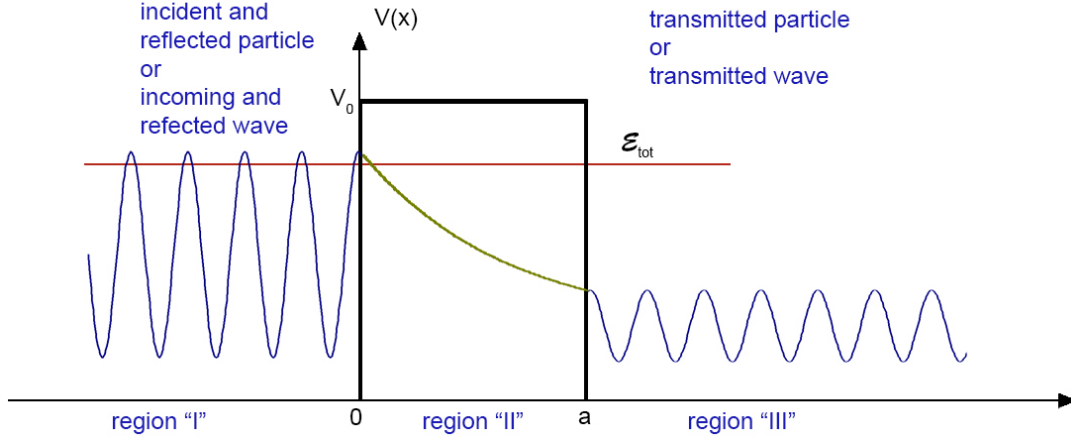


Figure 3.17: Quantum tunneling. A wave function corresponding to an incident particle with a total energy of \mathcal{E}_{tot} arrives at the potential barrier from the left. Some of it is reflected back, and some of it travels through the barrier. The total energy of the electron does not change only the amplitude of the transmitted part of the wave function is smaller in region III than in region I.

$$\varphi_{II} = C e^{-qx} + D e^{qx} \quad q = \frac{\sqrt{2m(V_0 - \mathcal{E})}}{\hbar} \quad (3.7.1b)$$

$$\varphi_{III} = G e^{ikx} \quad (3.7.1c)$$

The 4 boundary conditions are the usual ones:

$$\varphi_I(0) = \varphi_{II}(0) \quad (3.7.2a)$$

$$\varphi'_I(0) = \varphi'_{II}(0) \quad (3.7.2b)$$

$$\varphi_{II}(a) = \varphi_{III}(a) \quad (3.7.2c)$$

$$\varphi'_{II}(a) = \varphi'_{III}(a) \quad (3.7.2d)$$

We are interested in the *transmission coefficient* T , which is defined as

$$T := \left| \frac{G}{A} \right|^2$$

and can be calculated from the boundary conditions. As it turns out it is easier to calculate $1/T$ instead. From equations (3.7.1) and (3.7.2):

$$\frac{A}{G} = \left[\frac{1}{2} + \frac{i}{4} \left(\frac{q}{k} - \frac{k}{q} \right) \right] e^{(ik+q)a} + \left[\frac{1}{2} + \frac{i}{4} \left(\frac{k}{q} - \frac{q}{k} \right) \right] e^{i(k-q)a} \quad (3.7.3)$$

If we assume that $V_0 \gg \mathcal{E}$ then it follows that $q \gg k$ and $\frac{q}{k} - \frac{k}{q} \approx \frac{q}{k}$. Furthermore let a be so large that $q \cdot a > 1$, which results in $e^{qa} \gg e^{-qa}$. Using these simplifications

$$T = \left| \left(\frac{1}{2} + \frac{i q}{4 k} \right) e^{(ik+q)a} \right|^2 = \frac{16}{4 + \left(\frac{q}{k} \right)^2},$$

which, with $\left| \frac{q}{k} \right|^2 = \frac{V_0 - \mathcal{E}}{\mathcal{E}}$ leads to the result:

$$T \approx e^{-2qa} = e^{-2 \frac{\sqrt{2m(V_0 - E)}}{\hbar} a} \quad (3.7.4)$$

Example 3.8. *In an aluminum–aluminum oxide–aluminum layer structure a current of electrons with energies of 1 eV flows through the 0.5 nm thick insulating oxide boundary, which we represent as a square potential barrier with $V_0 = 10$ eV. What is the probability of an electron to pass through the barrier?* **Solution**

$$\begin{aligned} q &= \frac{\sqrt{2m_e(V_0 - E)}}{\hbar} = \frac{\sqrt{18 \cdot 10^{-19} \cdot 9.1 \cdot 10^{-31}}}{1.055 \cdot 10^{-34}} \\ &= 1.0537 \cdot 10^{10} \text{ m}^{-1} \\ qa &= 7.685 \\ T &\approx e^{-ga} = 2.11 \cdot 10^{-7} \end{aligned}$$

When the total energy of a classical particle is larger than the height of the potential barrier ($\mathcal{E} > V_0$) it always passes through it. However particles with quantum mechanical descriptions may be reflected (i.e. $B \neq 0$) even in this case, *except* if their energy is such that

$$\begin{aligned} qa &= n\pi && \text{i.e.} \\ \mathcal{E}_n &= V_0 + \frac{h^2}{8ma} \end{aligned}$$

when they pass through without reflection (i.e. $B = 0$). This is called *resonance tunneling*.

Nitrogen inversion in ammonia

Quantum tunneling can describe the inversion of an ammonia²⁶ (NH_3) molecule called *nitrogen inversion*. As you see in Fig 3.18 the N atom in the molecule may be placed in two equivalent equilibrium positions. It can move to and fro between these through the plane of the hydrogen atoms. Because the equilibrium positions correspond to the

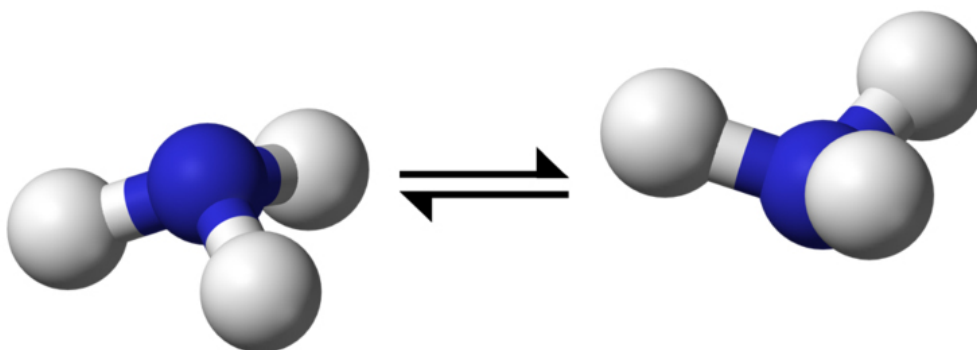


Figure 3.18: The N atom in an ammonia molecules may undergo a geometrical inversion.

minima of the potential there is a potential barrier between them created by the hydrogen atoms. See Fig 3.19. The total energy of the N atom is close to the minimum energy, therefore it can only move from one position to the other by quantum tunneling. In this system there are two kinds of oscillations present. First there is an oscillation around either of the minima, then there is a slower oscillation between the two minima. The wave function of the nitrogen atom is the superposition of these two oscillating states. The height of the potential barrier is about 24.7 kJ/mol (0.256 eV/atom) and the resonance frequency is 23.79 GHz ($\lambda = 1.260 \text{ cm}$) in the microwave range²⁷.

Field emission

Inside a metal (conductor) the potential the electrons move in is the superposition of the potentials from the ion cores and the other electrons. The resulting potential inside the conductor is lower than that of the free electron, called the vacuum level (which is usually considered to be 0) and we set it approximately constant. At the surface the potential climbs to the vacuum level. The (conduction) electrons in a conductor may move freely

²⁶'Household ammonia' or 'ammonium hydroxide' is a 5-10 wight% solution of ammonia in water.

²⁷The absorption at this frequency was the first microwave spectrum to be observed.

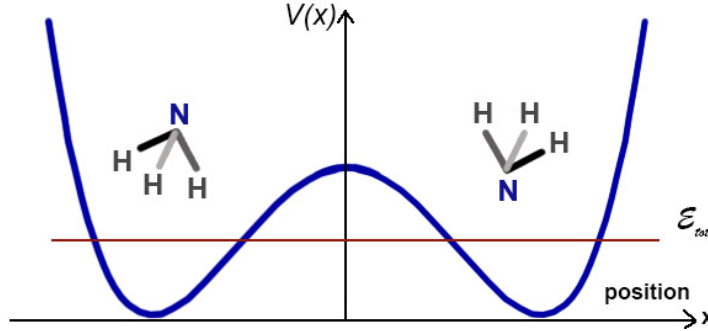


Figure 3.19: Potential of the nitrogen atom in an ammonia molecule. Around the equilibrium the potential can be approximated by the quadratic potential of the linear harmonic oscillator.

around²⁸, but ordinarily they cannot leave it because outside the solid the potential is $\Delta\Phi$ higher than the total energy of any electron inside it. To remove an electron from

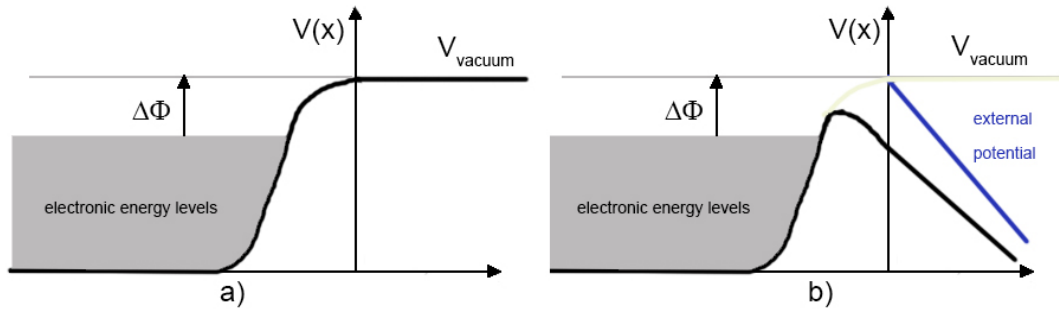


Figure 3.20: Schematic potential energy of electrons in a conductor a), and field emission b)

the solid therefore we either provide it with an additional $\Delta\mathcal{E} = e \Delta\Phi$ energy by thermal excitation or photon impact or we apply an external static electric field which distorts the potential (as in Fig 3.20 forming a potential barrier b). In the latter case we are talking about *field emission*.

²⁸The picture we present here is a simplification, you can find a more exact discussion in Chapter 14 and 15.

Chapter 4

Time dependent Schrödinger equation

4.1 Solutions of the time dependent Schrödinger equation

In section 3.5 we saw that the general solution of the time-dependent Schrödinger equation can be written using the solutions (eigenfunctions) of the time-independent Schrödinger equation (see (3.5.9)):

$$\psi(x, t) = \sum_n C_n \varphi_n(x) e^{-i \varepsilon_n / \hbar t} \quad (4.1.1)$$

where the C_n numbers are complex. The probability density will be time dependent too (the asterisk denotes the complex conjugate):

$$\begin{aligned} \mathcal{P}(x, t) &= |\psi|^2 = \psi^* \cdot \psi = \left(\sum_n C_n^* \varphi_n^*(x) e^{+i \frac{\varepsilon_n}{\hbar} t} \right) \cdot \left(\sum_m C_m \varphi_m(x) e^{-i \frac{\varepsilon_m}{\hbar} t} \right) = \\ &= \sum_{n,m} C_n^* C_m \varphi_n^* \varphi_m e^{i \frac{(\varepsilon_n - \varepsilon_m)}{\hbar} t} = \\ &= \sum_n C_n^* C_n \varphi_n^* \varphi_n + \sum_{n \neq m} C_n^* C_m \varphi_n^* \varphi_m e^{i \frac{(\varepsilon_n - \varepsilon_m)}{\hbar} t} = \end{aligned}$$

I.e.

$$\mathcal{P}(x, t) = \sum_n |C_n|^2 |\varphi_n|^2 + \sum_{n \neq m} C_n^* C_m \varphi_n^* \varphi_m e^{i \frac{(\varepsilon_n - \varepsilon_m)}{\hbar} t} \quad (4.1.2)$$

The first sum is the weighted sum of the probability densities of the eigenfunctions and the second sum is the interference term. As you see the probability density function is

not constant as it would be in the stationary case, but contains terms oscillating with angular frequencies

$$\omega_{n,m} = \frac{\mathcal{E}_n - \mathcal{E}_m}{\hbar} \quad (4.1.3)$$

These oscillations describe transitions between stationary states. The frequency of these oscillations is the frequency of photons the system can absorb or emit:

$$h \nu = \mathcal{E}_n - \mathcal{E}_m \quad (4.1.4)$$

At $t = 0$ (4.1.2) further simplifies to

$$\mathcal{P}(x, 0) = \sum_n |C_n|^2 |\varphi|^2$$

Example 4.1. *As an example let us use a wave function which is the linear combination of only two eigenfunctions¹*

$$\begin{aligned} \psi(x, t) &= C_1 \psi_1 + C_2 \psi_2 = C_1 \varphi_1 e^{-i \frac{\mathcal{E}_1}{\hbar} t} + C_2 \varphi_1 e^{-i \frac{\mathcal{E}_2}{\hbar} t} \\ \mathcal{P}(x, t) &= |\psi|^2 = |C_1 \psi_1 + C_2 \psi_2|^2 = \\ &= |C_1|^2 |\psi_1|^2 + |C_2|^2 |\psi_2|^2 + C_1^* C_2 \psi_1^* \psi_2 + C_1 C_2^* \psi_2^* \psi_1 = \\ &= |C_1|^2 |\varphi_1|^2 + |C_2|^2 |\varphi_2|^2 + C_1 C_2^* \varphi_1^* \varphi_2 e^{i \frac{\mathcal{E}_1 - \mathcal{E}_2}{\hbar} t} + \\ &\quad C_1 C_2^* \varphi_1 \varphi_2^* e^{-i \frac{\mathcal{E}_1 - \mathcal{E}_2}{\hbar} t} \end{aligned}$$

That is the probability density function now oscillates with a single angular frequency of

$$\omega = \frac{\mathcal{E}_1 - \mathcal{E}_2}{\hbar} \quad (4.1.5)$$

The integral of the probability density function \mathcal{P} for the whole space is the probability of

¹This is a special case of (4.1.2) when $C_n = 0$ for $n \neq 1$, or 2

the particle being somewhere in space, therefore it must be equal to 1, independent of t :

$$\begin{aligned}
 1 &= \int_{-\infty}^{\infty} \mathcal{P}(x, t) dx = \int_{-\infty}^{\infty} \mathcal{P}(x, 0) dx = \\
 &= \int_{-\infty}^{\infty} \psi^*(x, 0) \psi(x, 0) dx = \\
 &= \int_{-\infty}^{\infty} (C_1 \psi_1 + C_2 \psi_2)^* \cdot (C_1 \psi_1 + C_2 \psi_2) dx \tag{4.1.6}
 \end{aligned}$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} (|C_1|^2 |\psi_1|^2) dx + \int_{-\infty}^{\infty} (|C_2|^2 |\psi_2|^2) dx + \\
 &\quad \int_{-\infty}^{\infty} (C_1^* C_2 \psi_1^* \psi_2 + C_1 C_2^* \psi_1 \psi_2^*) dx = \\
 &= |C_1|^2 \int_{-\infty}^{\infty} |\psi_1|^2 dx + |C_2|^2 \int_{-\infty}^{\infty} |\psi_2|^2 dx + \\
 &\quad C_1^* C_2 \int_{-\infty}^{\infty} \psi_1^* \psi_2 dx + C_1 C_2^* \int_{-\infty}^{\infty} \psi_1 \psi_2^* dx \tag{4.1.7}
 \end{aligned}$$

If you recall the (3.5.13) definition of the scalar product, you can write (4.1.6) and the equivalent (4.1.7) in a shorter form as

$$\langle \psi | \psi \rangle = \langle C_1 \psi_1 + C_2 \psi_2 | C_1 \psi_1 + C_2 \psi_2 \rangle = 1$$

so

$$\begin{aligned}
 &|C_1|^2 \langle \psi_1 | \psi_1 \rangle + |C_2|^2 \langle \psi_2 | \psi_2 \rangle + \\
 &C_1^* C_2 \langle \psi_1 | \psi_2 \rangle + C_2^* C_1 \langle \psi_2 | \psi_1 \rangle = 1
 \end{aligned}$$

From which you can see the properties of the scalar product:

Important 4.1.1.

$$\begin{aligned}
 |\psi|^2 &\equiv \langle \psi | \psi \rangle \geq 0 \\
 \langle \psi_1 | \psi_2 \rangle &= \langle \psi_2 | \psi_1 \rangle^* \\
 \langle \psi_1 + \psi_2 | \psi_3 \rangle &= \langle \psi_1 | \psi_3 \rangle + \langle \psi_2 | \psi_3 \rangle \\
 \langle \psi | C \psi \rangle &= C \langle \psi | \psi \rangle \\
 \langle C \psi_1 | \psi_2 \rangle &= C^* \langle \psi_1 | \psi_2 \rangle
 \end{aligned} \tag{4.1.8}$$

If φ_1 and φ_2 were normalized wave functions themselves then

$$\begin{aligned}\langle \psi_1 | \psi_1 \rangle &= \langle \varphi_1 | \varphi_1 \rangle = 1 \\ \langle \psi_2 | \psi_2 \rangle &= \langle \varphi_2 | \varphi_2 \rangle = 1\end{aligned}$$

and because they are different eigenfunctions of the same Schrödinger equation they are orthogonal to each other

$$\langle \psi_1 | \psi_2 \rangle = \langle \psi_2 | \psi_1 \rangle = 0$$

therefore the normalization condition gives

$$|C_1|^2 + |C_2|^2 = 1 \quad (4.1.9)$$

4.1.1 Free electron in 1D

The energy spectrum of a free electron is continuous and the wave function of a free electron is

$$\psi(x, t) = \sum_{k \geq 0} \left(C_k^{(+)} e^{i(kx - \omega t)} + C_k^{(-)} e^{-i(kx + \omega t)} \right)$$

where '+' and '-' refers to the waves traveling in the positive and negative x direction respectively. We may incorporate this sign into k and notice that because k is continuous we must add values with infinitely close k indices, and this is what the integral calculus were invented for:

$$\psi(x, t) = \sum_k C_k e^{i(kx - \omega t)} \Rightarrow \int_{-\infty}^{\infty} C(k) e^{i(kx - \omega t)} dk \quad (4.1.10)$$

4.1.2 Particle in a 1 dimensional potential box

The eigenfunctions in this case can be written as (see section 3.5.3)

$$\psi_n(x, t) = A_n \sin \frac{n\pi}{L} x \cdot e^{-i \frac{E}{\hbar} t} = \frac{A_n}{2i} \left[e^{i \left(\frac{n\pi}{L} x - \frac{E}{\hbar} t \right)} - e^{-i \left(\frac{n\pi}{L} x - \frac{E}{\hbar} t \right)} \right] \quad (4.1.11)$$

This standing sine wave is a result of two waves of the same wavelength $\lambda_n = \frac{2L}{n}$ and amplitude moving in the opposite directions. This is a stationary wave because the probability

$$|\psi|^2 = |\varphi|^2$$

density does not depend on time.

The wave functions we used are not the only ones possible for a potential box, because any linear combination of them is also a solution.

4.2 Perturbation theory

We learned in the previous section that if the wave function of the system is a *superposition* (i.e. linear combination) of eigenfunctions, then the probability density contains oscillating terms, which describe the transition from one state to the other. In this section we will discuss how and why can such a transition occur.

To make the system jump from one stationary state to an other one we need some external interaction, called *perturbation*. This can be in the form of excitation by the radiation field (photons) or excitation by thermal vibrations for instance. But how can we include this excitation into the Schrödinger equation?

Suppose that we can solve the Schrödinger equation for the $V(x)$ potential, i.e.

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi_n(x, t)}{\partial x^2} + V(x) \psi_n(x, t) = i \hbar \frac{\partial \psi_n(x, t)}{\partial t} \Rightarrow \psi_n(x, t) = \varphi_n(x) e^{-i \frac{\mathcal{E}_n}{\hbar} t}$$

where the mutually orthogonal $\varphi_n(x)$ eigenfunctions are known. We describe the perturbation as an additional $K(x, t)$ potential which is small compared to V . Then in the time dependent Schrödinger equation:

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + (V(x) + K(x, t)) \psi(x, t) = i \hbar \frac{\partial \psi(x, t)}{\partial t} \quad (4.2.1)$$

The small $K(x)$ potential only slightly modifies – we say *perturbates* – the potential in the system.

Although we probably cannot solve this modified equation exactly we may derive an equation from it which can be solved by iteration. First we try to write the unknown solution as a linear combination of the known eigenfunctions of the *unperturbed* Schrödinger equation. For this we use (4.1.1) modified in a simple way: we introduce time dependency into the complex coefficients C_n

$$\psi(x, t) = \sum_n C_n(t) \varphi_n(x) e^{-i \frac{\mathcal{E}_n}{\hbar} t} \quad (4.2.2)$$

The $C_n(t)$ coefficients satisfy the following equation²:

$$\frac{dC_m(t)}{dt} = -\frac{i}{\hbar} \sum_n K_{mn}(t) C_n(t) e^{i \omega_{mn} t} \quad (4.2.3)$$

where $\omega_{mn} = \frac{\mathcal{E}_m - \mathcal{E}_n}{\hbar}$ and $K_{mn}(t)$ is called the m, n -th *matrix element* of the potential $K(x, t)$ and defined by

$$K_{mn}(t) := \int_{-\infty}^{\infty} \varphi_m^*(x) K(x, t) \varphi_n(x) dx \quad (4.2.4)$$

²You find the derivation in Appendix 22.8

Using the notation introduced for the scalar product (4.2.4) can be written in the form:

$$K_{mn}(t) := \langle \varphi_m(x) | K(x, t) \varphi_n(x) \rangle = \langle \varphi_m(x) | K(x, t) | \varphi_n(x) \rangle \quad (4.2.5)$$

too, where the second notation is called *the m, n matrix element of the potential $K(x, t)$* . In this second form $\langle \varphi |$ corresponds to the factor φ^* in the integral. The complicated (4.2.3) equation may be solved by *successive approximation*. The r -th approximation is (see (22.8.2)):

$$C_m^{(r)}(t) = C_m^{(r-1)}(0) - \frac{i}{\hbar} \sum_n \int_{-\infty}^{\infty} K_{mn}(\tau) e^{i\omega_{mn}\tau} C_n^{(r-1)}(\tau) d\tau \quad (4.2.6)$$

4.3 Transition probabilities and selection rules

Example 4.2. *Let us suppose that we have a system with only two eigenstates $\psi_1(x, t)$ and $\psi_2(x, t)$. Under the influence of the external K perturbation it jumps into the superposition of these eigenstates. In this case*

$$\psi(x, t) = C_1(t) \varphi_1(x) e^{-\frac{i}{\hbar} \mathcal{E}_1 t} + C_2(t) \varphi_2(x) e^{-\frac{i}{\hbar} \mathcal{E}_2 t}$$

If ψ is normalized then $|C_1(t)|^2 + |C_2(t)|^2 = 1$. The equations for C_1 and C_2 are

$$\begin{aligned} \frac{dC_1(t)}{dt} &= \frac{i}{\hbar} \left[K_{11}(t) C_1(t) + K_{12}(t) C_2(t) e^{-\frac{\mathcal{E}_2 - \mathcal{E}_1}{\hbar} t} \right] \\ \frac{dC_2(t)}{dt} &= \frac{i}{\hbar} \left[K_{22}(t) C_2(t) + K_{21}(t) C_1(t) e^{\frac{\mathcal{E}_2 - \mathcal{E}_1}{\hbar} t} \right] \end{aligned}$$

If at $t = 0$ the system was in its $\psi_1(x, t)$ eigenstate, with $C_1(0) = 1$ and $C_2(0) = 0$, and after a certain t_1 time it is found in its $\psi_2(x, t)$ eigenstate, i.e. $C_1(t_1) = 0$, $C_2(t_1) = 1$, then the system underwent a transition from $\psi_1(x, t)$ to $\psi_2(x, t)$.

In the first approximation therefore

$$C_2^{(1)}(t) = -\frac{i}{\hbar^2} \int_0^t K_{21}(\tau) e^{i\omega_{21}\tau} d\tau$$

$|C_2(t)|^2$ is the probability that the system is in state “2” at time t . In other words the probability of the transition between states “1” and “2” is

$$W(1 \rightarrow 2) = |C_2(t)|^2 = \frac{1}{\hbar^2} \left| \int_0^t K_{21}(\tau) e^{i\omega_{21}\tau} d\tau \right|^2 \quad (4.3.1)$$

If at $t = 0$ the system was in its n -th eigenstate (i.e. $C_k(0) = \delta_{kn}$ for $k = 1, 2, \dots$) and after the transition it will be in its m -th eigenstate (i.e. $C_k(0) = \delta_{km}$ for $k = 1, 2, \dots$), then in the first approximation ($C_m(t) \approx C_m^{(1)}(t)$)

$$W(n \rightarrow m) = |C_m(t)|^2 = \frac{1}{\hbar^2} \left| \int_0^t K_{mn}(\tau) e^{i\omega_{mn}\tau} d\tau \right|^2 \quad m \neq n \quad (4.3.2)$$

The transition probabilities depend on the matrix elements K_{mn} . It is possible that although both the m -th and the n -th states are stationary eigenstates of the system still $K_{mn} = 0$ may hold therefore no $n \rightarrow m$ transition is possible *in the first approximation*. The no $n \rightarrow m$ transition is called a *forbidden transition*. Because for the perturbation theory to remain applicable the $K(x, t)$ “potential” must be small relative to $V(x)$ higher approximations give only small corrections to the first approximation. But when a transition is prohibited in the first approximation these small corrections may become observable and the transition may occur although with a very small probability. Of course it is also possible that a transition is prohibited in all approximations.

Example 4.3. Determine the transition probability in a 1 dimensional two level system under the influence of an external electromagnetic field. In this case the perturbation is of the form:

$$K(x, t) = \mathcal{K}(x) \cdot \cos \omega t,$$

where ω is very close to ω_{21} in the sense³ that

$$\omega_{21} + \omega \gg |\omega_{21} - \omega|$$

and both are in the optical range ($\approx 10^{14} \text{ Hz}$). What is the range of validity of the perturbation theory in this case? What interesting behavior will you find and why? **Solution From (4.3.1)**

$$W(1 \rightarrow 2) = \frac{1}{\hbar^2} \left| \int_0^t \mathcal{K}_{21} \cos(\omega \tau) e^{i\omega_{21}\tau} d\tau \right|^2$$

where $\omega_{21} = (\mathcal{E}_2 - \mathcal{E}_1)/\hbar$ and $\mathcal{K}_{21} \equiv \int_{-\infty}^{\infty} \varphi_2^*(x) \mathcal{K} \varphi_1^*(x) dx$. Because $\cos \omega \tau = (e^{i\omega \tau} +$

³This is not a serious limitation, because perturbations with other frequencies have a negligible probability to cause a transition anyway.

$$e^{-i\omega\tau})/2$$

$$\begin{aligned} W(1 \rightarrow 2) &= \frac{|\mathcal{K}_{21}|^2}{2\hbar^2} \left| \int_0^t (e^{i(\omega_{21}+\omega)\tau} + e^{i(\omega_{21}-\omega)\tau}) d\tau \right|^2 = \\ &= \frac{|\mathcal{K}_{21}|^2}{4\hbar^2} \left| \frac{e^{i(\omega_{21}+\omega)t} - 1}{\omega_{21} + \omega} + \frac{e^{i(\omega_{21}-\omega)t} - 1}{\omega_{21} - \omega} \right|^2 \end{aligned}$$

Now because of our assumptions for ω and ω_{21} the first term in the absolute sign may be neglected as it is much smaller than the second one (the numerator is of the same magnitude, while the denominator of the first term is much greater than in the second one)

$$\begin{aligned} W(1 \rightarrow 2) &\approx \frac{|\mathcal{K}_{21}|^2}{4\hbar^2} \left| \frac{e^{i(\omega_{21}-\omega)t} - 1}{\omega_{21} - \omega} \right|^2 = \\ &= \frac{|\mathcal{K}_{21}|^2}{4\hbar^2} \left| \frac{e^{i(\omega_{21}-\omega)t/2} - e^{-i(\omega_{21}-\omega)t/2}}{\omega_{21} - \omega} \right|^2 = \\ &= \frac{|\mathcal{K}_{21}|^2}{\hbar^2} \frac{\sin^2[(\omega_{21} - \omega)t/2]}{(\omega_{21} - \omega)^2} \end{aligned}$$

If $|\omega_{21} - \omega| \ll 1$ then the sine can be approximated with its argument and the maximum of $W \propto |\mathcal{K}|^2 t^2 / \hbar^2$ which increases with t . However the assumption that this is a small perturbation will become invalid long before this maximum reaches 1. Therefore our result is only valid for relatively small t .

The most interesting feature of this solution is that the transition probability *oscillates* sinusoidally as a function of time between 0 and a maximum value which is still much less than 1, otherwise this would not be a *small* perturbation. When $t = \frac{2\pi n}{|\omega_{21} - \omega|}$, where $n = 1, 2, 3, \dots$ the particle will be back in the lower state.

The reason for this behavior is that although ψ_1 and ψ_2 are eigenfunctions (i.e. stationary states) of the non-perturbed system they are not eigenfunctions of the perturbed system.

4.3.1 Selection rules

Selection rules determine whether a transition is possible (i.e. $W(1 \rightarrow 2) > 0$). In some cases there is no need to calculate the integral, because these are connected to conservation laws (e.g. conservation of angular momentum). Symmetry arguments can also be used. For instance when the $f(x) := \varphi_m(x) K(x, t) \varphi_n(x)$ function under the integral

is *antisymmetric*, i.e. $f(-x) = -f(x)$ the integral is always zero and the corresponding transition is forbidden.

In the case of the linear harmonic oscillator the selection rule states that only photons with the eigenfrequency of the oscillator can be absorbed or emitted ($\hbar\omega = \mathcal{E}_m - \mathcal{E}_n$), therefore the selection rule in this case states that

$$m = n \pm 1 \quad \Rightarrow \quad \Delta n = \pm 1$$

4.4 Radiative transitions

Particles in a particular potential are considered a quantum mechanical system. Let us consider a system with two energy levels and the transitions between them. When an electromagnetic radiation of suitable frequency ($\hbar\nu = \mathcal{E}_m - \mathcal{E}_n$) interacts with this system and the transition is allowed the system may absorb the photon and go to the higher lying energy level with the transition probability of $W(n \rightarrow m)$. If the transition in the reverse direction is also possible the system may return to its original state by emitting a photon with the same frequency as of the absorbed one. Why would it do this?

If a higher energy level is occupied and a lower one is empty this is not a stable state. It is only metastable. There are two possible processes for photon emission from this state. First a second photon with the same frequency as the first may disturb the metastable state. According to our formulas the role of levels n and m are interchangeable⁴. Therefore a photon of the same frequency is needed to excite and de-excite the system. In this case two photons of the same frequency will leave the system: the emitted, and the perturbing one. This is called *induced emission* or *stimulated emission*. Therefore the principle of light amplification is simple: take a number of atoms, excite them to a higher metastable energy state by any means and then use a single incident photon to force one of them to return to the lower energy state. This will produce 2 photons, which can de-excite two atoms which in turn will emit 4 photons, etc. This will generate an enormous number of photons with the same frequency that leave the system at almost the same time. Of course you must be sure that most of the atoms are excited, because excitation and stimulated emission are competing processes. The photo amplification described serves the fundamental mechanism for *lasers*⁵.

The other possible process is called *spontaneous emission*. In that case seemingly nothing disturbs the system but it still returns to its lower energy state by emitting a single photon. Spontaneous emission can only be understood with Quantum Electrodynamics, which gives an uncertainty relation between the fields \mathbf{E} and \mathbf{B} . As a consequence there is no such thing as an “empty space”, the electromagnetic field has zero point vibrations. This is similar to the zero point vibrations of a linear harmonic oscillator and

⁴In the example above we only need to change ω_{21} to $\omega_{12} = -\omega_{21}$ and neglect the *second* term instead of the first.

⁵Acronym for “Light Amplification by Stimulated Emission of Radiation”. See Section 10.3

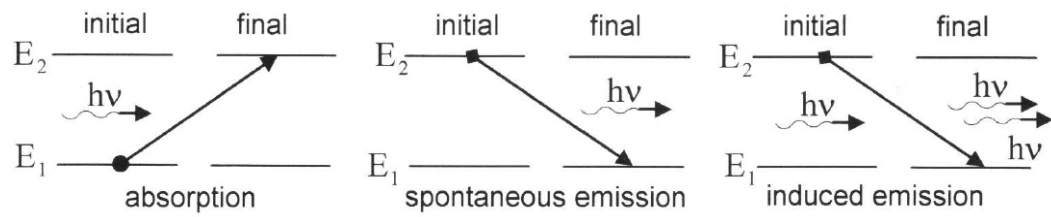


Figure 4.1: Possible radiative processes

usually called the *zero point energy* of the vacuum. This means that strictly speaking there is no such thing as a really spontaneous emission because this emission is caused by the interaction of these zero point vibrations of the electromagnetic field called *vacuum fluctuations*.

Chapter 5

Formal quantum mechanics

5.1 Formal quantum mechanics. Operators

We have become familiar with many features of quantum mechanics that describes the behavior of microscopic particles which are vastly different from the things we were accustomed to in classical physics:

1. The energy of spatially restricted (bound) particles may have only discrete values.
2. Particles are not described with coordinates and velocities, but with wave functions, which are solutions of the Schrödinger equation.
3. Some physical quantities (like the position and momentum) form so called conjugate pairs and there is an uncertainty relation between these pairs: the value of either member of such a pair can be determined with any degree of accuracy on the expense of decreasing accuracy (or increasing uncertainty) of the other member. They never can have exact values.
4. Classical electrodynamics is still valid¹, e.g. accelerating charged particles do emit electromagnetic radiation but e.g. electrons in stationary quantum states in atoms do not accelerate therefore do not radiate.

Most of the mathematical basis of formal quantum mechanics have been covered already, but there are some “new” mathematics involved too, like the use of linear vector spaces or matrix calculus.

Let us summarize the mathematical foundations in three dimensions²:

¹Quantum Electrodynamics not included.

²In previous discussions we used simpler 1D equations, but for the completeness of this section we will use 3 dimensions notation whenever it is not too inconvenient.

- The state $|\psi\rangle$ of a quantum mechanical system is described by the wave function $\psi(\mathbf{r}, t)$, which is³
 - complex valued,
 - continuous,
 - continuously differentiable (except where the potential has an infinite jump),
 - and quadratically integrable, functions. I.e. :

$$\iiint_{-\infty}^{\infty} \psi^*(\mathbf{r}, t) \psi(\mathbf{r}, t) d\mathbf{r} < \infty \quad (5.1.1)$$

Functions with all of these properties are called *regular functions*.

- Wave functions are solutions of the time dependent Schrödinger equation⁴:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}, t) + V(\mathbf{r}) \psi(\mathbf{r}, t) = i \hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t}$$

- Some wave functions can be written in the form

$$\psi(\mathbf{r}, t) = \varphi(\mathbf{r}) e^{-i\mathcal{E}/\hbar}, \quad (5.1.2)$$

where $\varphi(\mathbf{r})$ is the solution of the stationary Schrödinger equation:

$$-\frac{\hbar^2}{2m} \nabla^2 \varphi(\mathbf{r}) + V(\mathbf{r}) \varphi(\mathbf{r}) = \mathcal{E} \varphi(\mathbf{r}) \quad (5.1.3)$$

and \mathcal{E} is the energy eigenvalue of the system. These are called eigenfunctions of the stationary Schrödinger equation⁵.

Therefore the only new mathematical notion we have to introduce here is that of the *operator*.

³If there are more than one particle in the system, then the wave function is a function of all of the coordinates of all of the particles of the system: $|\psi\rangle = \psi(\mathbf{r}_1, \mathbf{r}_2, \dots, t)$

⁴ $\nabla \varphi \equiv \left(\frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial z} \right)$, and $\nabla^2 \varphi \equiv \left(\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} \right)$

⁵The meaning of the term “eigenfunction” is explained below

5.1.1 Operators

Most of the functions we are used to associate a single number or vector to every number or vector in a set, called *domain* by a formula. The set of the values associated to the elements of the domain is called *codomain*. Examples: the real valued $f(x) = \sqrt{x}$ function on the set of the real numbers associates a single real value for all positive real x numbers and 0, while the complex valued $g(z) = \sqrt{x}$ function on the set of the real numbers associates a single complex value to any real number:

x		f(x)	g(x)
1	→	1	1
2	→	4	4
3	→	9	9
-1	→	—	i
-2	→	—	4i
-3	→	—	9i

Functions are defined by their domain and codomain together with the rule of assignment.

Important 5.1.1. *An operator \hat{O} assigns a new function to an existing one:*

$$g(x) = \hat{O} f(x)$$

or in other words it transforms one function ($f(x)$) to an other one ($g(x)$). An operator is defined by its domain and its codomain of functions and the rule of the assignment or transformation (exactly as a function is defined).

E.g. differentiation is described by the formula $g(x) = \frac{d f(x)}{dx}$. For every differentiable functions $f(x)$ it associates its derivative $g(x)$, therefore it too may be written as an operator expression:

$$g(x) = \hat{D}_x f(x) \equiv \frac{d f(x)}{dx}$$

$f(x)$		$g(x) = \frac{d f(x)}{dx}$
const	→	0
x	→	1
x^2	→	x^2
x^3	→	x^3
$\sin x$	→	$\cos x$

Now we do a strange thing: we formally separate the symbols of the differentiation from the function and make a correspondence between them and the \hat{D} operator:

$$\hat{D}_x f(x) \equiv \frac{d}{dx} f(x) \quad \Rightarrow \quad \hat{D}_x \equiv \frac{d}{dx}$$

We can therefore say that the *operator of differentiation with respect to x* is $\frac{d}{dx}$. In this case the 'hat' symbol is not used. What are the properties of this operator? The same as the properties of the differentiation, i.e. if

$$\begin{aligned} f(x) &= C_1 f_1(x) + C_2 f_2(x) \quad \text{and} \\ g_1(x) &= \hat{D}_x f_1(x) \equiv \frac{d}{dx} f_1(x), \quad \text{and} \\ g_2(x) &= \hat{D}_x f_2(x) \equiv \frac{d}{dx} f_2(x) \end{aligned}$$

then

$$\begin{aligned} g(x) &= \hat{D}_x (C_1 f_1(x) + C_2 f_2(x)) \equiv \\ &\frac{d}{dx} (C_1 f_1(x) + C_2 f_2(x)) \equiv \frac{d (C_1 f_1(x) + C_2 f_2(x))}{dx} = \\ &= C_1 \frac{d f_1(x)}{dx} + C_2 \frac{d f_2(x)}{dx} = \\ &= C_1 \frac{d}{dx} f_1(x) + C_2 \frac{d}{dx} f_2(x) = \\ &= C_1 g_1(x) + C_2 g_2(x) = \\ &= \underline{C_1 \hat{D}_x f_1(x) + C_2 \hat{D}_x f_2(x)} \end{aligned}$$

i.e.

$$\hat{D}_x (C_1 f_1(x) + C_2 f_2(x)) = C_1 \hat{D}_x f_1(x) + C_2 \hat{D}_x f_2(x) \quad (5.1.4)$$

The \hat{D}_x operator transforms a linear combination of functions to the same linear combination of the transformed functions. Such operators are called *linear operators*. Examples of linear operators:

- n-th order ordinary or partial differentiation with respect to any a variable:

$$\hat{D}_x^{(n)} := \frac{d^n}{dx^n}, \quad \hat{D}_l^{(n)} := \frac{\partial^n}{\partial x_l^n}$$

Example:

$$\begin{aligned} \hat{D}_x^{(n)} [A f(x) + B g(x)] &= \frac{d^n [A f(x) + B g(x)]}{dx^n} = \\ &= A \frac{d^n f(x)}{dx^n} + B \frac{d^n g(x)}{dx^n} = \\ &= A \hat{D}_x^{(n)} f(x) + B \hat{D}_x^{(n)} g(x) \end{aligned}$$

- integral of a function:

$$\hat{I} := \int \dots dx$$

Example:

$$\begin{aligned}\hat{I} [A f(x) + B g(x)] &= \int [A f(x) + B g(x)] dx = \\ &= A \int f(x) dx + B \int g(x) dx = \\ &= A \hat{I} f(x) + B \hat{I} g(x)\end{aligned}$$

- multiplication with a number or with a function:

$$\hat{V} := V(x).$$

Example:

$$\begin{aligned}\hat{V} [A f(x) + B g(x)] &= V(x) [A f(x) + B g(x)] = \\ &= A V(x) f(x) + B V(x) g(x) = \\ &= A \hat{V} f(x) + B \hat{V} g(x)\end{aligned}$$

In contrast the operator of squaring is *not a linear operator*, because

$$\hat{S} \cdot (f(x) + g(x)) = \hat{S} \cdot f(x) + 2 \cdot f(x) \cdot g(x) + \hat{S} \cdot g(x) \neq \hat{S} \cdot f(x) + \hat{S} \cdot g(x)$$

Example 5.1. We define some operators with the formulas:

$$\hat{O}_1 \mathbf{v}(t) := v_x \quad - x \text{ coordinate of the velocity vector}$$

$$\hat{O}_2 f(t) := A \sin f(t) \quad - \text{sine of a time dependent function, e.g } f(t) = \omega t$$

$$\hat{O}_3 f(k) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(k) e^{-ikx} dk \quad - \text{Fourier transform of } f(k)$$

$$\hat{O}_4 f(x) := \sqrt[3]{f(x)}$$

Which of these are the linear operators? **Solution** \hat{O}_1 and \hat{O}_3

One advantage of using operators is that we can define operations between them too. This makes our calculations easier. Some possible operations involving operators are:

Important 5.1.2. Addition of operators *The sum of two operators is also an operator. This operator transforms a function $f(x)$ to the sum of the two transformed functions:*

$$\text{if } \hat{O} = \hat{O}_1 + \hat{O}_2, \text{ then } \hat{O} f(x) = \hat{O}_1 f(x) + \hat{O}_2 f(x)$$

Multiplication of operators *The product of two operators is also an operator. This operator is defined by the successive application of the two operators in the given order:*

$$\text{if } \hat{O} = \hat{O}_1 \cdot \hat{O}_2, \text{ then } \hat{O} f(x) = \hat{O}_1 (\hat{O}_2 f(x))$$

The order of the operators in a product may be important. If

$$\hat{O} \hat{P} \neq \hat{P} \hat{O}$$

then the operators do not commute:

Multiplication of an operator with a (complex) number *Operators can be multiplied with a (complex) number. The result is also an operator:*

$$\text{if } \hat{O} = C \cdot \hat{O}_1, \text{ then } \hat{O} f(x) = C \cdot \hat{O}_1 f(x)$$

An example for *non-commuting operators* is the differentiation of a scalar function with respect to x ($\hat{D}_x = \frac{d}{dx}$) and the multiplication with the coordinate x ($\hat{X} = x \cdot$), because:

$$\begin{aligned} (\hat{D}_x \hat{X}) f(x) &= \frac{d}{dx} (x \cdot f(x)) = \\ &= f(x) + x \cdot \frac{df(x)}{dx} \\ (\hat{X} \hat{D}_x) f(x) &= x \cdot \frac{df(x)}{dx} \end{aligned}$$

which means that

$$\begin{aligned} \hat{D}_x \hat{X} f(x) &\neq \hat{X} \hat{D}_x f(x) \\ \hat{D}_x \hat{X} &\neq \hat{X} \hat{D}_x \end{aligned}$$

5.1.2 Operators in Quantum Mechanics. Angular momentum

One axiom of formal quantum mechanics is, that

Important 5.1.3. *In quantum mechanics every physical quantity is represented by a linear, self-adjoint (or Hermitian⁶) operator, which acts on complex valued functions. This operator can be found using the definition of the corresponding quantity from classical mechanics and substituting all of the classical physical quantities in it with the corresponding operators. As all classical mechanical quantities can be expressed with a combination or as a function of the momentum \mathbf{p} and position vector \mathbf{r} , we can determine their operators if we know the operators $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$.*

This means that we have some freedom in selecting operators for the momentum and position, but the operators of other physical quantities must be calculated using these.

We can use the stationary Schrödinger equation to determine suitable operators for \hat{p} and \hat{x} :

Take the classical mechanical formula for the total energy of a particle and replace \mathcal{E} , \mathbf{p} and $V(r)$ with operators $\hat{\mathcal{E}}$, $\hat{\mathbf{p}}$ and $\hat{\mathbf{V}}$ respectively. This will result in an operator equation. Apply both sides of this equation to the wave function $\varphi(x)$ then compare the resulting formula with the stationary Schrödinger equation to determine a possible representation of these operators. In one dimension:

$$\mathcal{E} = \frac{p^2}{2m} + V(x) \quad \text{classical mechanics}$$

$$\hat{\mathcal{E}} = \frac{\hat{p}^2}{2m} + \hat{V}(x) \quad \text{operator equation in quantum mechanics}$$

$$\hat{\mathcal{E}}\varphi(x) = \frac{\hat{p}^2}{2m}\varphi(x) + \hat{V}(x)\varphi(x) \quad \text{applied to the wave function}$$

$$\mathcal{E} = -\frac{\hbar^2}{2m} \frac{\partial^2 \varphi}{\partial x^2} + V(x)\varphi \quad \text{the stationary Schrödinger equation}$$

From this (using the rules of the sum and product of operators) we obtain the definitions of the operators \hat{p} and \hat{x} :

$$\hat{\mathcal{E}}\varphi \equiv E\varphi \Rightarrow \hat{\mathcal{E}} = E. \quad (5.1.5)$$

$$\frac{\hat{p}^2}{2m}\varphi \equiv -\frac{\hbar^2}{2m} \frac{d^2 \varphi}{dx^2} \Rightarrow \hat{p} = \frac{\hbar}{i} \frac{d}{dx} \hat{V}(x)\varphi \equiv V(x)\varphi \Rightarrow \hat{V}(x) = V(x). \quad (5.1.6)$$

The last one for $V(x) = x$ implies that

$$\hat{x} = x. \quad (5.1.7)$$

⁶For the meaning of the terms “self-adjoint” and “Hermitian” see (5.1.21).

Using these definitions it is easy to see (c.f. end of previous section) that the operators of p and x do not commute, and

$$\hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar \quad (5.1.8)$$

To characterize the commutativity of operators we introduce a notation:

Important 5.1.4. *For any two quantum mechanical operators the quantity*

$$[\hat{O}, \hat{P}] := \hat{O}\hat{P} - \hat{P}\hat{O} \quad (5.1.9)$$

*is called the commutator of \hat{O} and \hat{P} .
If the two operators commute $[\hat{O}, \hat{P}] = 0$.*

Commutators are useful. In Appendix 22.6 for instance we used only commutators to derive the possible energies of the linear harmonic oscillator.

The commutator of the position and momentum in 1D is

$$[\hat{x}, \hat{p}] = i\hbar (\neq 0) \quad (5.1.10)$$

and we know there is an uncertainty relation between x and p . This is a general principle.

Important 5.1.5. *If the commutator of two operators is not zero then there exists an uncertainty relation between the corresponding physical quantities.*

In 3 dimensions we must define operators for all three components of the momentum and position vectors. We have to use partial derivatives in this case. These operators may also be combined into a *vector operator*

$$\hat{\mathbf{p}} := (\hat{p}_x, \hat{p}_y, \hat{p}_z) = -\frac{\hbar}{i} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad (5.1.11)$$

In this case the operator of \mathbf{p} is a constant multiplied *symbolic vector*, for which there is a special notation, the ∇ symbol, called *nabla* or *del*:

$$\nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad (5.1.12)$$

Although this is not a real vector in many cases we may use it as one. For instance the square of it, which is called the *Laplace operator* and denoted by Δ is:

$$\Delta := \nabla^2 \equiv \nabla \cdot \nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) = \quad (5.1.13)$$

$$= \left(\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial y^2}, \frac{\partial^2}{\partial z^2} \right) \quad (5.1.14)$$

Similarly the position operator \mathbf{r} is a symbolic vector:

$$\hat{\mathbf{r}} := (\hat{x}, \hat{y}, \hat{z}) = (x\cdot, y\cdot, z\cdot) \quad (5.1.15)$$

The commutators of the components of $\hat{\mathbf{p}}$, and $\hat{\mathbf{r}}$ then can be written in a single formula:

Important 5.1.6.

$$[x_k, p_l] = i \hbar \delta_{kl} \quad \text{where } k, l = 1, 2, 3 \text{ and e.g. } x_2 \equiv y, p_2 \equiv p_y \quad (5.1.16)$$

i.e. different components of the position and the momentum commute, but the same components do not, therefore there is no uncertainty relation between different components, only between the same components of $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$.

The combination of the operators \hat{p} and \hat{V} in the 1D and 3 dimensions Schrödinger equation are operators themselves. They are called the (1D and 3 dimensions) Hamilton operator or *Hamiltonian* of the system:

$$\hat{H} := -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x). \quad \text{1D} \quad (5.1.17a)$$

$$\hat{H} := -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}). \quad \text{3 dimensions} \quad (5.1.17b)$$

Therefore the one and three dimensional stationary Schrödinger equations both can be written as an operator equation:

$$\begin{aligned} \hat{H} \varphi(x) &= \mathcal{E} \varphi(x) \text{ and} \\ \hat{H} \varphi(\mathbf{r}) &= \mathcal{E} \varphi(\mathbf{r}) \end{aligned} \quad (5.1.18)$$

When dealing with the quantum mechanical problem of an electron in a centrally symmetric potential e.g. in the hydrogen atom it will be much easier to solve the Schrödinger equation in an (r, θ, ϕ) spherical coordinate system than in a Cartesian one. In this case the operator form of the equation will not change, although the formula for the Hamiltonian will change significantly (see Appendix 22.9):

$$\begin{aligned} \hat{H} = & -\frac{\hbar^2}{2m} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right] + \\ & + V(r). \end{aligned} \quad (5.1.19)$$

Important 5.1.7. *If we use the operator form of an equation it will remain the same independent of the coordinate system, only the representation of the operator will change. This is another important and very useful property of the operators.*

For the use in quantum physics it is important to know the behavior of operators in a scalar product.

Let \hat{O} an operator that acts on wave functions. We can calculate the scalar product of a function φ_2 with $\hat{O} \varphi_1$. In 1D:

$$\langle \varphi_2 | \hat{O} \varphi_1 \rangle \equiv \int_{-\infty}^{\infty} \varphi_2^* \cdot (\hat{O} \varphi_1) dx \quad (5.1.20)$$

Let us introduce a *new* operator \hat{O}^\dagger , called the *adjoint* of \hat{O} with the definition:

$$\langle \hat{O}^\dagger \varphi_2 | \varphi_1 \rangle = \langle \varphi_2 | \hat{O} \varphi_1 \rangle \quad (5.1.21a)$$

$$\int_{-\infty}^{\infty} (\hat{O}^\dagger \varphi_2)^* \cdot \varphi_1 dx = \int_{-\infty}^{\infty} \varphi_2^* \cdot (\hat{O} \varphi_1) dx \quad (5.1.21b)$$

Some operators are self-adjoint, which means that they are equal to their adjoint: $\hat{O}^\dagger = \hat{O}$. Self-adjoint operators are also called Hermitian⁷.

Example 5.2. Determine the adjoint of the operators \hat{p} , \hat{x} and \hat{H} ! **Solution a) adjoint of the momentum operator**

According to the definition of the adjoint operator:

$$\begin{aligned} \langle \hat{p}^\dagger \varphi_2 | \varphi_1 \rangle &= \langle \varphi_2 | \hat{p} \varphi_1 \rangle \quad \text{i.e.} \\ \langle \hat{p}^\dagger \varphi_2 | \varphi_1 \rangle &= \langle \varphi_2 | \frac{\hbar}{i} \frac{d\varphi_1}{dx} \rangle \\ \int_{-\infty}^{\infty} (\hat{p}^\dagger \varphi_2)^* \cdot \varphi_1 dx &= \int_{-\infty}^{\infty} \varphi_2^* \cdot (\hat{p} \varphi_1) dx \quad \text{or} \\ \int_{-\infty}^{\infty} (\hat{p}^\dagger \varphi_2)^* \cdot \varphi_1 dx &= \int_{-\infty}^{\infty} \varphi_2^* \cdot \frac{\hbar}{i} \frac{d\varphi_1}{dx} dx \end{aligned}$$

The right hand side can be calculated with integration by parts:

$$\int_{-\infty}^{\infty} \varphi_2^* \cdot \frac{\hbar}{i} \frac{d\varphi_1}{dx} dx = \frac{\hbar}{i} [\varphi_2^* \cdot \varphi_1]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\hbar}{i} \frac{d\varphi_2^*}{dx} \cdot \varphi_1 dx$$

⁷For the sake of completeness: all self-adjoint operators are Hermitian, but not all Hermitian operators are self-adjoint, but for our purposes the two notions are equivalent.

Because both φ_1 and φ_2 are physical wave functions they must be square integrable, therefore they must vanish when $x \rightarrow \infty$, so the first term is zero

$$\int_{-\infty}^{\infty} (\hat{p}^\dagger \varphi_2)^* \cdot \varphi_1 dx = - \int_{-\infty}^{\infty} \frac{\hbar}{i} \frac{d\varphi_2^*}{dx} \cdot \varphi_1 dx$$

Because $(p^\dagger \varphi)^* = (p^\dagger)^* \varphi^* = \left(-\frac{\hbar}{i} \frac{d}{dx}\right)^* \varphi^*$:

$$\hat{p}^\dagger \equiv \frac{\hbar}{i} \frac{d}{dx} = \hat{p}$$

The momentum operator is *self-adjoint*.

b) adjoint of the position operator

This is much simpler, because $\hat{x} \equiv x \cdot$ is a multiplication with a real number (or vector in 3 dimensions) and it commutes with the wave functions, therefore

$$\hat{x}^\dagger = \hat{x}$$

The position operator is self adjoint too.

c) the Hamiltonian

The Hamiltonian is a linear combination of the operators $\hat{p}^2 = -\hbar^2 \frac{d^2}{dx^2}$ and $V(x) = V(x) \cdot$. It is easy to prove that the product and sum of self-adjoint operators is also a self-adjoint operator.

Because the operator of the potential is a multiplication with a function it is self-adjoint, and $\hat{p}^2 = \hat{p} \hat{p}$ is a product of the self-adjoint \hat{p} with itself, the Hamiltonian is also self-adjoint:

$$\hat{H}^\dagger = \hat{H}$$

5.2 Measurement in quantum mechanics

In gaining information about any object in the world experimental physics characterizes objects by *observables*, quantities which can be measured in a physical experiment. One of the tasks of physics is to predict the result of such a measurement. We have already seen that the possible energy levels of the system can be calculated using the stationary Schrödinger equation. These energy levels are the eigenvalues of the Schrödinger equation. We also know that these energy levels correspond to stationary states of the system, which are eigenstates of the Schrödinger equation and that *those are the only energy states we can observe*.

Important 5.2.1. For any operator \hat{O} the equation

$$\hat{O} \varphi_\lambda = \lambda \varphi_\lambda \quad (5.2.1)$$

is called an eigenvalue equation, φ_λ is the eigenfunction for the eigenvalue λ .

The stationary Schrödinger equation is an eigenvalue problem of the Hamiltonian of the system. Therefore the possible stationary states and energy values are the eigenstates and eigenvalues of the Hamiltonian respectively. The eigenfunctions are orthogonal in the sense that

$$\langle \varphi_j | \varphi_k \rangle = \text{const} \cdot \delta_{jk}$$

Any possible state of the system can be written as a linear combination of the possible eigenfunctions of \hat{H} . Because all observables are represented in quantum mechanics by an operator, we can write separate eigenvalue equations to any of them.

Example 5.3. Determine the eigenfunctions and eigenvalues for the 3D momentum operator! **Solution**

$$\begin{aligned} \hat{\mathbf{p}} \varphi_{\mathbf{p}}(\mathbf{r}) &= \mathbf{p} \varphi_{\mathbf{p}}(\mathbf{r}) \\ \frac{\hbar}{i} \nabla \varphi_{\mathbf{p}}(x, y, z) &= \mathbf{p} \varphi_{\mathbf{p}}(x, y, z) \\ \frac{\hbar}{i} \left(\frac{\partial \varphi_{\mathbf{p}}(x, y, z)}{\partial x}, \frac{\partial \varphi_{\mathbf{p}}(x, y, z)}{\partial y}, \frac{\partial \varphi_{\mathbf{p}}(x, y, z)}{\partial z} \right) &= (p_x, p_y, p_z) \varphi_{\mathbf{p}}(x, y, z) \\ \varphi_{\mathbf{p}}(x, y, z) &= e^{i(p_x x + p_y y + p_z z)/\hbar} = e^{i \mathbf{p} \cdot \mathbf{r}/\hbar} \end{aligned}$$

i.e. the eigenfunctions of the $\hat{\mathbf{p}}$ operator are plane waves with eigenvalues corresponding to a continuous set of exact momenta. Because these functions are not quadratically integrable, they can not describe any physical state of the system separately. As we saw (Section 3.3) we must use wave packets created as a linear combination of an infinite number of these eigenstates (\Rightarrow Fourier transformation.) to describe a physical state.

If the eigenfunctions of an operator of an observable \hat{O} are normalized, i.e. $\langle \varphi_o | \varphi_o \rangle = 1$ then the eigenvalue can be determined by multiplying the

$$\hat{O} \varphi_o = \lambda \varphi_o$$

eigenvalue equation from the left by φ_o^* and integrating it for the whole space. In 1D⁸:

$$\int_{-\infty}^{\infty} \varphi_o^* (\hat{O} \varphi_o) dx = \int_{-\infty}^{\infty} \varphi_o^* \lambda \varphi_o = \lambda \int_{-\infty}^{\infty} \varphi_o^* \varphi_o dx = \lambda,$$

or

$$\lambda = \int_{-\infty}^{\infty} \varphi_o^* (\hat{O} \varphi_o) dx \quad (5.2.2)$$

According to the definition of the adjoint operator \hat{O}^\dagger

$$\begin{aligned} \int_{-\infty}^{\infty} \varphi_o^* (\hat{O} \varphi_o) dx &= \int_{-\infty}^{\infty} (\hat{O}^\dagger \varphi_o)^* \varphi dx \\ \int_{-\infty}^{\infty} \varphi_o^* (\lambda \varphi_o) dx &= \int_{-\infty}^{\infty} (\lambda^\dagger \varphi_o)^* \varphi dx \\ \lambda \int_{-\infty}^{\infty} \varphi_o^* \varphi_o dx &= \lambda^{\dagger*} \int_{-\infty}^{\infty} \varphi_o^* \varphi dx \\ \lambda &= \lambda^{\dagger*} \end{aligned}$$

where λ^\dagger is the eigenvalue of the adjoint operator. Because an observable is a measurable physical quantity and the result of all measurements should be real and not just complex, λ is a real number. And the same must be true for λ^\dagger . This means that all eigenfunctions and eigenvalues of the operator \hat{O} of an observable are the same as those of its adjoint \hat{O}^\dagger , which means that:

Important 5.2.2. *Operators of observables must be self-adjoint.*

Let us suppose that φ_n is an eigenfunction of the 1D Hamiltonian \hat{H} with the eigenvalue \mathcal{E}_n , where the n *quantum number* can go over all positive integer numbers:

$$\hat{H} \varphi_n = \mathcal{E}_n \varphi_n \quad n = 1, 2, \dots \quad (5.2.3)$$

⁸In the shorthand notation

$$\begin{aligned} \langle \varphi_o | \hat{O} \varphi_o \rangle &= \langle \varphi_o | \lambda \varphi_o \rangle = \lambda \langle \varphi_o | \varphi_o \rangle = \lambda \\ \lambda &= \langle \varphi_o | \hat{O} \varphi_o \rangle \end{aligned}$$

Any possible wave functions of the system can be expressed as a linear combination of eigenfunctions:

$$\varphi = C_1 \varphi_1 + C_2 \varphi_2 + \dots = \sum_{n=1} C_n \varphi_n \quad (5.2.4)$$

where for normalized wave functions the sum of the absolute square of the C_n coefficients must be 1:

$$\sum_n |C_n|^2 = 1 \quad (5.2.5)$$

Now apply the Hamiltonian to this wave function:

$$\hat{H} \varphi = \hat{H} \sum_{n=1} C_n \varphi_n \quad (5.2.6)$$

because the \hat{H} operator is linear:

$$\hat{H} \varphi = \hat{H} \sum_{n=1} C_n \varphi_n = \sum_{n=1} C_n \hat{H} \varphi_n = \sum_{n=1} C_n \mathcal{E}_n \varphi_n \quad (5.2.7)$$

The state of the system after the measurement of its energy is one of the possible stationary states of the system and that the value we measure is one of the eigenvalues. Calculate the probability that *after* the measurement we find that the energy of the particle is \mathcal{E}_m and its wave function is φ_m ? Multiply (5.2.7) with φ^* and integrate for the whole of space:

$$\begin{aligned} \mathcal{P}_{\uparrow} &= \int_{-\infty}^{\infty} \varphi^* (\hat{H} \varphi) dx = \int_{-\infty}^{\infty} \varphi^* \hat{H} \varphi dx = \\ &= \int_{-\infty}^{\infty} \left(\sum_{m=1} C_m^* \varphi_m \right) \hat{H} \left(\sum_{n=1} C_n \mathcal{E}_n \varphi_n \right) dx = \\ &= \sum_{n,m=1} C_m^* C_n \mathcal{E}_n \underbrace{\left(\int_{-\infty}^{\infty} \varphi_m^* \varphi_n dx \right)}_{=\delta_{nm}} = \\ &= \sum_{n=1} C_n^* C_n \mathcal{E}_n = \sum_{n=1} |C_n|^2 \mathcal{E}_n \end{aligned} \quad (5.2.8)$$

We see that the $|C_n|^2$ coefficients are the probabilities that after a measurement the wave function will be the n -th eigenfunction of \hat{H} and that the measured value is \mathcal{E}_n . We may generalize this result.

Important 5.2.3. Let $\varphi_n^{(O)}$ denote the set of eigenfunctions of an observable \hat{O} . Any state of the system can be written as a linear combination of these eigenfunctions. If the state of the system before a measurement is :

$$|\varphi\rangle = \sum_n C_n \varphi_n^{(O)}$$

then as the result of a measurement it will reduce to one of the possible eigenfunctions for that observable with a probability of $|C_n|^2$. The measured value then is the corresponding eigenvalue. It is impossible to tell exactly beforehand which one of these eigenvalues will be measured. The expectation value (average) of the measured observable can be calculated by

$$\langle O \rangle = \frac{\langle \varphi | \hat{O} \varphi \rangle}{\langle \varphi | \varphi \rangle} = \frac{\int_{-\infty}^{\infty} \varphi^* (\hat{O} \varphi) dx}{\int_{-\infty}^{\infty} \varphi^* \varphi dx} \quad (5.2.9)$$

Mathematically a measurement is represented by this formula.

As we discussed if the commutator of two observables is not 0 then there is an uncertainty relation between the two observables. This means that the two observables can not be measured simultaneously with arbitrary accuracy, therefore the two operator can not have the same set of eigenfunctions⁹.

Important 5.2.4. Two observables can have the same set of eigenfunctions and both can be measured with any accuracy (no uncertainty relation between them), if, and only if their commutator is 0.

For instance the commutator of the Hamiltonian and the angular momentum in a 3D centrally symmetric potential (e.g in the hydrogen atom) is 0 (see Chapter 6):

$$[\hat{H}, \hat{\mathbf{L}}] = 0$$

therefore they have a *joint system of eigenfunctions*. The eigenfunction system of \hat{H} is said to be *degenerate*, because more than one different eigenfunctions have the same energy, but a different angular momentum for the centrally symmetric potential. The eigenfunctions of \hat{H} and $\hat{\mathbf{L}}$ for a bound state are discrete and can be labeled by the index of the corresponding eigenvalues of both operators, in this case with n for the energy values, l for the maximum of the z-component of the angular momentum and m for the actual z component. These are 3 *quantum numbers* (see Chapter 6):.

⁹Let \hat{A} and \hat{B} two observables with a non zero commutator: $[\hat{A}, \hat{B}] = \hat{C} \neq 0$. For any φ function then $[\hat{A}, \hat{B}] \varphi = \hat{C} \varphi$ which is never zero because the only operator that results in 0 when applied to any function is the $\hat{0}$ operator. However if φ was an eigenfunction of both operators, with eigenvalues λ_A and λ_B then $[\hat{A}, \hat{B}] \varphi = \hat{A} \hat{B} \varphi - \hat{B} \hat{A} \varphi = \lambda_B \lambda_A \varphi - \lambda_A \lambda_B \varphi = 0$, which is a contradiction.

Chapter 6

Central potential. The hydrogen atom.

An atom consists of a positively charged nucleus that restricts the motion of the negatively charged electrons by the Coulomb force that is attractive and centrally (spherically) symmetric. Therefore motion of particles in centrally symmetric potentials is crucial understanding atoms.

6.1 Angular momentum.

Classically angular momentum is a conserved physical quantity for objects moving in centrally symmetric potentials. Using the general principles introduced above we determine the quantum mechanical operators for the three components of the angular momentum. Starting from the classical formulas:

$$\begin{aligned}\mathbf{L} &= \mathbf{r} \times \mathbf{p} \\ L_x &= y p_z - z p_y \\ L_y &= z p_x - x p_z \\ L_z &= x p_y - y p_x\end{aligned}$$

and substituting all quantities with the corresponding operators we get the quantum mechanical operators for the components of the angular momentum:

$$\begin{aligned}\hat{L}_x &= \hat{y} \hat{p}_z - \hat{z} \hat{p}_y = \frac{\hbar}{i} \left(y \frac{d}{dz} - z \frac{d}{dy} \right) \\ \hat{L}_y &= \hat{z} \hat{p}_x - \hat{x} \hat{p}_z = \frac{\hbar}{i} \left(z \frac{d}{dx} - x \frac{d}{dz} \right) \\ \hat{L}_z &= \hat{x} \hat{p}_y - \hat{y} \hat{p}_x = \frac{\hbar}{i} \left(x \frac{d}{dy} - y \frac{d}{dx} \right)\end{aligned}$$

Example 6.1. Determine whether there exists an uncertainty formula for the different components of the angular momentum operator. **Solution** An uncertainty formula between two physical quantities exists only if their commutator is not 0. Let us calculate $[\hat{L}_x, \hat{L}_y]$! This requires simple algebra and not higher mathematics. We do not even have to know the concrete form of the operators, because their commutators show exactly how their products can be rearranged.

$$\begin{aligned}
[\hat{L}_x, \hat{L}_y] &= \hat{L}_x \hat{L}_y - \hat{L}_y \hat{L}_x = \\
&= (\hat{y} \hat{p}_z - \hat{z} \hat{p}_y) (\hat{z} \hat{p}_x - \hat{x} \hat{p}_z) - (\hat{z} \hat{p}_x - \hat{x} \hat{p}_z) (\hat{y} \hat{p}_z - \hat{z} \hat{p}_y) = \\
&= \underbrace{\hat{y} \hat{p}_z \hat{z} \hat{p}_x}_{\text{-----}} - \underbrace{\hat{y} \hat{p}_z \hat{x} \hat{p}_z}_{\text{=====}} - \underbrace{\hat{z} \hat{p}_y \hat{z} \hat{p}_x}_{\text{~~~~~}} + \underbrace{\hat{z} \hat{p}_y \hat{x} \hat{p}_z}_{\text{++++++}} - \\
&\quad - \underbrace{\hat{z} \hat{p}_x \hat{y} \hat{p}_z}_{\text{-----}} + \underbrace{\hat{z} \hat{p}_x \hat{z} \hat{p}_y}_{\text{~~~~~}} + \underbrace{\hat{x} \hat{p}_z \hat{y} \hat{p}_z}_{\text{=====}} - \underbrace{\hat{x} \hat{p}_z \hat{z} \hat{p}_y}_{\text{++++++}}
\end{aligned}$$

where the different “underlines” mark terms from which common factors may be pulled out, because some or all of the operators in them commute and therefore their order is not important. E.g. $\hat{y} \hat{p}_z \hat{z} \hat{p}_x \equiv \hat{y} \hat{p}_x \hat{p}_z \hat{z}$, because \hat{p}_x commutes with all other operators in this term. But the order of \hat{z} and \hat{p}_z is important as they do not commute.

$$\begin{aligned}
[\hat{L}_x, \hat{L}_y] &= (\hat{y} \hat{p}_x \hat{p}_z \hat{z} - \hat{y} \hat{p}_x \hat{z} \hat{p}_z) + (\hat{x} \hat{y} \hat{p}_z \hat{p}_z - \hat{x} \hat{y} \hat{p}_z \hat{p}_z) + \\
&\quad (\hat{z} \hat{z} \hat{p}_y \hat{p}_x) - \hat{z} \hat{z} \hat{p}_y \hat{p}_x (\hat{x} \hat{p}_y \hat{z} \hat{p}_z - \hat{x} \hat{p}_y \hat{p}_z \hat{z}) = \\
&\quad \underbrace{\hspace{10em}}_{\text{~~~~~}} \underbrace{\hspace{10em}}_{\text{+++++++}} = \\
&= \hat{y} \hat{p}_x (\hat{p}_z \hat{z} - \hat{z} \hat{p}_z) + 0 + 0 + \hat{x} \hat{p}_y (\hat{z} \hat{p}_z - \hat{p}_z \hat{z}) = \\
&= (\hat{x} \hat{p}_y - \hat{y} \hat{p}_x) (\hat{z} \hat{p}_z - \hat{p}_z \hat{z})
\end{aligned}$$

Because $(\hat{x} \hat{p}_y - \hat{y} \hat{p}_x) = \hat{L}_z$ and $(\hat{z} \hat{p}_z - \hat{p}_z \hat{z}) = [\hat{z}, \hat{p}_z] = i\hbar$

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z \quad (6.1.1a)$$

Similar formulas could be derived for the commutator of any two components:

$$[\hat{L}_y, \hat{L}_z] = i\hbar \hat{L}_x \quad (6.1.1b)$$

$$[\hat{L}_z, \hat{L}_x] = i\hbar \hat{L}_y \quad (6.1.1c)$$

Because their commutator is not zero, the different components of the angular momentum may not be determined with arbitrary accuracy simultaneously. There is an uncertainty relation between them.

This presents the most striking differences between the physics of the angular momentum in classical and quantum mechanics:

Important 6.1.1. *In quantum mechanics the length of the angular momentum vector is always larger than the maximum of any of its components and a non-zero angular momentum (e.g. in an atom) does not necessarily imply the acceleration of particles.*

Example 6.2. *Determine the eigenvalues and eigenfunctions of \hat{L}_z !* **Solution** This problem is best dealt with in a spherical polar coordinate system. The form of the \hat{L}_z operator in spherical polar coordinates is (see Appendix 22.9)

$$\hat{L}_z = \frac{\hbar}{i} \frac{\partial}{\partial \phi} \quad (6.1.2)$$

In such a system the form of the eigenvalue equation of \hat{L}_z becomes

$$\frac{\hbar}{i} \frac{d\varphi}{d\phi} = \lambda \varphi \quad \Rightarrow \quad \varphi(\phi) = C e^{\frac{i}{\hbar} \lambda \phi},$$

where the C normalization constant is determined from the equation

$$\begin{aligned} \int_0^{2\pi} |\varphi(\phi)|^2 d\phi &= 1 \\ \int_0^{2\pi} |C|^2 d\phi &= 2\pi |C|^2 = 1 \\ C &= \frac{1}{\sqrt{2\pi}} \end{aligned}$$

and because φ is periodic in ϕ :

$$\begin{aligned} \varphi(\phi + 2\pi) &= \varphi(\phi) \\ e^{\frac{i}{\hbar} \lambda 2\pi} &= 1 \\ \frac{\lambda}{\hbar} &= m, \quad \text{where } m \text{ is an integer, i.e.} \\ \lambda &= m \hbar, \quad m = 0, \pm 1, \pm 2, \pm 3, \dots \end{aligned}$$

i.e. the z component of the angular momentum is quantized, its value is an integer multiple of \hbar and it can only change in \hbar increments.

It is easy to prove that the magnitude (or the square of the magnitude) of the angular momentum however commute with any of its components, therefore with \hat{L}_z too. Consequently both \hat{L}_z and \hat{L}^2 can be measured simultaneously with any accuracy required.

The operator of \hat{L}^2 in spherical polar coordinates (see 22.9.7):

$$\hat{L}^2 = -\hbar^2 \left[+ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right] \quad (6.1.3)$$

The eigenfunctions of this equation are usually denoted by $Y(\theta, \phi)$ and are called *spherical harmonics*. The eigenvalue equation is:

$$\hat{L}^2 Y(\theta, \phi) = L^2 Y(\theta, \phi) \quad (6.1.4)$$

$$-\hbar^2 \left[+\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right] Y_\ell^m(\theta, \phi) = L^2 Y_\ell^m(\theta, \phi) \quad (6.1.5)$$

where ℓ and m are two quantum numbers, ℓ equals to the maximum possible value of L_z/\hbar and is an integer¹ and m is

$$m = 0, \pm 1, \pm 2, \dots \pm \ell \quad (6.1.6)$$

$Y_\ell^m(\theta, \phi)$ is called a spherical harmonic function of degree ℓ and order m .

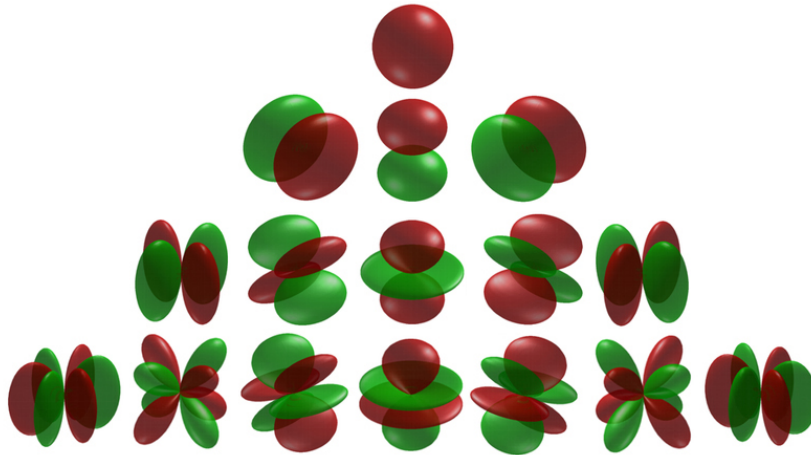


Figure 6.1: Visual representation of the first few spherical harmonics. Red denotes regions where Y_ℓ^m is positive, green where it is negative. ℓ is 0 at the top and increased by one in every row below it.

The eigenvalue L^2 in (6.1.4) is

$$L^2 = \ell(\ell + 1)\hbar^2 \quad (6.1.7)$$

This formula and (6.1.6) together state that the angle of the angular momentum with the z-axis can only take discrete values.

¹At least for the orbital angular momentum. We will discuss the *spin* momentum later in Section 6.3

The formula for the calculation of the spherical harmonics is

$$Y_\ell^m(\theta, \phi) = \sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}} P_\ell^m(\cos\theta) e^{im\phi} \quad (6.1.8)$$

where $P_\ell^m(x)$ are the *associated Legendre polynomials*².

6.2 The hydrogen atom.

Before the advent of quantum physics people imagined atoms as miniature solar systems with the positively charged heavy nucleus in the center and negatively charged electrons orbiting around it, bound by the electric Coulomb field of the nucleus. The simplest atom in the Universe is the hydrogen atom. It only contains a positively charged proton as a nucleus and a single negatively charged electron. The proton is 1836 times as heavy as the electron, so according to the classical model the proton does not move perceptibly while the electron is orbiting it³. Because an accelerating charge loses energy by radiation this is not a classically consistent model. For a little generalization let us consider *hydrogen like ions*, i.e. atoms with Z protons in their nucleus ionized until only a single electron remains.

Example 6.3. *After the discovery of de Broglie that electrons may have wave like properties the danish physicist Niels Bohr proposed a simple model for the hydrogen atoms which we now call the Bohr model. He proposed that the electron will not emit radiation if it moves on such (classical) circular orbits where the circumference is an integer multiple of the wavelength of the electron (c.f. 3.1):*

$$2r\pi = n\lambda_e = n\frac{h}{p}, \quad \text{where } n = 1, 2, \dots$$

²Associated Legendre polynomials are the solution of the differential equation

$$(1-x^2)y'' - 2xy' + (l(l+1) - \frac{m^2}{1-x^2})y = 0;$$

The first few associated Legendre polynomials are

$$P_0^0(x) = 1, P_1^0(x) = x, P_1^1(x) = -\sqrt{1-x^2},$$

³We could get rid of this assumptions by using the *reduced mass* of this 2-body problem instead of the electron mass m_e :

$$m = \frac{m_e m_N}{m_e + m_N},$$

where m_N is the mass of the nucleus. For hydrogen $m = 0.99945 m_e$.

This is the same as the statement that the orbital angular momentum of the electron is quantized, as with simple rearrangement

$$r p = \frac{h}{2\pi} n = n \hbar \quad (6.2.1)$$

and in our case the momentum vector is perpendicular to the radius vector, therefore $|\mathbf{L}| = |\mathbf{r} \times \mathbf{p}| = r p = r m_e v$. This semiclassical result is different from (6.1.6) (not known to Bohr at his time).

The electron is held in orbit by the Coulomb force, which provides the necessary centripetal force:

$$\frac{1}{4\pi\epsilon_o} \frac{Z e^2}{r^2} = \frac{m_e v^2}{r}. \quad (6.2.2)$$

This gives the total energy of the hydrogen atom to be half of its potential energy

$$\mathcal{E}_{tot} = \frac{1}{2} m_e v^2 - \frac{1}{4\pi\epsilon_o} \frac{Z e^2}{r} = -\frac{1}{8\pi\epsilon_o} \frac{Z e^2}{r} \quad (6.2.3)$$

From (6.2.1) and (6.2.2)

$$\frac{n^2 \hbar^2}{m_e} = \frac{1}{4\pi\epsilon_o} Z e^2 r \quad \text{therefore} \quad (6.2.4)$$

i.e. the electron can orbit only at discrete radii r_n .

$$r_n = \frac{(4\pi\epsilon_o) \hbar^2}{m_e Z e^2} n^2 \quad (6.2.5)$$

The radius of the first orbit r_1 in a hydrogen atom ($Z = 1$) is denoted by a_o

$$a_o = \frac{4\pi\epsilon_o \hbar^2}{m_e e^2} \approx 0.0529 \text{ nm} \quad (6.2.6)$$

and is called the Bohr radius⁴. For a hydrogen like atom:

$$r_n = a_o \frac{n^2}{Z}$$

⁴Sometimes a_o is written as

$$a_o = \frac{\hbar}{m_e c \alpha},$$

where c is the speed of light in vacuum and α is called the *fine structure constant*, introduced by Arnold Sommerfeld in 1916, which is the coupling constant characterizing the strength of the electromagnetic interaction. Being a dimensionless quantity, it has the same numerical value in all systems of units.

The current recommended value of α is $7.2973525698(24) \cdot 10^{-3} = 1/137.035999074(44)$

The total energy then

$$\mathcal{E}_{tot,n} = -\frac{1}{4\pi\epsilon_o} \frac{Ze^2}{2r_n} = -\left[\frac{(Ze^2)^2 m_e}{2(4\pi\epsilon_o)^2 \hbar^2} \right] \frac{1}{n^2}, \text{ or} \quad (6.2.7)$$

$$\mathcal{E}_{tot,n} = -\frac{Z^2}{2a_o^2 m_e} \frac{1}{n^2} = -R_{\mathcal{E}} \frac{1}{n^2}, \quad (6.2.8)$$

where $R_{\mathcal{E}} \approx 13.6 \text{ eV}$ is called the Rydberg constant or the Rydberg unit of energy and the index n denotes the n -th electron orbit.

In quantum mechanics to get the possible energy levels and wave functions of the electron in the H atom we must solve the 3 dimensional time independent Schrödinger equation (3.5.4) for the centrally symmetric Coulomb potential of the proton:

$$\begin{aligned} \hat{H} \varphi(x, y, z) &= \mathcal{E} \varphi(x, y, z) \\ -\frac{\hbar^2}{2m_e} \nabla^2 \varphi + V \varphi &= \mathcal{E} \varphi \\ -\frac{\hbar^2}{2m_e} \left(\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} \right) + V(x, y, z) \varphi(x, y, x) &= \mathcal{E} \varphi \end{aligned}$$

where

$$V(x, y, z) = -\frac{1}{4\pi\epsilon_o} \frac{Ze^2}{r} \quad (\text{where } r = \sqrt{x^2 + y^2 + z^2}) \quad (6.2.9)$$

Because this potential is spherically symmetric we will be better off if we write this equation in spherical coordinates (see (5.1.19), or (22.9.6)):

$$\begin{aligned} \hat{H} \varphi(r, \theta, \phi) &= \mathcal{E} \varphi(r, \theta, \phi) \\ -\frac{\hbar^2}{2m_e} \left[\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \varphi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \varphi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \varphi}{\partial \phi^2} \right] + \\ &+ V(r) \frac{e^2}{r} \varphi = \mathcal{E} \varphi \end{aligned} \quad (6.2.10)$$

Note that the sum of the 2^{nd} and 3^{rd} terms in the square brackets is equal to $\frac{1}{r^2} \hat{L}^2 \varphi$, and remember, the eigenfunctions and eigenvalues of \hat{L}^2 we already know. Therefore we can separate the radial and angular dependence of φ . Let $\varphi(r, \theta, \phi) = R(r) Y_{\ell}^m(\theta, \phi)$, where $Y_{\ell}^m(\theta, \phi)$ is a spherical harmonics, the eigenfunction of \hat{L}^2 , with the eigenvalue $\ell(\ell+1) \hbar^2$. With this

$$-\frac{\hbar^2}{2m_e} \left[\frac{d^2 R}{dr^2} Y_{\ell}^m + \frac{2}{r} \frac{dR}{dr} Y_{\ell}^m + R \frac{\hat{L}^2}{\hbar^2 r^2} Y_{\ell}^m \right] + V(r) \frac{e^2}{r} R Y_{\ell}^m = \mathcal{E} R Y_{\ell}^m \quad (6.2.11)$$

Now multiply both sides with r^2 and divide them with $\varphi = R Y_\ell^m$ then reorder the terms containing only the radial and only the angular part on the opposite sides of the equal sign:

$$-\frac{\hbar^2}{2m_e} \frac{r^2}{R} \left[\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} \right] + V(r) e^2 r - r^2 \mathcal{E} = \frac{1}{2m_e} \frac{1}{Y_\ell^m} \hat{L}^2 Y_\ell^m \quad (6.2.12)$$

Because Y_ℓ^m is the eigenfunction of \hat{L}^2 with the eigenvalue $\ell(\ell+1)\hbar^2$ the right hand side becomes

$$\frac{\hbar^2}{2m_e} \ell(\ell+1)$$

therefore the equation for the radial part of the wave function is

$$-\frac{\hbar^2}{2m_e} \left[\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} - \frac{\ell(\ell+1)}{r^2} R \right] + V(r) R - \mathcal{E} R = 0 \quad (6.2.13)$$

Note that this is “only” an ordinary second order differential equation, which nevertheless contains the unknown eigenvalue \mathcal{E} , i.e this is also an eigenvalue equation. A mathematical trick leads to a much simpler form. Let us introduce a new function $u(r)$ with the formula: $R(r) = \frac{u(r)}{r}$. After some easy mathematical steps we get:

$$-\frac{\hbar^2}{2m_e} \frac{d^2 u(r)}{dr^2} + \left[V(r) + \frac{\hbar^2 \ell(\ell+1)}{2m_e r^2} \right] u(r) = \mathcal{E} u(r) \quad (6.2.14)$$

which is a 1D Schrödinger equation with the *effective potential*

$$V_{eff}(r) = V(r) + \frac{\hbar^2 \ell(\ell+1)}{2m_e r^2} = \frac{\hbar^2 \ell(\ell+1)}{m_e r^2} - \frac{1}{4\pi\epsilon_o} \frac{Z e^2}{r}$$

The non-Coulomb part of this is sometimes called the *centrifugal potential*. This potential is repulsive. The resulting effective potential have a minimum where $\frac{\partial V_{eff}}{\partial r} = 0$.

$$r_{min} = \frac{4\pi\epsilon_o \hbar^2 \ell(\ell+1)}{m_e Z e^2} = \frac{1}{Z} a_o \ell(\ell+1) \quad \text{for } l > 0 \quad (6.2.15)$$

(where a_o is defined in (6.2.6)).

Even after such simplifications this equation is hard to solve and we will not attempt to do it here. For our purposes it is sufficient to show the eigenvalues and eigenfunctions.

The eigenfunctions can be characterized by 2 *quantum numbers*: a positive integer $n = 1, 2, \dots$, which determines the energy and ℓ which describes what spherical harmonics

n	ℓ	m	$R_{n\ell}$	$Y_{\ell m}$	$\psi_{n\ell m} = R_{n\ell} Y_{\ell m}$
1	0	0	$2 \left(\frac{1}{a_0}\right)^{3/2} e^{-r/a_0}$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{\sqrt{\pi}} \left(\frac{1}{a_0}\right)^{3/2} e^{-r/a_0}$
2	0	0	$\left(\frac{1}{2a_0}\right)^{3/2} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{4\sqrt{2\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$
2	1	0	$\left(\frac{1}{2a_0}\right)^{3/2} \frac{1}{\sqrt{3}} \frac{r}{a_0} e^{-r/2a_0}$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \cos \theta$	$\frac{1}{4\sqrt{2\pi}} \left(\frac{1}{a_0}\right)^{3/2} \frac{r}{a_0} e^{-r/2a_0} \cos \theta$
2	1	± 1	$\left(\frac{1}{2a_0}\right)^{3/2} \frac{1}{\sqrt{3}} \frac{r}{a_0} e^{-r/2a_0}$	$\pm \frac{1}{2} \sqrt{\frac{3}{\pi}} \sin \theta e^{\pm i\varphi}$	$\frac{1}{8} \sqrt{\frac{1}{\pi}} \left(\frac{1}{a_0}\right)^{3/2} \frac{r}{a_0} e^{-r/2a_0} \sin \theta e^{\pm i\varphi}$
3	0	0	$2 \left(\frac{1}{3a_0}\right)^{3/2} \left(1 - \frac{2}{3} \frac{r}{a_0} + \frac{2}{27} (r/a_0)^2\right) e^{-r/3a_0}$	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{81\sqrt{3\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(27 - 18 \frac{r}{a_0} + 2(r/a_0)^2\right) e^{-r/3a_0}$
3	1	0	$\left(\frac{1}{3a_0}\right)^{3/2} \frac{4\sqrt{2}}{3} \left(1 - \frac{1}{6} \frac{r}{a_0}\right) \frac{r}{a_0} e^{-r/3a_0}$	$\frac{1}{2} \sqrt{\frac{3}{\pi}} \cos \theta$	$\frac{1}{81\sqrt{2\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(6 - \frac{r}{a_0}\right) \frac{r}{a_0} e^{-r/3a_0} \cos \theta$
3	1	± 1	$\left(\frac{1}{3a_0}\right)^{3/2} \frac{4\sqrt{2}}{3} \left(1 - \frac{1}{6} \frac{r}{a_0}\right) \frac{r}{a_0} e^{-r/3a_0}$	$\pm \frac{1}{2} \sqrt{\frac{3}{\pi}} \sin \theta e^{\pm i\varphi}$	$\frac{1}{81\sqrt{\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(6 - \frac{r}{a_0}\right) \frac{r}{a_0} e^{-r/3a_0} \sin \theta e^{\pm i\varphi}$
3	2	0	$\left(\frac{1}{3a_0}\right)^{3/2} \frac{2\sqrt{2}}{27\sqrt{5}} \left(\frac{r}{a_0}\right)^2 e^{-r/3a_0}$	$\frac{1}{4} \sqrt{\frac{5}{\pi}} (3 \cos^2 \theta - 1)$	$\frac{1}{81\sqrt{6\pi}} \left(\frac{1}{a_0}\right)^{3/2} \frac{r^2}{a_0^2} e^{-r/3a_0} (3 \cos^2 \theta - 1)$
3	2	± 1	$\left(\frac{1}{3a_0}\right)^{3/2} \frac{2\sqrt{2}}{27\sqrt{5}} \left(\frac{r}{a_0}\right)^2 e^{-r/3a_0}$	$\pm \frac{1}{2} \sqrt{\frac{15}{\pi}} \sin \theta \cos \theta e^{\pm i\varphi}$	$\frac{1}{81\sqrt{\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(\frac{r}{a_0}\right)^2 e^{-r/3a_0} \sin \theta \cos \theta e^{\pm i\varphi}$
3	2	± 2	$\left(\frac{1}{3a_0}\right)^{3/2} \frac{2\sqrt{2}}{27\sqrt{5}} \left(\frac{r}{a_0}\right)^2 e^{-r/3a_0}$	$\frac{1}{4} \sqrt{\frac{15}{\pi}} \sin^2 \theta e^{\pm 2i\varphi}$	$\frac{1}{162\sqrt{\pi}} \left(\frac{1}{a_0}\right)^{3/2} \left(\frac{r}{a_0}\right)^2 e^{-r/3a_0} \sin^2 \theta e^{\pm 2i\varphi}$

Figure 6.2: Radial and angular parts of the full wave functions and the total wave function for the first three energy level in a hydrogen atom ($Z=1$).

belong to this function⁵. In Fig. 6.2 we summarized the different wave functions for the first three energy levels ($n = 1, 2$ and 3) in a hydrogen atom.

⁵The form of the radial part is

$$R_{n,\ell}(r) = \sqrt{\left(\frac{2Z}{na_0}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!}} e^{-Zr/na_0} \left(\frac{2Zr}{na_0}\right)^\ell L_{n-\ell-1}^{2\ell+1}\left(\frac{2Zr}{na_0}\right), \quad (6.2.16)$$

where $L_n^{(k)}$ are the generalized (or associated) Laguerre polynomials. These can be defined as:

$$L_n^{(k)} = \frac{e^x x^{-k}}{n!} \frac{d^n}{dx^n} (e^{-x} x^{n+k}) \quad (6.2.17)$$

The first few generalized Laguerre polynomials are

$$L_0^{(k)}(x) = 1 \quad (6.2.18)$$

$$L_1^{(k)}(x) = -x + k + 1 \quad (6.2.19)$$

$$L_2^{(k)}(x) = \frac{x^2}{2} - (k+2)x + \frac{(k+2)(k+1)}{2} \quad (6.2.20)$$

$$L_3^{(k)}(x) = \frac{-x^3}{6} + \frac{(k+3)x^2}{2} - \frac{(k+2)(k+3)x}{2} + \frac{(k+1)(k+2)(k+3)}{6} \quad (6.2.21)$$

In quantum mechanics the classical notion of a trajectory or orbit does not apply⁶. What quantum mechanics have instead is called an *atomic orbital* in an atom and *molecular orbital* in a molecule.

Important 6.2.1. *An atomic or molecular orbital is a one electron wave function in an atom or a molecule respectively.*

If an atom has more than one electron than both the Hamiltonian and its eigenfunctions contain the coordinates of all electrons. In this case the eigenfunctions are usually written as a linear combination of atomic orbitals.

Fig. 6.3 shows the radial part of the wave function for $n = 1, 2$ and 3 .

Using the correct quantum mechanical calculation for a hydrogen like ion gives exactly the same energy levels as the Bohr model:

$$\mathcal{E}_n = - \left[\frac{(Z e^2)^2 m_e}{2 (4 \pi \epsilon_o)^2 \hbar^2} \right] \frac{1}{n^2} \quad (6.2.22)$$

But, in contrast with the Bohr model, in the ground state the angular momentum of the electron is 0, the electron is not orbiting the nucleus! Similar states exist for at every n , because ℓ can be 0. And even in cases where the total angular momentum is not zero the electron does not move on classical orbits.

Important 6.2.2. *The reason electrons in an atom or molecule do not emit electromagnetic radiation (except when exited from one stationary state to an other one) is that they do not move around the nucleus on classical orbits, therefore they do not accelerate in their stationary states.*

The solutions of the original 3D Schrödinger equation of the hydrogen atom are characterized by 3 *quantum numbers*:

- the *principal quantum number* $n = 1, 2, \dots$, that determines the energy level
- the (orbital) *azimuthal (or angular momentum) quantum number* $\ell = 0, 1, 2, \dots, n-1$, which corresponds to the length of the angular momentum (*sub-level* or *subshell*), and
- the *magnetic quantum number* $m = \pm 1, \pm 2, \dots, \pm \ell$, which determines the z-components of the angular momentum⁷.

⁶When chemists talk about the “orbit” of an electron in atoms or molecules they usually refer to the range in space where the electron can be found with an (arbitrarily chosen) 90% probability.

⁷There is a 4th quantum number, called *spin* which we will discuss later.

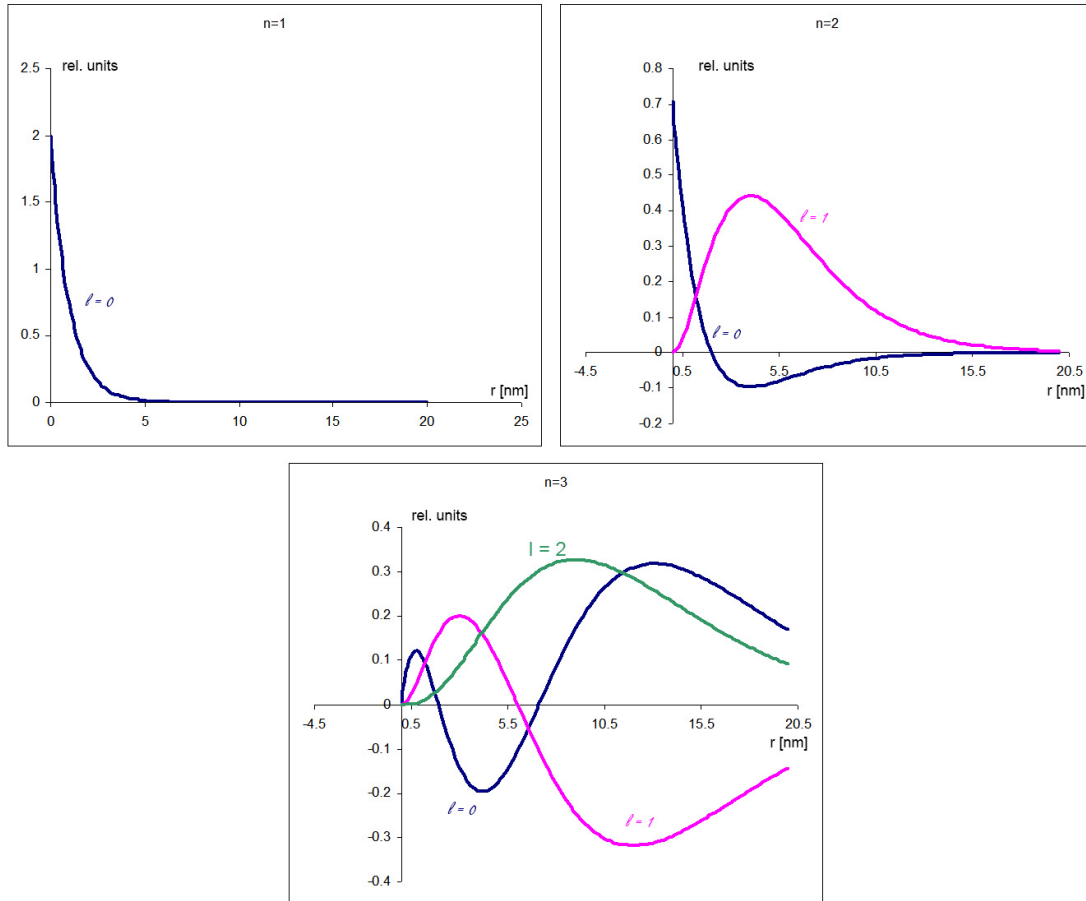


Figure 6.3: Radial part $R_{n,\ell}$ of the hydrogen wave function for the first three principal quantum numbers

Because all possible orbitals with the same principal quantum number have the same energy most of the energy levels are degenerate.

For historical reasons the states with different ℓ values have single letter names⁸ as well, as summarized in the following table.

ℓ	name
0	s
1	p
2	d
3	f
\vdots	\vdots

Table 6.1: Names for the different angular momentum states

The expressions $1s^2 2s^1 2p^5, \dots$ etc denotes the orbitals or *electron shells* where the number before the letter is the value of n , the letter determines the *subshell* (the value of ℓ) and the “exponent” is the number of electrons occupying that subshell in the ground state of the atom. The electron structure of an atom is then written as a series of these expressions. E.g. hydrogen has an electron structure of $1s^1$, helium with 2 electrons is $1s^2$ and argon with 10 electrons is $1s^2 2s^2 2p^6$.

As you can see the radial part of the ground state (the lowest energy) wave function

$$R_{1,0} = 2 \left(\frac{Z}{a_o} \right)^{3/2} e^{-Zr/a_o} \quad (6.2.23)$$

neither have a minimum nor a maximum around a_o/Z . The corresponding $Y_0^0(\theta, \phi) = \frac{1}{2\sqrt{\pi}}$ is constant. So not only $R_{1,0}$ does not have a minimum, but neither have $\varphi_{1,0,0} = R_{1,0}Y_0^{(0)}$. So what does the Bohr-radius a_o correspond to?

The probability that the electron is found in a dr range around the distance r from the nucleus is

$$\mathcal{P}_{n\ell m}(r) dr = \int_{\text{angular part}} \mathcal{P}_{n\ell m}(\mathbf{r}) dV$$

⁸Letters are abbreviations for “*sharp*”, “*principal*”, “*diffuse*” and “*fundamental*”. These are historical names used for the spectroscopic lines.

where $dV \equiv d^3r = r^2 \sin \theta d\phi d\theta dr$:

$$\begin{aligned}
\mathcal{P}_{n\ell m}(r) dr &= \left(\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \mathcal{P}_{n\ell m}(r, \theta, \phi) r^2 \sin \theta d\phi d\theta \right) dr = \\
&= \left(\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} |\varphi_{n\ell m}(r)|^2 r^2 \sin \theta d\phi d\theta \right) dr = \\
&= |R_{n\ell m}(r)|^2 r^2 dr \cdot \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} |Y_\ell^{(m)}|^2 \sin \theta d\phi d\theta
\end{aligned}$$

For the state with $\ell = 0$ (therefore $m = 0$ as well) this gives

$$\begin{aligned}
\mathcal{P}_{n\ell m}(r, dr) &= |R_{n\ell m}(r)|^2 r^2 dr \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{4\pi} \sin \theta d\phi d\theta = \\
&= |R_{n\ell m}(r)|^2 r^2 dr,
\end{aligned}$$

since the integral equals to $\frac{1}{4\pi} 4\pi = 1$. This, however has a maximum. For the ground state of a hydrogen like ion, substituting $\mathcal{P}_{1,0,0}$:

$$\mathcal{P}_{1,0,0} = r^2 \left| 2 \left(\frac{Z}{a_o} \right)^{3/2} e^{-Zr/a_o} \right|^2 = \frac{4Z^3}{a_o^3} r^2 e^{-2Zr/a_o}$$

The probability may have an extremum where $d\mathcal{P}_{1,0,0}/dr = 0$

$$\frac{d\mathcal{P}_{1,0,0}}{dr} = \frac{4Z^3}{a_o^3} \left[2r - \frac{2Z}{a_o} r^2 \right] e^{-2Zr/a_o} = 0,$$

which can only be zero if the expression in the square brackets are zero.

$$2r - \frac{2Z}{a_o} r^2 = 0 \quad \Rightarrow \quad 1 - \frac{Z}{a_o} r = 0 \quad \Rightarrow \quad r = \frac{a_o}{Z},$$

i.e. the maximum of the radial probability for an electron in the ground state of a hydrogen like ion is at a distance $r_{max\ probab} = a_o/Z$. For hydrogen $r_{max\ probab} = a_o$.

In Fig. 6.4 we depicted the spatial probability as a function of the distance for the 5 lowest orbitals.

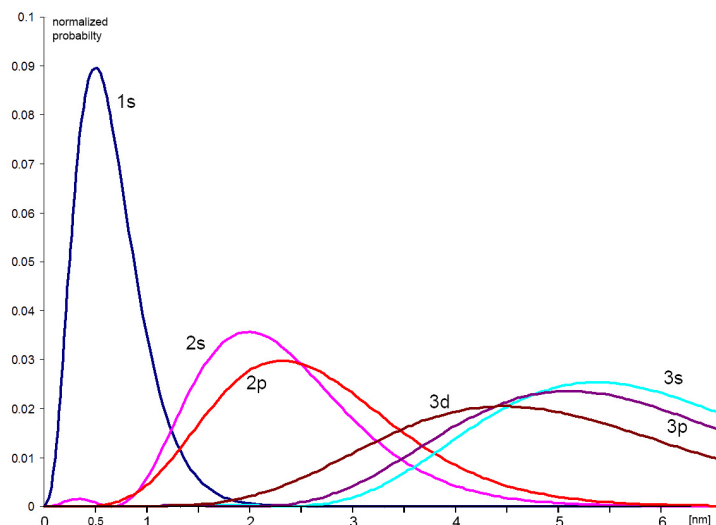


Figure 6.4: Radial probability functions in a hydrogen atom.

6.3 Electron spin

Before going into detailed quantum mechanical description of the hydrogen spectrum let us take a little detour to discuss the problem of the *spin*, especially the *electron spin*, the inherent internal angular momentum of the electron.

In 1922 two German physicists, Otto Stern and Walther Gerlach, performed an experiment to explain the *doublets* (see below) observed in atomic spectra, which lead to the discovery of a strictly quantum mechanical phenomena, something that has no counterpart in classical physics at all.

In the experiment illustrated in Fig. 6.5 a beam of neutral silver atoms is shot into a region of strong inhomogeneous magnetic field.

The electronic structure of silver is similar to that of H in that, that out of its 47 electrons 46 are “compensated” (their resulting angular momenta is 0, see section 6.7), one $5s^1$ electron is on an s orbit with zero angular momentum ($\ell = 0$). However according to classical physics (and the Bohr model) the $5s^1$ electron orbits around the nucleus, therefore even when the momentum of all other electrons are “compensated” the angular momentum of the outermost electron is not. In general when a charged particle, like the electron, has non-zero angular momentum it possesses a magnetic moment as well, therefore according to classical physics the silver atom must have a non-zero angular momentum, while quantum mechanics predicts that it does not have one. In the inhomogeneous magnetic field a beam of silver atoms (magnetic dipoles) is deflected, depending on the orientation of the magnetic moment. If the orientation of the magnetic moments are random and continuous (not discrete) the deflected particles will create a

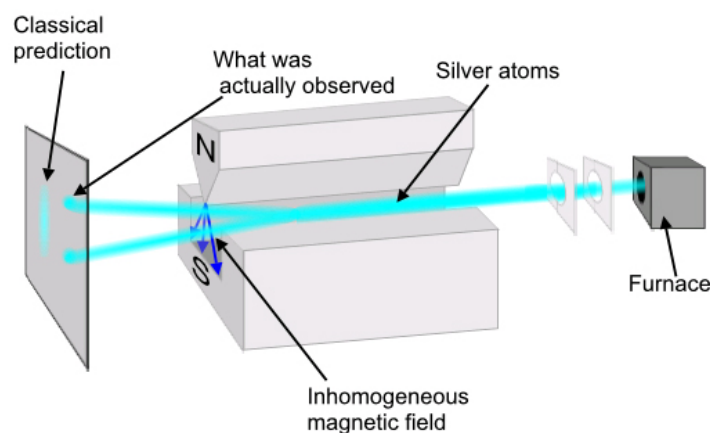


Figure 6.5: Schematics of the Stern-Gerlach experiment. Both the classical physical prediction and the observed behavior is shown.

continuous band on the screen.

Quantum mechanically no deflection is expected, resulting in a single spot on the screen.

However Stern and Gerlach observed two spots on the screen corresponding to 2 discrete deflection angles. But this means that for the Ag atom

$$2\ell + 1 = 2 \quad \Rightarrow \quad \ell = \frac{1}{2}$$

$$L_{electron} = \ell_{electron} \hbar = s \hbar = \frac{\hbar}{2}$$

This magnetic moment and angular momentum must belong to the electron because the total orbital angular momentum of the silver atoms is zero. This self angular momentum vector of the electron is called *electron spin*⁹ and denoted by \mathbf{S} . The corresponding spin operator is $\hat{\mathbf{S}}$ and the eigenvalues of its z component are $s\hbar$, where $s = \pm\frac{1}{2}$. Not just the electron, but all other particles have spins (which can be of different discrete values including 0).

Important 6.3.1. *Elementary particles possess an internal angular momentum called spin, which has no classical physical equivalent. Particles whose spin angular momentum is an odd multiple of $\frac{1}{2}\hbar$ (e.g. $s = \frac{1}{2}, \frac{3}{2}$, etc) are called half-integer spin particles. Other*

⁹If we had tried to explain the spin of an electron as a rotation of the particle, then the linear velocity of the circumference of the electron would exceed the speed of light in vacuum. The spin has no classical physical equivalent.

particles are called integer spin particles. Electrons are half-integer spin particles with a spin of $s = 1/2$, photons are integer spin particles with a spin of $s = 1$.

The magnetic moments $\boldsymbol{\mu}_L$ and $\boldsymbol{\mu}_S$ associated with the orbital and spin angular momentum \mathbf{L} and \mathbf{S} respectively are multiples of

$$\mu_B := \frac{e \hbar}{2 m_e} \quad (6.3.1)$$

the Bohr-magneton:

$$\boldsymbol{\mu}_L = -g_L \frac{\mu_B}{\hbar} \mathbf{L}, \text{ or} \quad (6.3.2a)$$

$$\boldsymbol{\mu}_S = -g_S \frac{\mu_B}{\hbar} \mathbf{S} \quad \text{where } g_S \equiv -g_e, \quad (6.3.2b)$$

g is called the gyromagnetic factor (or g-factor for short). The orbital g-factor, g_L is exactly 1. The electron g-factor, g_e is less than zero ($g_S = -g_e = |g_e|$) and is about -2 . The z-component of the magnetic moments are then

$$\mu_z^{(L)} = -g_L \mu_B m_\ell, \quad (6.3.3)$$

$$\mu_z^{(S)} = -g_S \mu_B m_s, \quad \text{where } m_s = \pm \frac{1}{2} \quad (6.3.4)$$

The electron g-factor $g_e \approx 2$ is one of the most precisely measured values in physics with its uncertainty beginning at the twelfth decimal place¹⁰.

The magnitude of the spin vector cannot be measured, it can only be calculated from the formula

$$|\mathbf{S}| = \sqrt{s(s+1)} \hbar = \frac{\sqrt{3}}{2} \hbar \quad (6.3.5)$$

where s is the maximum of the z-component of the spin in \hbar units.

The spin operator $\hat{\mathbf{S}}$ has very similar properties to the angular momentum operator. For instance it has a similar commutator between its elements (see:(6.1.1))

$$[\hat{S}_j, \hat{S}_k] = \epsilon_{jkn} \hat{S}_n. \quad (6.3.6)$$

The eigenvalue equations for the electron spin are analogous to the ones for the orbital angular momentum:

$$\hat{S}^2 \chi_{m_s} = \frac{3}{4} \hbar^2 \chi_{m_s} \quad (6.3.7)$$

$$\hat{S}_z \chi_{m_s} = m_s \hbar \chi_{m_s} \quad (6.3.8)$$

But unlike orbital angular momentum the χ eigenfunctions are not spherical harmonics, as there is no angle dependence in $\hat{\mathbf{S}}$.

¹⁰Its value is $g_e = -2.0023193043622$. This accuracy is surpassed by the Rydberg constant whose value is measured with 13 decimal digits accuracy

Important 6.3.2. *If we measure the spin along any direction in space we can still only get the values $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$, but calculations are simplest if we select the z -axis. Instead of writing fully the z -component of the spin, usually the \hbar is omitted and simply the m_s quantum number is used. The state where $m_s = \frac{1}{2}$ is called the up, the $m_s = -\frac{1}{2}$ the down state. The \uparrow and \downarrow symbols are also used for these.*

The complete wave function of the electron describing stationary states in a hydrogen atom then is written using 4 quantum numbers as

$$\varphi_{n,\ell,m,m_s}(r, \theta, \phi) = R(r)_{m\ell} Y_{\ell}^m(\theta, \phi) \chi_{m_s} \quad (6.3.9)$$

There is a third g -factor g_J in an atom which connects the total angular momentum ($\mathbf{J} \equiv \mathbf{L} + \mathbf{S}$) of the electron with its total magnetic moment. It is called the *Landé g -factor* and its formula is:

$$\boldsymbol{\mu} = -g_J \frac{\mu_B}{\hbar} \mathbf{J} \quad (6.3.10)$$

6.3.1 Addition of angular momenta

How can we calculate the possible values for the sum of two or more total angular momenta? Let

$$\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)}$$

where $\mathbf{J}^{(1)}$ and $\mathbf{J}^{(2)}$ can be any combination of orbital and spin momenta. Then the z -component of \mathbf{J} is

$$J_z = J_z^{(1)} + J_z^{(2)} \quad (6.3.11)$$

Because the z -component of $\mathbf{J}^{(1)}$ can only assume $(2j^{(1)} + 1)$ different discrete values between $-j^{(1)}$ and $+j^{(1)}$ and similar formulas are true for $j^{(2)}$ and j , to determine all possible values of J_z we must add together all possible values of $\mathbf{J}_z^{(1)}$ and $\mathbf{J}_z^{(2)}$. $\mathbf{J}_z^{(1)}$ can vary from $-(j^{(1)} + j^{(2)})$ to $+(j^{(1)} + j^{(2)})$ by steps of 1, so j , the maximum of the z component (a positive number) must be between the values $|j^{(1)} - j^{(2)}|$, for anti-parallel, and $(j^{(1)} + j^{(2)})$, for parallel $J^{(1)}$ and $J^{(2)}$ momenta. As an example in Table 6.2 we summarized the result for

$$\mathbf{J}^{(1)} \equiv \mathbf{L}, \quad \text{and} \quad (6.3.12)$$

$$\mathbf{J}^{(2)} \equiv \mathbf{S} \quad (6.3.13)$$

In this case

ℓ	0	1		2		3	
j	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{5}{2}$	$\frac{7}{2}$
	$s_{1/2}$	$p_{1/2}$	$p_{3/2}$	$d_{3/2}$	$d_{5/2}$	$f_{5/2}$	$f_{7/2}$

Table 6.2: Possible values for an atomic angular momentum. The third row contains the name of the given state.

$$j = \begin{cases} \ell \pm 1/2 & \ell > 0 \\ 1/2 & \ell = 0 \end{cases}$$

Please note that for $\ell = 0$ there is no “parallel” direction, therefore only $s = 1/2$ is possible.

6.4 Quantum mechanical analysis of the spectrum of the H atom. Spin-orbit coupling.

As we saw in Section 2.1 we can measure with a spectrometer the spectrum of absorption and emission of electromagnetic radiation that correspond to the transitions (according to Chapter 4) between stationary states characterized by the quantum numbers. Both the Bohr model and quantum mechanics gives the same formula for the energy of the stationary states. But the Bohr model needs special additional condition to ensure that electrons in stationary states do not emit electromagnetic radiation, while this condition is fulfilled automatically in quantum mechanics.

In the previous sections we learned that electron states are degenerate¹¹. Without the spin there is one possible state for $n = 1$, the 1s state $\varphi_{1,0,0}$. There are 4 possible degenerate states for $n = 2$: one 2s state $\varphi_{2,0,0}$ and three 2p states: $\varphi_{2,1,-1}$, $\varphi_{2,1,0}$ and $\varphi_{2,1,1}$. The angular part of the wave function for the first 3 values of ℓ is in Fig. 6.6. The surfaces correspond to $Y_\ell^{(m)} = \text{const}$.

Because for a given n there are n different values for ℓ ($\ell = 0, \dots, n-1$) and for every ℓ we have $2\ell + 1$ states for m ($m = -\ell, \dots, 0, \dots, \ell$) the total number of possible states (without the spin) is

$$N = \sum_{\ell=0}^{n-1} (2\ell + 1) = n + 2 \frac{(n-1)n}{2} = n^2 \quad (6.4.1)$$

If we take the spin into account too the number of the states must be multiplied by

¹¹All hydrogen states, including the φ_{100} state are degenerate when we take the spin quantum number into consideration.

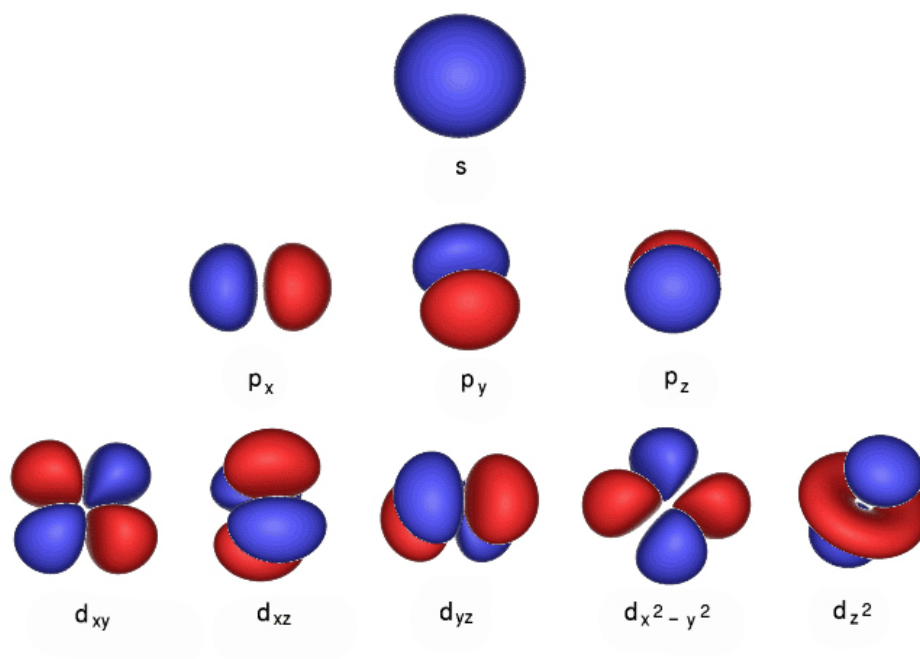


Figure 6.6: Angular part of the hydrogen wave function for the first 3 angular momenta (s,p and d orbitals)

2, which means that on the n -th shell there are $2n^2$ stationary states for electrons with different quantum numbers.

Electrons may emit radiation only when they move from one state to the other. In hydrogen the law of energy conservation states that during a transition from the state with principal quantum number n to a state with principal quantum number $m < n$ the electron emits a photon with energy equal to the energy difference of these two states:

$$h\nu = \mathcal{E}_n - \mathcal{E}_m = R_{\mathcal{E}} \left(\frac{1}{m^2} - \frac{1}{n^2} \right) \quad (6.4.2)$$

This is called the *Rydberg formula*. The hydrogen spectrum is non-continuous it features separate lines (non-zero intensities at discrete wavelengths Fig. 6.8). Each line corresponds to a possible transition.

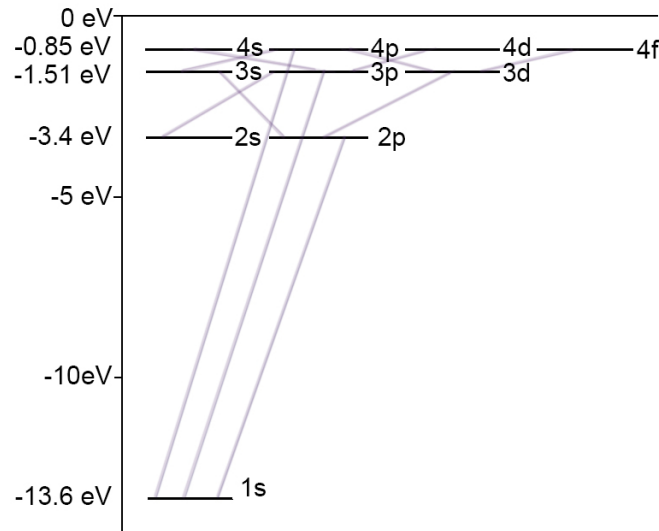


Figure 6.7: The first 4 energy levels of hydrogen and the possible transitions determined by the selection rule $\Delta \ell = \pm 1$. A single photon is emitted or absorbed in any possible transition.

In Fig. 6.7 the lowest lying 4 energy levels of the hydrogen atom are shown together with the possible transitions. There exists a selection rule that states the angular momentum difference between the original and resulting states must be¹² $\Delta \ell = \pm 1$ (i.e. $\Delta L = \pm \hbar$). The emission spectrum lines of hydrogen are classified in 6 *series* depending

¹²This condition corresponds to the law of the conservation of angular momentum, because the *spin* of the photon is 1.

on the principal quantum number of the state the emissive transition ends. The series of spectrum lines in the *visible* and in the *ultraviolet* part of the spectrum where the transition ends at the $m = 2$ level is called the *Balmer series*. Table 6.3 shows the data of the Balmer series¹³.

Trans.	$3 \rightarrow 2$	$4 \rightarrow 2$	$5 \rightarrow 2$	$6 \rightarrow 2$	$7 \rightarrow 2$	$8 \rightarrow 2$	$9 \rightarrow 2$	$\infty \rightarrow 2$
Name	$H - \alpha$	$H - \beta$	$H - \gamma$	$H - \delta$	$H - \epsilon$	$H - \zeta$	$H - \eta$	
$\lambda (nm)$	656.3	486.1	434.1	410.2	397.0	388.9	383.5	364.6
Color	Red	Cyan	Blue	Violet	UV	UV	UV	UV

Table 6.3: Some lines of the Balmer series

In Fig. 6.8 the measured hydrogen spectrum lines are shown.

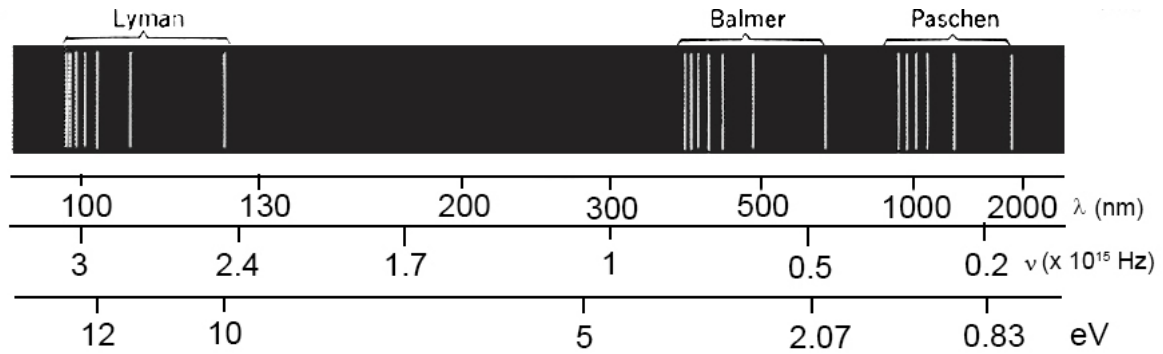


Figure 6.8: Measured hydrogen spectrum.

6.5 Spin-orbit coupling

Table 6.2 confirms what we had already discussed: the spin momentum of an electron for $\ell > 0$ can have two orientations relative to the orbital angular momentum: $j = \ell \pm 1/2$.

Important 6.5.1. *There is an associated magnetic moment for both the spin and the orbital angular momentum, proportional with them. Because magnetic moments interact, there is a small interaction energy between these two moments which then depends on*

¹³The *Lyman series* corresponds to transitions from the n -th level to the ground level ($m = 1$) and lies in the ultraviolet range. The *Paschen* (or Bohr) series are in the infrared band with $m = 3$, The *Brackett*, *Pfund* and *Humpfreys* series correspond to $m = 4, 5, 6$ respectively

the product of μ_L and μ_S , which, in turn are proportional to \mathbf{S} and \mathbf{L} respectively:

$$\mathcal{E}_{SL} = a \mathbf{S} \mathbf{L} \quad (6.5.1)$$

This is called spin-orbit coupling or spin-orbit interaction.

Depending on the relative orientation of \mathbf{S} and \mathbf{L} the interaction energy \mathcal{E}_{SL} ($j = \ell \pm 1/2$), is either positive or negative. This interaction energy should be added to the \mathcal{E}_n energy level to get the actual energy, i.e.

$$\mathcal{E}'_n = \mathcal{E}_n + \mathcal{E}_{SL}$$

Spin-orbit coupling leads to a splitting of every, but the s levels into two close lying levels. The split will materialize as a split of the spectral lines into two close lying *doublets*. For instance sodium has 2 yellow (or D) lines with wavelenghts of 589.0 nm and 589.6 nm ¹⁴, which is a 0.002 eV split.

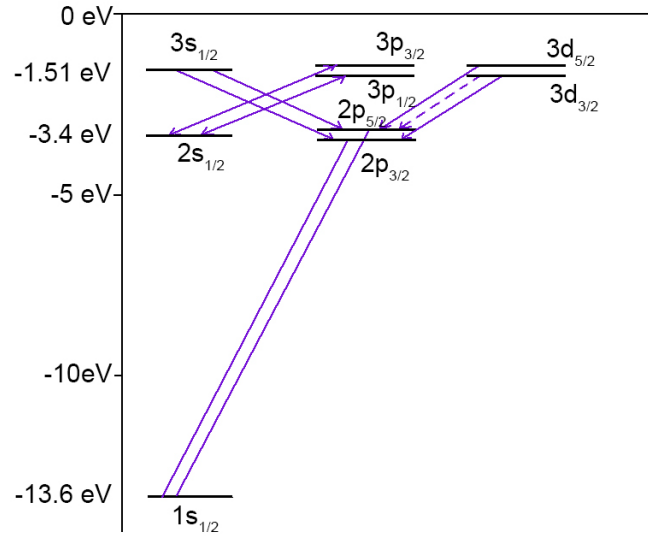


Figure 6.9: The first 4 energy levels of hydrogen and the possible transitions in the presence of spin-orbit coupling. The dashed lines correspond to $\Delta j = 0$ and has a low probability, therefore most transition appears as a doublet. The splitting indicated is not to scale (magnified).

¹⁴The study of these doublets led to the discovery of the electron spin, which only have two different orientations.

Spin-orbit coupling modifies the possible transitions depicted in Fig. 6.7, because of the creation of doublets. The selection rules that include spin-orbit interaction for hydrogen will become the following:

$$\Delta \ell = \pm 1, \quad \Delta j = 0, \pm 1 \quad \text{and} \quad \Delta m = 0, \pm 1 \quad (6.5.2)$$

From these the $\Delta j = 0$ has a low probability, because it requires $\Delta \ell = \pm 1$ together with $\Delta s = \mp 1$. Fig. 6.9 shows the possible transitions. As you see all lines with $\ell > 0$ in Fig 6.7 are split into doublets.

We may think that states $2s_{1/2}$ $2p_{1/2}$ should have the same energy. However this is not the case. The small difference, caused by the interaction of the electron with the so called *vacuum fluctuations*, called the *Lamb shift*. The deviation from the theoretical spectrum caused by the electron spin and relativistic corrections is very small so it appears as the *fine structure* of the spectrum.

6.6 The structure of atoms

Atoms consist of a nucleus containing Z protons (Z is the atomic number) and $A - Z$ neutrons, where A is the *atomic mass number* (a.k.a as *mass number* or *nucleon number*) and the *electron cloud* of Z electrons surrounding it. Atoms with the same number of protons but a different number of neutrons in their nucleus are called *isotopes*. The name “electron cloud” is used to emphasize that electrons are not classical particles.

The electrons can occupy states corresponding to the eigenfunctions in a centrally symmetric potential. The states with the same energy are called together as *shells*. The electrons do not all occupy the lowest energy shells for reasons we will discuss shortly, but are forced to fill in the shells in a strict order.

The electrons on the highest energy (“outermost”) shells determines the chemical properties of the atom.

The Schrödinger equation of the electrons of an atom contains the attractive Coulomb potential between electrons and the nucleus and the repulsive Coulomb potential between the electrons. The wave function of the electrons in the atom therefore must contain the coordinates of all electrons: $\varphi(\mathbf{r}_1, \mathbf{r}_2, \dots)$. This is a very complicated system, whose eigenvalue equation cannot be solved analytically.

6.7 He atom. Independent particle model. Pauli exclusion principle.

Hydrogen is the lightest element and also the most abundant element in the observable Universe. The next lightest and most abundant element is helium, discovered in 1868 in the spectrum of the Sun. It is about 24 % of the total elemental mass of the Universe,

which is more than 12 times the mass of all the heavier elements combined, but it is a relatively rare element on Earth ($5.2 \cdot 10^{-4} \%$ by volume). It is also the element with the second simplest structure as it consists of only two protons, one or two neutrons¹⁵ and two electrons. Although the Schrödinger equation for helium cannot be solved exactly the approximation methods we use here can give solutions in perfect agreement with experimental results. The Hamiltonian of helium ($Z = 2$) is

$$\hat{H} = -\frac{\hbar^2}{2m_e} \nabla_1^2 - \frac{\hbar^2}{2m_e} \nabla_2^2 + \frac{2e^2}{4\pi\epsilon_o} \left(-\frac{1}{r_1} - \frac{1}{r_2} \right) + \frac{1}{4\pi\epsilon_o} \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \quad (6.7.1)$$

where the third term containing the bracket is the potential energy of the electrons in the field of the nucleus while the last term is the interaction energy between the electrons. The wave function of this system of two electrons $\varphi(\mathbf{r}_1, \mathbf{r}_2)$ depends on the \mathbf{r}_1 and \mathbf{r}_2 position vectors (coordinates) of both electrons, and

$$\nabla_j = \left(\frac{\partial}{\partial x_j}, \frac{\partial}{\partial y_j}, \frac{\partial}{\partial z_j} \right) \quad j = 1, 2$$

This is a very complicated equation that cannot be solved analytically, therefore we will use a two step approach. First we will neglect the electron-electron interaction term and consider the electrons independent of each other. Then we will use perturbation calculus and take this term into account as a shielding of the field of the nucleus.

6.7.1 Independent particle model

If the two electrons are independent then the total wave function of the system can be written as the product of two atomic wave functions:

$$\varphi(\mathbf{r}_1, \mathbf{r}_2) = \varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) \quad (6.7.2)$$

The energy is then also written as the sum of two energies: $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$. Substituting these into the eigenvalue equation, dividing both sides with φ and reordering the terms gives:

$$\left(-\frac{\hbar^2}{2m_e} \nabla_1^2 \varphi_1 - \frac{2e^2}{4\pi\epsilon_o r_1} \varphi_1 - \mathcal{E}_1 \varphi_1 \right) + \left(-\frac{\hbar^2}{2m_e} \nabla_2^2 \varphi_2 - \frac{2e^2}{4\pi\epsilon_o r_2} \varphi_2 - \mathcal{E}_2 \varphi_2 \right) = 0 \quad (6.7.3)$$

¹⁵The two isotopes of He on Earth are ³He - 0.000137% and ⁴He - 99.999863%

Both brackets contain the Schrödinger equation for a single electron in a hydrogen like atom as expected, therefore

$$\begin{aligned}\varphi_1 &\equiv \varphi_{n\ell ms} \\ \varphi_2 &\equiv \varphi_{n'\ell' m' s'} \\ \mathcal{E}_{1,n} = \mathcal{E}_{2,n} &= 2^2 * \mathcal{E}_n^{hydrogen} = \frac{-4 * 13.6 \text{ eV}}{n^2} = \frac{-54.4 \text{ eV}}{n^2} \\ \mathcal{E} &= -108.8 \text{ eV} \frac{1}{n^2}\end{aligned}$$

The measured ground state ($n=1$) energy however is -79 eV and not -108.8 eV ! So the independent electron model is not good enough.

6.7.2 Shielding potential

In this approximation the effect of the electron–electron interaction energy is considered a perturbation. However the normal perturbation theory cannot be applied here, because this is not a small perturbation as it must add $+29.8 \text{ eV}$ to the total energy. What we can do is to average the effect of the other electron on the one we examine by approximating it with a central averaged potential, which partially shields the charge of the nucleus from the electron. Supposing the shielding decreases the charge acting on the selected electron by Z_{sh} the value of this shielding factor can be calculated from the required total ground state energy of -79 eV :

$$\begin{aligned}-79 \text{ eV} &= 2 (Z - Z_{sh})^2 \mathcal{E}_1^{hydrogen} = 2 (2 - Z_{sh})^2 (-13.6 \text{ eV}) \\ Z_{sh} &= 0.32\end{aligned}$$

6.7.3 The Pauli exclusion principle

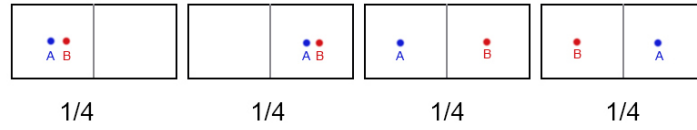
If we try to write the two complete wave function of the system as the product of two single electron wave functions as in (6.7.2) we run into a problem. The two wave functions may not describe the same single electron state ($\varphi_1(\mathbf{r}_1) \neq \varphi_2(\mathbf{r}_2)$), therefore we must determine which electron is in which state.

So the question arises: can we distinguish between the two electrons, can we identify them?

Example 6.4. *Please note that we did not ask how we would do this, we asked whether it is possible at all. This is a very subtle difference, which has an enormous effect. A simple thought experiment may illustrate it. Let us consider a box which contains two objects. These objects (e.g. gas molecules, electrons) can move around in the box randomly. Let us further suppose that we are unable to distinguish between these objects. Divide the*

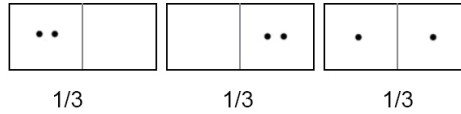
box into two equal partitions. Now determine the probability of finding one object in each partition!

Refer to the two partitions of the box as “left” and “right” respectively. If we assume the objects are distinguishable – in principle, (but maybe not in practice) – then we can label them, let us say with A and B. There are 4 different possibilities of the distribution of the objects in the boxes, each with the same probability as seen in the next figure:



Therefore the probability of one object in both partitions is $1/4 + 1/4 = 1/2$. So if we measure this probability as $1/2$, we can conclude that the objects in the box are, in fact distinguishable.

Now let the object be indistinguishable even in principle. In this case we cannot label the objects and the following figure shows the different possibilities.



Therefore the probability of one object in both partitions is $1/3$. So if the measured probability is $1/3$, we can conclude that the objects in the box we cannot distinguish are, in fact indistinguishable.

Important 6.7.1. Experiments show that electrons are indistinguishable objects in regions where their wave functions overlap.

This means that we can distinguish between two electrons when they are e.g. in two separate free atoms, but we cannot distinguish between electrons of the same atom. This we can take into account by writing the wave function as a linear combination of all possible ordering of the one-electron wave functions

$$\varphi(\mathbf{r}_1, \mathbf{r}_2)_{\pm} = \frac{1}{\sqrt{2}} (\varphi_1(\mathbf{r}_1) \varphi_2(\mathbf{r}_2) \pm \varphi_1(\mathbf{r}_2) \varphi_2(\mathbf{r}_1)) \quad (6.7.4)$$

where the $1/\sqrt{2}$ factor is the normalization constant. We can only measure the absolute square of the wave function and for both signs it is the same

$$|\varphi_+(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\varphi_-(\mathbf{r}_2, \mathbf{r}_1)|^2 \quad (6.7.5)$$

The combination φ_+ is a *symmetric* function for the exchange of the coordinates of the two electrons

$$\varphi_+(\mathbf{r}_1, \mathbf{r}_2) = \varphi_+(\mathbf{r}_2, \mathbf{r}_1),$$

while

$$\varphi_-(\mathbf{r}_1, \mathbf{r}_2) = -\varphi_-(\mathbf{r}_2, \mathbf{r}_1)$$

is *antisymmetric*. It is easy to see from (6.7.4) that the antisymmetric wave function is $\varphi_- \approx 0$, when the two electrons are in the same atom and described by the same n, ℓ, m quantum numbers and $\mathbf{r}_1 \approx \mathbf{r}_2$. This means that the probability density of two electron atoms with antisymmetric (spatial) wave functions is higher when the two electrons with the same quantum numbers are apart from each other, i.e. this is the preferred state. If the electrons are further apart the effective shielding of the nucleus will be smaller, the other electron will bound to the nucleus tighter with lower energy. For symmetric (spatial) wave functions no such restriction apply. Therefore *the energy of the electrons with antisymmetric wave functions is smaller than that for symmetric wave functions*. (Note that these energy levels can be measured spectroscopically.)

The complete wave function of the electron in an atom however includes the spin as well. The spin of the two electrons can be parallel ($s = s_1 + s_2 = 1$) or anti-parallel ($s = s_1 - s_2 = 0$). In the first case the z-component can be $m_s = -1, 0, 1$, which is called a *triplet* and has *symmetric spin functions*, while in the second case $m_s = 0$, which is a *singlet* that is an *anti-symmetric spin function*. If the function for “up” spin of the first electron is $\chi_{1,\uparrow}$ and for “down” spin is $\chi_{1,\downarrow}$ and for the second the index 2 is used then the complete spin function can be one of the following combinations

$$\begin{aligned} \chi_s &= \begin{cases} \chi_{1,\uparrow}\chi_{2,\uparrow} \\ \frac{1}{\sqrt{2}}(\chi_{1,\uparrow}\chi_{2,\downarrow} + \chi_{1,\downarrow}\chi_{2,\uparrow}) \\ \chi_{1,\downarrow}\chi_{2,\downarrow} \end{cases} & \text{triplet} \\ \chi_a &= \frac{1}{\sqrt{2}}(\chi_{1,\uparrow}\chi_{2,\downarrow} - \chi_{1,\downarrow}\chi_{2,\uparrow}) & \text{singlet} \end{aligned}$$

The complete wave function also can be symmetric or antisymmetric:

$$\varphi_s(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} \varphi_+(\mathbf{r}_1, \mathbf{r}_2) \chi_s \\ \varphi_-(\mathbf{r}_1, \mathbf{r}_2) \chi_a \end{cases} \quad (6.7.6)$$

$$\varphi_a(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} \varphi_+(\mathbf{r}_1, \mathbf{r}_2) \chi_a \\ \varphi_-(\mathbf{r}_1, \mathbf{r}_2) \chi_s \end{cases} \quad (6.7.7)$$

Spectroscopic measurements show that the symmetric spatial φ_+ state is always a singlet, which requires the antisymmetric spin function χ_a , while φ_+ is always a triplet, i.e.

it must have the symmetric spin function χ_s .

It follows the complete electron wave function *must* always be antisymmetric.

This principle is true not only in an atom and not only for electrons, but for any systems consisting of the same half-integer spin particles (e.g. electrons, protons or neutrons).

Important 6.7.2. *The complete wave function, which includes the spin, of a system of any half-integer spin particles must be antisymmetric for the exchange of the coordinates of any two particles.*

This is called the Pauli exclusion principle or simply the exclusion principle.

The name exclusion principle emphasizes the consequence of the antisymmetric nature of the complete wave function, – and not only in an atom – namely that no two electrons can be in the same state, where all of their quantum numbers would be equal.

This is another, equivalent phrasing of the Pauli exclusion principle.

In an atom the exclusion principle says that no two electron wave functions may have all of their four quantum number to be the same. But if we consider two non interacting free atoms, than it is possible that they have electrons with the same n, ℓ, m, m_s quantum numbers.

Important 6.7.3. *The Pauli exclusion principle results in a strong repulsive force between electrons of a multi-electron system that does not allow them to occupy the same state. This force is electrostatic in nature, therefore the exclusion principle does not introduce any new kind of force.*

The wave function of an N-electron system (e.g. a multi-electron atom) depends on the quantum numbers and coordinates of all of the electrons. To make the equations easier to read we denote a given combinations of the four quantum numbers n, ℓ, m, m_s with a single letter and replace all arguments with a number, i.e.

$$\begin{aligned} \{n, \ell, m\} &\Rightarrow a \\ \{n', \ell', m'\} &\Rightarrow b \\ \varphi_{a,b,\dots}(\mathbf{r}_1, \mathbf{r}_2, \dots) &\Rightarrow \varphi_{a,b,\dots}(1, 2, \dots) \end{aligned}$$

It is easy to see that with this notation the antisymmetric spatial wave function of a two electron system can be written as a determinant:

$$\frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_a(1) & \varphi_b(1) \\ \varphi_a(2) & \varphi_b(2) \end{vmatrix} = \frac{1}{\sqrt{2}} (\varphi_a(1)\varphi_b(2) - \varphi_a(2)\varphi_b(1)) \quad (6.7.8)$$

Similarly for N electrons (variables)

$$\varphi_{abc\dots} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_a(1) & \varphi_b(1) & \varphi_c(1) & \cdots \\ \varphi_a(2) & \varphi_b(2) & \varphi_c(2) & \cdots \\ \varphi_a(3) & \varphi_b(3) & \varphi_c(3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} \quad (6.7.9)$$

Chapter 7

Electron structure of atoms.

7.1 The periodic table of elements.

The periodic table of elements by the Russian scientist and inventor Dmitri Ivanovich Mendeleev appeared in print in 1869. It contained all 63 elements known at that time grouped by their chemical properties. He developed his table to illustrate periodic trends in the properties of the then-known elements. What is more Mendeleev was able to *predict* the properties of then unknown elements that would be expected to fill gaps in his table, elements, most of which were discovered later with chemical properties predicted by him. The table has been expanded and refined with the discovery of further new elements, the ones after *californium* (atomic number 98) have only been synthesized in laboratory. The last one (so far) is *ununoctium* with an atomic number of 118¹. An example for the periodicity is shown in Fig. 7.2. The *ionization energy* is the energy needed to remove one electron² from a neutral atom. The periodicity of the ionization energy suggests that elements with similar ionization energies may have the same number of outermost electrons.

The periodic table was constructed phenomenologically. Only after the discovery of quantum mechanics and especially the Pauli exclusion principle became a physical explanation available.

We have seen that atomic orbitals may be characterized by 4 quantum numbers $\{n, \ell, m, m_s\}$, where $n = 1, 2, \dots$, $\ell = 0, 1, 2, \dots, n - 1$, $m = -\ell, \dots, 0, \dots, \ell$ and $m_s = \pm \frac{1}{2}$. For any ℓ (subshell) there are $2\ell + 1$ possible m values, which when combined with the spin produce $X = 2(2\ell + 1)$ different combinations of the quantum numbers for a given

¹The number of possible elements are not known. There are different estimates for it. It is of some interest that in the Bohr model atoms with atomic numbers above 137 would require the 1s electrons to travel faster than the speed of light in vacuum.

²As we will see soon the outermost electron of the atom will be removed.

Group →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
↓ Period																		
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba		72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra		104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
Lanthanides			57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
Actinides			89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

Figure 7.1: Standard form of the periodic table. It can be deconstructed into four rectangular box: the *s-block* - first two columns to the left, the *p-block* - columns 13-18 to the right, the *d-block* (columns 3 to 12) in the middle and the *f-block* - lanthanides and actinides below the d-block. The rows are called *periods*, the columns are called *groups*.

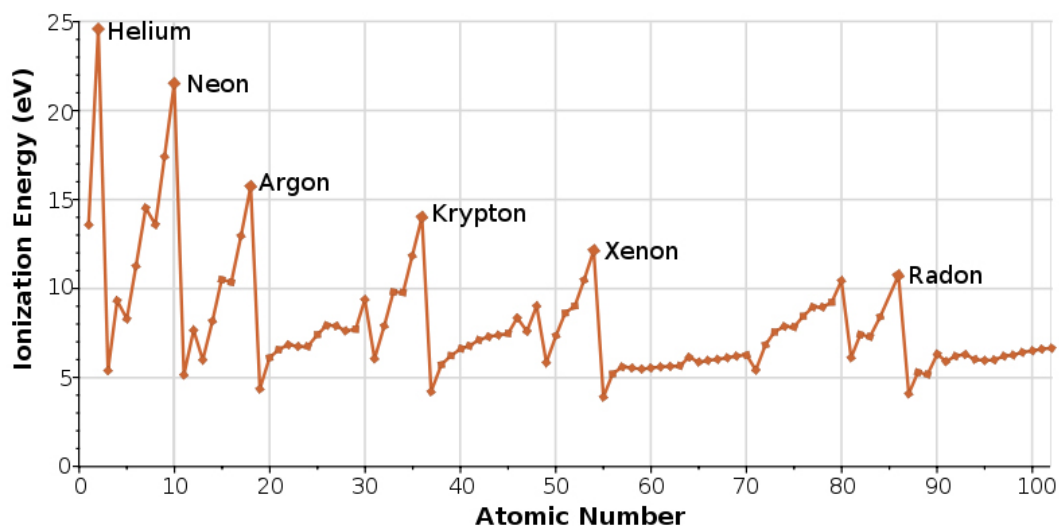


Figure 7.2: Ionization energies of neutral elements. The names indicates the noble gases in column 18th of the periodic table.

principal quantum number n . According to the Pauli principle the maximum number of electrons that can be put on this subshell (or in these states) for a given n is also $2(2\ell + 1)$. The total number of electrons that can be in the same shell n is $2n^2$.

As we have already introduced in Section 6.2 the usual notation for a given atomic *electron configuration* is a series of expressions $n\ell^X$, where the “exponent” X is the number of electrons occupying that subshell. For instance for hydrogen it is $1s^1$, for helium this is $1s^2$, for lithium $1s^2 2s^1$ and for beryllium $1s^2 2s^2$, etc.³ as $[He] 2s^2$. Similarly to the angular momentum (subshell) the shells themselves also have single names of capital letters. $n = 1 \Rightarrow K, n = 2 \Rightarrow L, n = 3 \Rightarrow M, n = 4 \Rightarrow N$, etc.

7.2 Hund’s rules.

Let us see how the atoms of the elements are built up starting with hydrogen at $Z = 1$. At any Z the electrons are added to the lowest energy state allowed by the exclusion principle. When the maximum number of electrons that can be accommodated by a shell (a given n) is reached the next new shell is used. As we saw with helium, electrons in the inner shells partially shield the charge of the nucleus from the outer electrons, therefore it is possible that a new shell is started before the shell below it is completely filled in. Fig. 7.3 shows the build up of the first 10 elements^{4, 5}.

Important 7.2.1. *Hund’s rule states the lowest energy atomic state is the one which maximizes the sum of the S values for all of the electrons in the open subshell. The orbitals of the subshell are each occupied singly with electrons of parallel spin before*

³This can also be abbreviated. The symbol in the square bracket is always the symbol of the *noble gas* with electron structure corresponding to that of the inner shells of the element in question.

⁴The complete table of atomic configuration is in Appendix ??.

⁵There are three general rules of thumb how to determine the term symbol (e.g. $2p^1$), called *Hund’s rules*⁶, which are the following:

1. For a given electron configuration, the term with maximum multiplicity has the lowest energy. The multiplicity is equal to $2S + 1$, where S is the total spin angular momentum for all electrons. The term with lowest energy is also the term with maximum S .
2. For a given multiplicity, the term with the largest value of the orbital angular momentum number L , has the lowest energy.
3. For a given term, in an atom with outermost subshell half-filled or less, the level with the lowest value of the total angular momentum quantum number J lies lowest in energy. If the outermost shell is more than half-filled, the level with the highest value of J , is lowest in energy.

From these the first one is the most important for chemistry and is often referred to as *Hund’s rule*.

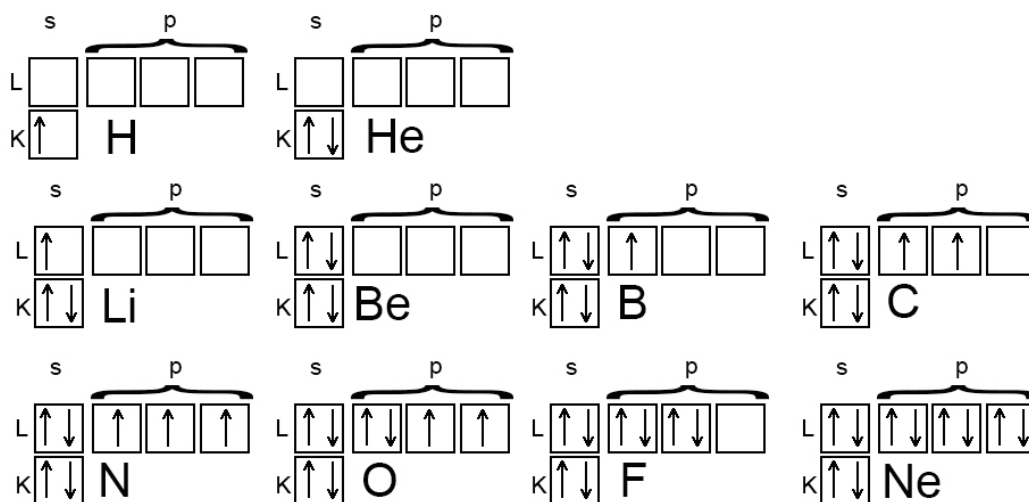


Figure 7.3: Buildup of the first 10 elements. The boxes represent the sub-shells, the arrows the unpaired and paired electron spins. Note that selecting the up state for an unpaired electron is just for the representation.

double occupation occurs⁷. If we have a look at Fig. 7.3 we see this rule in action. We have already explained the physical explanation for this phenomena for helium: electrons in singly occupied orbitals are less effectively shielded from the nucleus, so that such orbitals contract and electron–nucleus attraction can be more intense.

Physical and chemical properties of atoms are determined by the electronic configuration in the ground state and in the closely laying excited states.

Elements with completely filled shells (all of the electrons are paired) are the most stable configurations. They are called noble gases because their paired electrons are hard to excite. Although beryllium also has only paired electrons like helium it is not a noble gas as the L shell is not completely filled and only a small amount of energy is needed to excite one of its $2s$ electrons to one of the empty states of the p subshell.

When there is only a single unpaired electron on the outermost subshell (e.g. in Li, Na, K) it can easily be excited and this leads to metallic behavior.

7.3 Valence electrons

Total orbital and spin angular momenta of the closed shells are zero. The system of electrons on inner, completely filled (closed) shells is called the *core* or *kernel* and its

⁷This is occasionally called the "bus seat rule" since it is analogous to the behavior of bus passengers who tend to occupy all double seats singly before double occupation occurs.

electrons are called *core electrons*. It requires a lot of energy, so it is hard to excite core electrons of the atoms. Electrons on the outermost partially filled shells feel the shielded nuclear charge and they are more easy to excite. This means that only these outer shell electrons take part in most interactions that involves the atom. These electrons are responsible for chemical bonding too, for this reason they are called *valence electrons*.

As an example let us examine the electronic structure of the simplest single-valence electron atom lithium. The electron configuration is $1s^2 2s^1$, so the core is the two $1s$ electrons and there is one valence electron. As the charge distribution $(-e |r^2 R(r)|^2)$ of

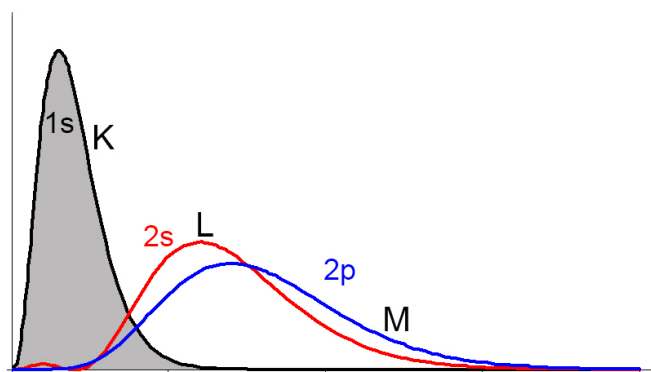


Figure 7.4: Radial charge distributions in lithium. The grayed area denotes the core electrons. Notice how the valence ($2s$ and $2p$) electron distributions intersects the core.

the valence electron penetrates into the area of the core the effective nuclear charge that acts on it will change. It will be $+e$ when the valence electron is far out, as the nuclear charge is shielded by the core electrons, and $+3e$ when it is deeply inside the kernel and the shielding is virtually non-existent. The corresponding energy then is between that for hydrogen and that for Li^{2+} . But the penetration depends on the angular momentum too. The smaller the angular momentum the larger the penetration: the s orbital penetrates deeper than the p orbital. This would be true for other single valence electron atoms too. So unlike to hydrogen like ions the energy of the single valence electron depends not only on the principal quantum number n but also on the angular momentum as you can see in Fig. 7.5

7.4 X-ray emission

As we said in the previous section core electrons are hard to excite. It requires a lot of energy but it is not impossible. For instance when high velocity electrons, with energies in the $> 10 \text{ keV}$ range, collide with a metal target they can kick out core electrons from their shells. Either a higher energy core electron or an electron from an outer shell, or

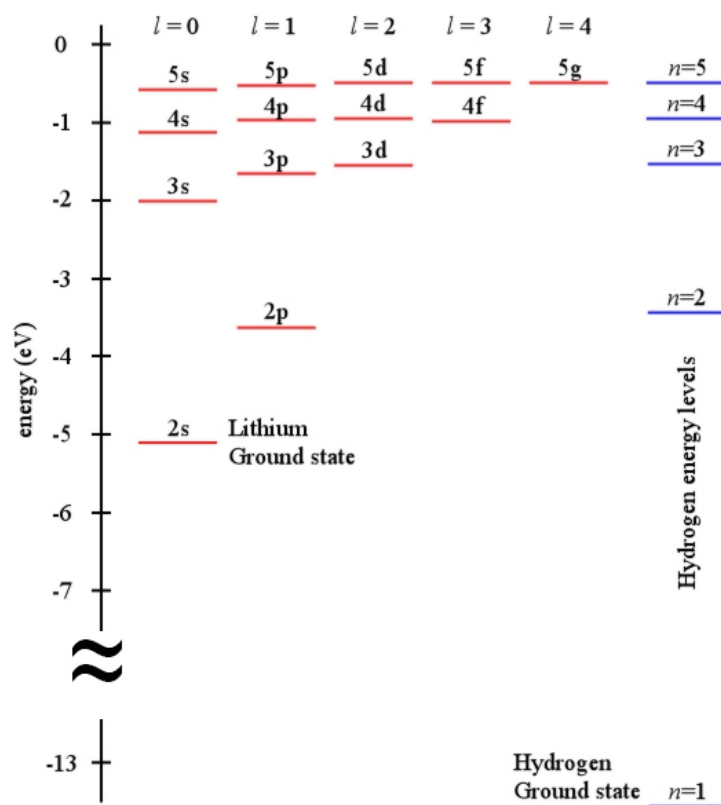


Figure 7.5: Lithium valence electron energy levels compared to hydrogen electron energy levels.

even a free electron may fall into the remaining *core hole* or *vacancy*, while emitting high energy electromagnetic radiation, called X-rays (Röntgen rays)

X-rays have a wavelength in the range of 0.01 to 10 nanometers, corresponding to frequencies in the range $3 \cdot 10^{16} \text{ Hz}$ to $3 \cdot 10^{19} \text{ Hz}$ and energies in the range 100 eV to 100 keV . They are shorter in wavelength than UV rays and longer than gamma rays⁸. X-rays with photon energies above $5 - 10 \text{ keV}$ (below $0.2 - 0.1 \text{ nm}$ wavelength), are called *hard X-rays*, these are used in medical imaging and crystallography, while those with lower energy are called *soft X-rays* as they are easily absorbed by air (but are also used in *mammography*).

X-ray photons carry enough energy to ionize atoms and disrupt molecular bonds. This makes X-rays a type of ionizing radiation and thereby harmful to living tissue. A very high radiation dose over a short amount of time causes radiation sickness, while lower doses can increase the risk of radiation-induced cancer. In medical imaging this increased cancer risk is generally greatly outweighed by the benefits of the examination.

The spectrum of the emitted X-rays contains one or more sharp peaks, the *characteristic X-ray peaks*, corresponding to the energy of the *recombination* of the electron with the hole and a lower intensity continuous part emitted by the electrons deflected by the electric field of the nucleus, called *brehmstrahlung* after the original German expression⁹. If the energy of the colliding electrons is not high enough no characteristic peaks can be observed as seen in Fig. 7.6.

The maximum energy of X-ray photons is determined by the energy of the colliding electrons, therefore below a *cutoff wavelength* no radiation is produced as it is shown in this figure¹⁰. The X-ray peak is labeled by the name of the shell where the hole is generated plus a greek letter, the latter defines the source level. E.g. K_{α} x-rays are produced when the hole is on the K shell ($n=1$) and the electron that fills this hole come from the $n=2$ shell, K_{β} rays when the electron comes from the $n=3$ shell.

X-ray emission is not the only possible way to lose the excess energy. It may be transferred to another electron of the atom which is then ejected from it. This process is called the *Auger process* and the emitted electron is the *Auger electron*. There is a material science method called Auger Electron Spectroscopy or AES based on this effect.

The design and actual implementation of the most widely used type side-window

⁸The distinction between X-rays and gamma rays is not universal. One distinction may be based on their origin: X-rays are emitted by electrons, while gamma rays are emitted by the atomic nucleus. The one we use is based on an arbitrarily chosen wavelength limit of 10^{-11} m , below which the radiation is called gamma rays

⁹“Bremsen” means “to break” and “strahlung” means radiation.

¹⁰The formula for the cutoff wavelength is

$$\lambda_{min} = \frac{hc}{eV} \approx \frac{1239.8 \cdot 10^{-9} \text{ m}}{V \text{ in kV}}$$

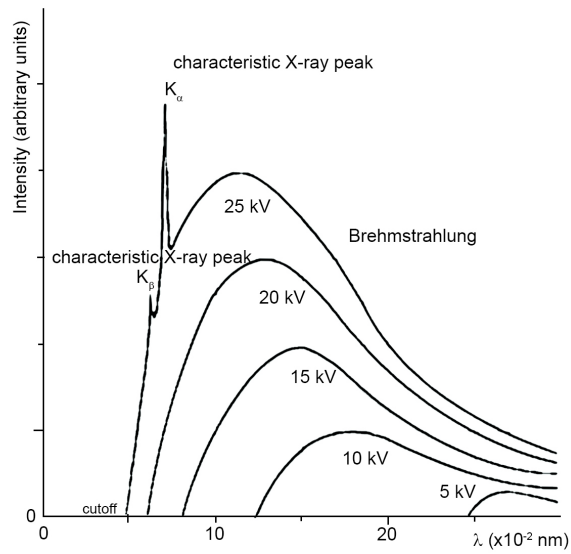


Figure 7.6: X-ray intensity vs wavelength function. Below a cutoff energy no X-ray peaks are produced.

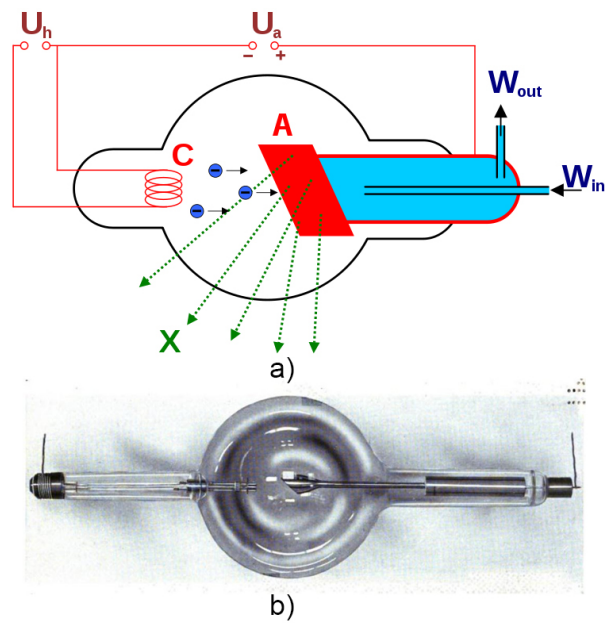


Figure 7.7: A typical side-window X-ray tube. Fig. a) schematics, Fig. b) photo of a Coolidge X-ray tube from around 1917.

X-ray *vacuum tube* can be seen in Fig. 7.7. Electrons created by a hot cathode “C” are focused and accelerated toward the anode “A”, made out of tungsten or molybdenum, by a high U_a voltage. Because both X-ray producing process is very inefficient (efficiency is about 1%) most of the energy of the colliding electron heats the anode up. To ensure proper operation the anode must be cooled by either a circulating coolant (e.g. water) or by mechanical rotation. X-rays are emitted essentially perpendicular to the electron current. The anode is angled at $1 - 20^\circ$ degrees off perpendicular to the electron current allowing some of the generated X-rays to leave the tube through a special side window. The power of such an X-ray tube is in the range of 0.1 to 18 kW¹¹.

¹¹X-rays may be generated by other processes, e.g. by *synchrotron radiation*, which is generated by particle accelerators. Its unique features are X-ray outputs many orders of magnitude greater than those of X-ray tubes.

Chapter 8

Molecules

8.1 H_2^+ - The hydrogen molecule ion

The notion of a molecule was first accepted in chemistry because of Dalton’s laws of “definite and multiple proportions” and Avogadro’s law. The word molecule means an electrically neutral group of two or more atoms held together by *chemical bonds*. But what does “chemical bond” mean in physics? This is the question we search the answer for in this section.

Molecules are components of matter and are common in organic substances. However, the majority of familiar solid substances on Earth, including most of the minerals that make up the crust, mantle, and core of the Earth, while they contain many chemical bonds, but are not made of identifiable molecules¹.

The answer to our question is best answered by using an example of a *molecular ion*, the *hydrogen molecule ion* or *dihydrogen cation*, H_2^+ . It consists of two protons and one electron. As this molecule has only one electron, understanding it is as fundamental in the study of the chemical bonds as was the hydrogen atom for the structure of the atoms of the periodic table of elements. The ion is commonly formed in molecular clouds in space, and is important in the chemistry of the interstellar medium.

Fig. 8.1 shows the schematics of this ion. The figure obviously is only meant as a base for the coordinate definitions and does not represent the real structure of the H_2^+ ion. The wave function of the system will depend on the positions of all three particles, therefore the Schrödinger equation of this system is

$$-\frac{\hbar^2}{2} \left(\frac{1}{M} \nabla_{R_1}^2 + \frac{1}{M} \nabla_{R_2}^2 + \frac{1}{m_e} \nabla_r^2 \right) \varphi + V(r_1, r_2, R) \varphi = \mathcal{E} \varphi \quad (8.1.1)$$

where $\varphi(r_1, R)$ is the complete wave function of the system, M and m_e are the proton and electron mass respectively and the lower index of the ∇ operators denotes the variable

¹As we will see in Chapter 11 most of the crystals are not made of molecules.

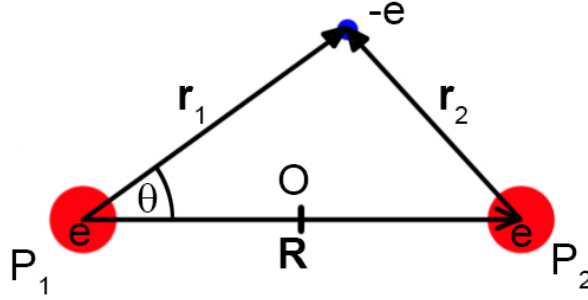


Figure 8.1: Schematic representation of the H_2^+ ion. The origin of the coordinate system O is in the middle between the protons, the positions of protons are therefore $\mathbf{R}_1 = -\mathbf{R}/2$ and $\mathbf{R}_2 = \mathbf{R}/2$, and the electron coordinate vector is \mathbf{r} . The proton-electron distances are r_1 and r_2 , where

the differentiation is respect to. The potential in this molecule is

$$V(r_1, r_2, R) = \frac{e^2}{4\pi\epsilon_0} \left(-\frac{1}{r_1} - \frac{1}{r_2} + \frac{1}{R} \right) \quad (8.1.2)$$

(8.1.1) is a complicated equation which we will try to solve by separating the movement of the electron and the protons based on their greatly different masses and assuming the protons are the equilibrium distance of 0.106 nm from each other. That is we disregard the movement of the protons and solve the eigenvalue equation of the electron only:

$$\begin{aligned} \hat{H}\varphi(r) &= \mathcal{E}\varphi(r) \quad \text{where} \\ \hat{H} &= -\frac{\hbar^2}{2m} \nabla^2 + V(r_1, r_2, R). \end{aligned}$$

To help understanding the H_2^+ molecule ion it can be thought to be formed by a combination of a neutral hydrogen atom and a single proton²: $H + H^+ \Rightarrow H_2^+$.

- Imagine H and H^+ are far apart. In this case the electron is localized at the H atom. There are two equivalent configurations as the H atom may be the one that contains either proton P_1 or P_2 and these are indistinguishable configurations. See Fig 8.2. In this case the ground state the electron wave function will be the $1s$ function in hydrogen: $\varphi_{1s} = \frac{1}{\sqrt{\pi} a_o^{3/2}} e^{-r/a_o}$

²This is not the natural process of the formation of this *dihydrogen* ion though. It is formed in nature when cosmic rays knock an electron off the hydrogen molecule leaving the cation behind. The ionization energy of the H_2 molecule is 15.603 eV .

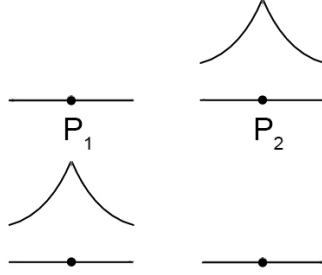


Figure 8.2: Two equivalent configurations of the $H + H^+$ system

- If we decrease the distance between the protons, the electron starts to feel the pull of the free proton.
- When the molecule finally formed φ must reflect the symmetry of the molecule. Because the molecule is symmetric to the midpoint between the two protons the electron wave function must be either symmetric or antisymmetric. We can *approximate* the wave function as the linear combination of two atomic orbitals, one centered on proton P_1 $\varphi_a(\mathbf{r}_1)$ and the other one centered on P_2 $\varphi_a(\mathbf{r}_2)$:

$$\varphi^\pm(\mathbf{r}) = C [\varphi_a(\mathbf{r}_1) \pm \varphi_a(\mathbf{r}_2)], \quad (8.1.3)$$

where the C complex number is the normalization constant. The wave functions and the corresponding probability distributions are schematically shown in Fig. 8.3.

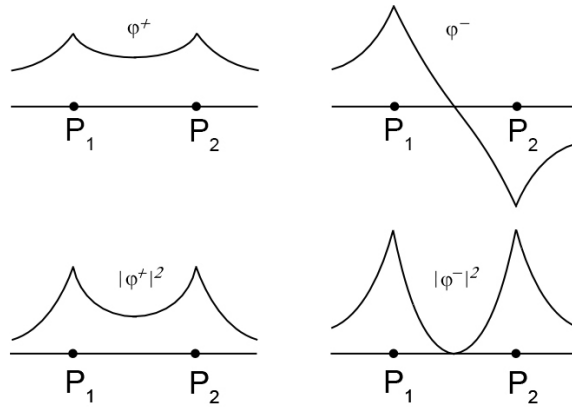


Figure 8.3: Possible wave functions and probability distributions in a H_2^+ ion.

The φ^\pm are called *molecular orbitals*

Looking at the probability distributions we can guess that the even φ^+ molecular orbital will have the lower energy, because according to the $|\varphi^+|^2$ probability distribution the electron can be found between the two protons with a high probability shielding the protons from the Coulomb repulsion of the other, thereby decreasing the total energy. For the odd combination this probability is much smaller, almost no shielding occurs and the repulsion of the protons is higher. Knowing the wave function is a linear combination of well known hydrogen atomic wave functions the total energy can be calculated by evaluating the integral

$$\mathcal{E} = \int (\varphi^\pm)^* \hat{H} \varphi^\pm dV = |C|^2 \int [\varphi_a(r_1) \pm \varphi_a(r_2)]^* \hat{H} [\varphi_a(r_1) \pm \varphi_a(r_2)] dV$$

Here, and in the following $dV \equiv d^3 r_1$. The result will have the following form:

$$\mathcal{E} = \mathcal{E}_a + \frac{e^2}{4\pi\epsilon_o} \frac{1}{R} - \frac{A \pm B}{1 \pm S}, \quad (8.1.4)$$

where E_a is the atomic energy for either $\varphi_a(\mathbf{r}_1)$ or $\varphi_a(\mathbf{r}_2)$, the second term is the Coulomb repulsion between the two protons. The symbols in the third term are:

$$A \equiv \frac{e^2}{4\pi\epsilon_o} \int \frac{|\varphi_a(r_1)|^2}{r_2} dV = \frac{e^2}{4\pi\epsilon_o} \int \frac{|\varphi_a(r_2)|^2}{r_1} dV$$

is the Coulomb attraction of the electron to the other proton.

$$B \equiv \frac{e^2}{4\pi\epsilon_o} \int \frac{\varphi_a^*(r_1)\varphi_a(r_2)}{r_1} dV = \frac{e^2}{4\pi\epsilon_o} \int \frac{\varphi_a^*(r_2)\varphi_a(r_1)}{r_2} dV$$

is a quantum mechanical term called the *resonance integral*, and

$$S \equiv \int \varphi_a(r_1) \varphi_a(r_2) dV$$

is the quantum mechanical *overlap integral*.

By either calculating the energy minimum for $\varphi_a = \varphi_b = \varphi_{1s}$ the corresponding energy - proton distance curve is shown in Fig 8.4 we can see that the green curve corresponding to the positive sign in (8.1.4) has a minimum, while the red curve for the negative sign does not. The first one is an *attractive* energy curve and leads to a formation of the molecule therefore the corresponding φ^+ molecular orbital is called *bonding molecular orbital*. The other one is a *repulsive* energy curve, which has no minimum, therefore no bonding is possible for the φ^- wave function. It is therefore called *anti-bonding molecular orbital*³

³Usually the symmetries of a given molecule are also denoted with the wave function. The ground state of the H_2^+ ion for instance is denoted with $\sigma_g 1s$, while the first excited state with $\sigma_u 2p$. The suffixes *g* and *u* are from the German words *gerade* and *ungerade* (meaning even and odd) which denote the symmetry under space inversion. Their use is standard practice for the designation of electronic states of diatomic molecules.

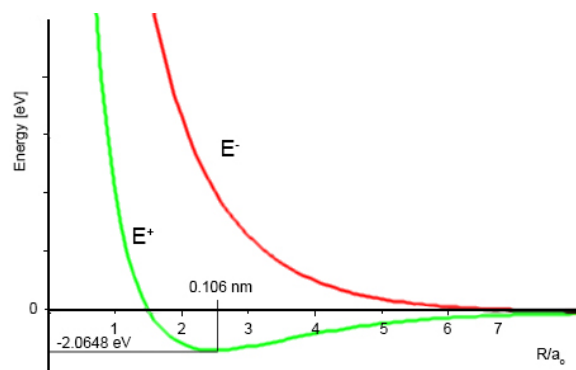


Figure 8.4: Energy of bonding (+) and anti-bonding (-) molecular orbitals as a function of the proton-proton distance for the ground state of the H_2^+ ion.

8.2 Diatomic homonuclear molecules. Molecular orbitals. Chemical bond.

After discussing the H_2^+ ion let us examine the H_2 molecule! Because both atoms in the molecule are of the same element their bonding is called *homonuclear*. And because the two nuclei determine a line in space the potential felt by the electrons is not centrally symmetric, it only has one symmetry axis. As a result the orbital angular momentum L is not conserved. Usually the line connecting the two nuclei is selected as the z axis with an origin halfway between the protons. The cylindrical symmetry then means only the $L_z = m_l \hbar$ ($m_l = 0, \pm 1, \pm 2, \dots$) component is preserved. The energy of the state depends only on the absolute value of m_l denoted by λ . The first few values of $\lambda = 0, 1, 2, 3, \dots$ are called σ, π, δ and ϕ states. Taking the $s = \pm \frac{1}{2}$ spin quantum number into account the possible number of electrons on these states are 2, 4, 4 and 4, respectively.

There are other homonuclear molecules such as N_2 or O_2 . The potential in all of them has cylindrical symmetry, which leads to wave functions, which also have cylindrical symmetry. Their symmetries are distinguished by their u -odd and g -even indices: $\sigma_g, \sigma_u, \pi_g, \pi_u$ etc. To understand how the chemical bond forms in this case we use an argument similar to the one followed for the H_2^+ ion: we imagine the two atoms far apart with wave functions ψ_1 and ψ_2 (orbitals) for the two electrons, and try to write up the wave function when the atoms get very close. The angular distribution of molecular orbitals formed from atomic orbitals are shown in Fig. 8.7, while the corresponding energy levels are in Fig. 8.5.

The electronic configuration of the H_2 molecule is still simple enough to analyze without actually solving the Schrödinger equation. The total electric potential of the

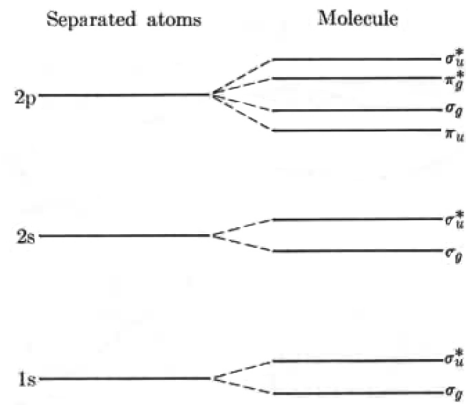


Figure 8.5: Energy levels before and after a diatomic molecule is formed. As before $*$ denotes the anti-bonding states, g the even and u the odd spatial wave function.

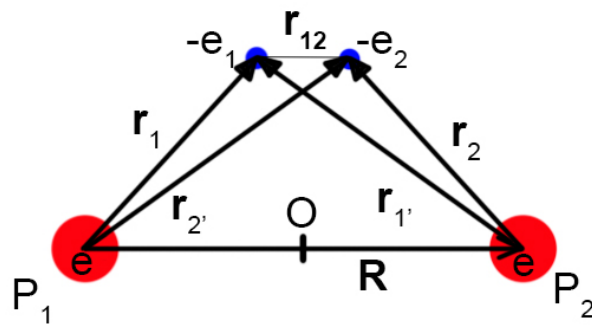


Figure 8.6: Electronic structure of the H_2 atom.

system is

$$V(r) = \frac{e^2}{4\varphi\epsilon_o} \left(-\frac{1}{r_1} - \frac{1}{r_{1'}} - \frac{1}{r_2} - \frac{1}{r_{2'}} + \frac{1}{r_{12}} + \frac{1}{R} \right), \quad (8.2.1)$$

where the meaning of the indices can be read from Fig. 8.6. Although in this formula and in the figures we have indices for the electrons, this is just to signal that one set of the vectors belong to one of the electrons, while the other one to the other electron, but as electrons are indistinguishable we cannot say to which one. The electrons in the molecule has loose their identities and now belong to both of the atoms, i.e. to the molecule.

Instead of solving the Schrödinger equation, however we will base our discussion to the results of the previous section.

According to the exclusion principle if the two electrons have opposite spins they can be accommodated in the bonding level $\sigma_g 1s$ giving the configuration $(\sigma_g 1s)^2$. This bonding state has an energy diagram like the one with the minimum in Fig. 8.4, just the equilibrium distance in this case is 0.074 nm and the bonding energy is -4.476 eV . If both electrons have the same spin one of them can be in the bonding state $\sigma_g 1s$, while the other one must occupy the anti-bonding⁴ $\sigma_u^* 1s$ state, so the resulting configuration will be $(\sigma_g 1s)(\sigma_u^* 1s)$. In this case the anti-bonding effect dominates and the corresponding energy diagram has no minimum (like the red curve in Fig. 8.4).

A similar train of thought can be followed for the He_2^+ molecule ion. It has three electrons two of which are in the bonding state $\sigma_g 1s$ and the third in the anti-bonding state $\sigma_u^* 1s$, so the configuration is $(\sigma_g 1s)^2(\sigma_u^* 1s)$. This results in a stable molecule with a dissociation energy equal to 2.5 eV ⁵.

The He_2 molecule has four electrons, two in the bonding and two in the anti-bonding state. The electronic configuration is $(\sigma_g 1s)^2(\sigma_u^* 1s)^2$, This configuration is not stable, which explains why helium is a monatomic gas. However an excited He_2 molecule, in which one of the anti-bonding electrons is excited up to the bonding state $\sigma_g 2s$, may exist with the electronic configuration $(\sigma_g 1s)^2(\sigma_u^* 1s)(\sigma_g 2s)$.

Fig. 8.8 shows the electronic configuration of homonuclear diatomic molecules up to Ne_2 . As you can see in the table in Fig. 8.8 molecular binding generally occurs when two electrons occupy bonding molecular orbitals, i.e. they concentrate in the region of between the two combining atoms. Their bonds are called *covalent* bonds.

However this is not a strict rule with no exceptions. Not only because the He_2^+ molecular ion contains only 3 electrons, as in both Be_2 and O_2 the last pair of electrons are in π orbitals with parallel spins. This behavior is due to the fact that both of these molecules have only two electrons in that energy level although the π orbitals may accommodate four. In the atomic case we found that the repulsion between the electrons

⁴The asterisk denotes anti-bonding

⁵The metastable He_2^- molecule ion has a very small dissociation energy, because of the relatively small nuclear charge, therefore its lifetime is very short ($\approx 18.2\mu s$) too.

favor the most antisymmetric spatial wave function, for which the exclusion principle requires the most symmetric spin function which results in the parallel spin state.

But having parallel spins means the molecule always has a fix angular momentum of \hbar , which gives rise to a permanent magnetic moment. Therefore O_2 is a paramagnetic gas, while most other diatomic molecules have no permanent magnetic moment, so they are diamagnetic⁶.

The stability of a molecule (expressed by the dissociation energy) depends on the relative number of bonding and anti-bonding pairs of electrons. This explains why He_2 and Be_2 are not stable and why the stability of molecules N_2 , O_2 , F_2 and Ne_2 decreases, because the difference between bonding and anti-bonding pairs for these molecules is 3, 2, 1 and 0 respectively.

The component of the total angular momentum of the electron along the molecular axis is given by

$$M_L = \sum_i m_\ell^{(i)}$$

and the energy of a given state depends on its absolute value Λ . In this case the states are labeled by capital letters $\Sigma, \Pi, \Delta, \Phi, \dots$. Given that the resultant spin is S the symbol of a state (or *term*) is

$$^{2S+1}\Lambda$$

which is in the last column in Fig. 8.8.

8.3 Heteronuclear molecules.

When different atoms form molecules, e.g. HCl , CO and $NaCl$, we are talking of *heteronuclear* molecules. There is no center of symmetry, not even in diatomic heteronuclear molecules, thus although the electronic states are still called σ, π, δ , etc, they are not classified as g or u . Only electrons in the outermost, unfilled shells must be considered when talking about chemical bond.

Take $NaCl$ as an example. Na has 11, Cl 17 electrons. The electrons in closed (filled) shells are so tightly bound to their respective nuclei that they are hardly affected by the presence of a second nucleus. Similarly electrons on unfilled shells with coupled opposite spin pairs are not expected to participate strongly in the binding of the atoms. This leaves us with only one electron of each atoms to consider: the one $3s$ electron of Na and the $3p$ electron in Cl .

The same logic we used for homonuclear diatomic molecules tells us that when a stable structure is produced these two electrons will be concentrated between the two

⁶See Section 19.2 for details on para- and diamagnetic behavior.

atoms. However since the nuclear charges differ the electronic configuration will not be symmetric, but it will be displaced toward the Cl nucleus, which produces a larger attractive field. The uneven charge distribution leads to an electric dipole moment measured $3.0 \cdot 10^{-29} \text{ C m}$. Were the $3s$ electron of Na completely transferred to the Cl atom, the dipole moment would be $4.0 \cdot 10^{-29} \text{ C m}$. We conclude that about 75% of the valence electron of Na is displaced toward the Cl atom, i.e. the probability distribution of the electrons is higher nearer to the Cl nucleus. Still, we may consider the $NaCl$ molecule as being composed of two ions held together by Coulomb attraction, so we may write it as $Na^+ Cl^-$. This type of bond is called an *ionic bond*. However Fig. 8.9 shows that a molecule composed of Na^+ and Cl^- ions would have a different energy vs distance curve. The reason the potential energy rises to infinity as the distance of the atoms (or ions) decreases is that it would require the wave functions (orbitals) of the closed shells of the two atoms overlap, which would violate the exclusion principle. Consequently very large electrostatic fields will arise that prevent this overlap.

The bonding in the majority of the heteronuclear molecules is neither purely covalent, nor purely ionic. E.g. the bond in CO is mostly covalent with an electric dipole moment of $4.0 \cdot 10^{-31} \text{ C m}$.

8.4 Polyatomic molecules.

When a molecule contains more than two atoms the geometrical arrangement of the nuclei and the electrons, i.e. the *molecular symmetry* becomes important. The shape of *polyatomic molecules* is determined by this symmetry, which is reflected in the shape of the molecular orbitals. We build the molecular orbital from superposition of atomic wave functions. Our guideline can be stated as:

Important 8.4.1. *A bond between two atoms occurs in the direction in which the representative atomic wave functions making up the molecular orbital are concentrated or overlap. The strength of the bond depends on the degree of overlap.*

Let us take the example of the water molecule (H_2O): it contains 10 electrons and 3 nuclei. (see Fig. 7.3)

In the first approximation we disregard all but the unpaired electrons, which, for the O atom leaves only the two unpaired p electrons on the L shell to consider. The spin part of these atomic wave functions is parallel, according to Hund's rule, which requires different spatial wave functions for the two unpaired electrons, according to the Pauli principle. Select these to be p_x and p_y (see Fig. 6.6). Then the other two paired electrons must occupy the state p_z . The two unpaired electrons are on two orbitals perpendicular to each other. The unpaired electrons of the hydrogen atoms will pair with them and the largest overlap will be in the direction of the p_x and p_y orbitals. So in first approximation the H_2O molecule should have a right angle state as in Fig 8.10.

The measured angle between the bonds in a water molecule is however larger, namely 104.5° . This is because we neglected the repulsion of the H atoms. Detailed calculations shows that the $1s$ electrons of hydrogen are pulled toward the oxygen atom, which produces a polarized molecule with a resultant electric dipole momentum of $6.2 \cdot 10^{-30} \text{ C m}$ along the line bisecting the bond angle.

The next example is the ammonia (NH_3) molecule. Its electronic structure is in Fig. 7.3 and Fig. 8.11. The N atom has three unpaired $2p$ electrons concentrated along the x-, y- and z-axes occupying all three p states forming a pyramidal structure with the N atom at one vertex and the three H atoms in the other vertices. The resulting dipole moment is $5.0 \cdot 10^{-30} \text{ C m}$

8.5 Hydrocarbon molecules. Hybridization.

As everyone knows carbon is the most important atom for the life on Earth. This is a result of the wonderful quantum mechanical behavior of the C atom. In its ground state (see Fig. 7.3) it has only two unpaired $2p$ electrons which is not enough to explain many carbon compounds. However its first excited state has one $2s$ and three unpaired $2p$ electrons and it can be used to explain its role in molecules like CH_4 , where the 4 electrons are identically bound to the C atom. In this state the 3 unpaired $2p$ electrons act like to those in ammonia, but the $2s$ electron has spherical symmetry therefore cannot produce a bond of the same strength! The energy of the $2s$ and the $2p$ electrons of carbon are slightly different, but from their linear combination we can create a set of four *hybridized* directional wave functions which have the same energy. This technique is called *hybridization* and it has different variations⁷.

In the case of ammonia it is called *sp³ hybridization*, and the four wave functions are

$$\psi_1 = \frac{1}{2}(s + p_x + p_y + p_z) \quad (8.5.1a)$$

$$\psi_2 = \frac{1}{2}(s + p_x - p_y - p_z) \quad (8.5.1b)$$

$$\psi_3 = \frac{1}{2}(s - p_x + p_y - p_z) \quad (8.5.1c)$$

$$\psi_4 = \frac{1}{2}(s - p_x - p_y + p_z). \quad (8.5.1d)$$

Because the s and p functions have different angular momenta these hybridized functions do not describe states with well defined angular momentum. The four *sp³ hybrid* form a tetrahedral structure. (Fig. 8.12 (a)). This means that ammonia is also tetrahedral (Fig. 8.12 (b)). The ethane (H_3C-C-H_3) molecule also contains *sp³ hybrids*, but in that

⁷Although we called hybridization a computational “technique” it is a measurable phenomena as well.

case two of them overlap and this is what holds the two carbon atom together (Fig. 8.12 (c)). Because this resembles the σ orbitals in diatomic molecules it is called a σ bond.

Naturally carbon is not the only element and hydrocarbons are not the only molecules that show hybridization. neither sp^3 is the only hybrid molecular orbital possible. For instance N^+ has the same electronic structure as carbon and the $N^+ H_4$ molecule ion is similar in geometry to that of methane. Two other possible hybrids (sp^2 and sp) are discussed in Appendix 22.11, while Appendix 22.12 is about *conjugated molecules*.

8.6 Rotation and vibration of molecules.

Up till now we considered molecules as rigid structures with fixed position nuclei. But the nuclei may rotate as a whole rigid structure or their relative positions in the molecule may change, for instance when the molecule vibrates.

8.6.1 Rotation of diatomic molecules

For diatomic molecules, one of the *principal axes of inertia* goes through the two nuclei, which we select as our z axis, while the other two are perpendicular to this and to each other and all three go through the center of mass of the molecule. In this coordinate system the three principal values of the *moment of inertia* are Θ_x, Θ_y and Θ_z . Because of the small size and mass of the nuclei the Θ_z component can be taken to be zero. And because of symmetry the two components of the moment of inertia for any axis perpendicular to the z axis will be the same, which we will denote simply by Θ . If the distance of the nuclei is r_o

$$\Theta = \frac{m_1 m_2}{m_1 + m_2} r_o^2 \quad (8.6.1)$$

The kinetic energy, which in our case equals to the rotational energy of the molecule, then can be written as

$$\mathcal{E}_{rot} = \frac{1}{2} \Theta \omega^2$$

and introducing the angular momentum with $L = \Theta \omega$

$$\mathcal{E}_{rot} = \frac{L^2}{2\Theta} \quad (8.6.2)$$

But according to quantum mechanics the angular momentum of the molecule is quantized:

$$L = \sqrt{\ell(\ell+1)} \hbar, \quad \ell = 0, 1, 2, \dots \quad (8.6.3)$$

therefore

$$\mathcal{E}_{rot} = \frac{\hbar^2}{2\Theta} \ell(\ell + 1) \quad (8.6.4)$$

where $\frac{\hbar^2}{2\Theta} \approx 10^{-4} \text{ eV}$, which is much smaller than the thermal energy at room temperature, which at 300 K is 0.0258 eV. Consequently many molecules are excited to rotational levels even at room temperature. The distance of two consecutive levels

$$\Delta \mathcal{E}_{rot} = \frac{\hbar^2}{\Theta} (\ell + 1) \quad (8.6.5)$$

is of the same magnitude. Because the selection rule in this case is $\Delta \ell = \pm 1$ the corresponding frequency spectrum ($h\nu = \Delta \mathcal{E}_{rot}$) contains equidistant lines with a frequency difference of $\Delta \nu = \hbar/2\pi\Theta$ as shown in Fig. 8.13. Each dip in the spectrum corresponds to a resonance absorption. Purely rotational spectra lie in the microwave or far infrared range of the electromagnetic spectrum. For a molecule to have purely rotational spectrum a constant electrical dipole moment is required. During absorption this interacts with the electromagnetic field. Therefore homonuclear molecules do not have rotational spectra.

As the rotational energy increases the shape of the molecule will be affected by this rotation and correct calculations require corrections to our formulas.

8.6.2 Vibration of molecules

If you have a look at the potential energy curve of e.g. the $NaCl$ molecule in Fig. 8.9 you can easily realize that the two nuclei in the molecule will vibrate around the equilibrium distance r_0 . The ground state, which will not be the minimum of the potential but the zero point energy of the molecular oscillator. Near the equilibrium distance the potential is approximately parabolic, therefore the vibrational levels can be approximated by the formula, which is strictly valid for harmonic oscillators:

$$\mathcal{E}_v = \left(v + \frac{1}{2}\right) \hbar\omega_o \quad \text{where } v = 0, 1, 2, \dots \quad (8.6.6)$$

In reality the potential is anharmonic so this is just an approximation. Because of the zero point energy, the dissociation energy is not the energy difference denoted by D , but $D - \mathcal{E}_o$. The selection rule is the same $\Delta v = \pm 1$ as it was for the linear harmonic oscillator⁸.

The value of $\hbar\omega_o$ is about 0.1 - 0.5 eV, so the vibrational transitions are in the infra-red region. But it is also possible that higher harmonics of the base vibrational

⁸In reality this selection rule is not so strict as the potential is not harmonic, but higher Δn values are still improbable.

modes can be observed. These harmonics may fall into the visible range. For instance the intrinsic blue color of clear water is the result of an absorption at the harmonic $\tilde{\omega} = \tilde{\omega}_1 + 3\tilde{\omega}_3 = 14\,318\text{cm}^{-1}$, which corresponds to a wavelength of 698 nm and looks red for us. When red light is absorbed the remaining light becomes turquoise blue. However the absorption of the third overtone is very weak, so a larger body of water is needed to make this color visible. This is what makes glaciers and icebergs blue and adds to the color of big bodies (lakes, seas, oceans) of water⁹.

The total energy of a system is the sum of the electron energies plus the vibrational and rotational energies:

$$\mathcal{E} = \mathcal{E}_e + \mathcal{E}_r + \mathcal{E}_v = \mathcal{E}_e + \left(v + \frac{1}{2}\right) \hbar\omega_o + \frac{\hbar^2}{2\Theta} \ell(\ell + 1) \quad (8.6.7)$$

When a molecule vibrates its shape and moment of inertia change, changing the rotational frequencies. This is the *vibration-rotation interaction effect*.

8.6.3 Vibration of polyatomic molecules

For polyatomic molecules the situation is even more complicated. The behavior of such systems can be described using so called *normal (vibrational) modes*. In normal modes all nuclei vibrate and the relative phases of the vibrations are constant. Consequently discrete frequencies can be associated with the normal modes. These are determined by the geometry of the molecule and because of molecular symmetries some of these frequencies may be degenerate, i.e. may belong to more than one normal mode. In Fig. 8.15 some normal vibrational modes of the CO_2 and water molecules can be seen.

The corresponding vibrational wave numbers are¹⁰: $\tilde{\omega}_1 = 1337\text{cm}^{-1}$, $\tilde{\omega}_2 = 667\text{cm}^{-1}$, $\tilde{\omega}_3 = 2349\text{cm}^{-1}$ for CO_2 and $\tilde{\omega}_1 = 3657\text{cm}^{-1}$, $\tilde{\omega}_2 = 1595\text{cm}^{-1}$, $\tilde{\omega}_3 = 3756\text{cm}^{-1}$ for H_2O .

8.6.4 Franck-Condon principle, absorption and emission for molecules

Fig. 8.16 combines the properties of molecular rotational and vibrational energy levels. The equilibrium distances in different states differ. Typical separation between two electron states $\mathcal{E}'' - \mathcal{E}'$ is about 1 - 10 eV, which means that the frequency of the radiation emitted or absorbed by the molecule is in the visible or ultraviolet range. Each electronic state contains many vibrational states, while to each of these there correspond several rotational states. The total energy in any state is given by (8.6.7). So the energy change

⁹The color of water also depends on the materials solved in it. This can make water greener or bluer in some areas.

¹⁰ $\tilde{\omega} \equiv \frac{2\pi\nu}{c} = \frac{2\pi}{\lambda}$ is the wave number used in spectroscopy, which is called there “angular frequency”.

in a transition is

$$\Delta \mathcal{E} = \mathcal{E}'' - \mathcal{E}' = \Delta \mathcal{E}_e + \Delta \mathcal{E}_v + \Delta \mathcal{E}_r = \quad (8.6.8)$$

$$= \Delta \mathcal{E}_e + (v'' + \frac{1}{2}) \hbar \omega'' - (v' + \frac{1}{2}) \hbar \omega' + \quad (8.6.9)$$

$$+ \frac{\hbar^2}{2\Theta''} \ell'' (\ell'' + 1) - \frac{\hbar^2}{2\Theta'} \ell' (\ell' + 1) \quad (8.6.10)$$

But this means that the frequency $\nu = \mathcal{E}/h$ can be written as a sum of three terms:

$$\nu = \frac{\mathcal{E}}{h} = \nu_e + \nu_v(v'', v') + \nu_r(\ell'', \ell) \quad (8.6.11)$$

As a result the observable spectrum consists of a series of *bands*, where each band corresponds to the possible values of v', v'' and ℓ', ℓ'' . The selection rules in this case can be stated for all types of transitions. For rotation

$$\Delta \ell = 0, \pm \ell, \quad \text{except } (\ell'' = 0) \leftrightarrow (\ell' = 0) \quad (8.6.12)$$

Note that now even $\Delta \ell = 0$ is allowed, because there can be a change in configuration of the molecule during the transition, except when both ℓ'', ℓ' are 0s. The spin of the photons is 1 and in this case it would be impossible to satisfy the conservation law of the angular momentum.

Because spin dependent forces involved in the electronic transition are not strong enough to change the spin of the electrons, therefore for ν_e , generally $\Delta S = 0$.

When a molecule absorbs or emits a photon the vibrational and electronic state of the molecule changes simultaneously. These transitions are sometimes called *vibronic*. Because the characteristic times of electronic transitions ($\approx 10^{-16}$ s) are much shorter than those of vibrational transitions ($\approx 10^{-13}$), during an electronic transition the actual nuclear separation is essentially constant. Consequently:

Important 8.6.1. *If the molecule is to move to a new vibrational level during the electronic transition, this new vibrational level must be instantaneously compatible with the nuclear positions and momenta of the vibrational level of the molecule in the original electronic state. This is the Franck-Condon principle.*

Fig. 8.17 shows a Franck-Condon energy diagram where the transition occurs between states $v' = 0$ and $v'' = 2$. As you see transitions do not conform to the usual selection rules of $\Delta v = \pm 1$.

8.6.5 Scattering of light by molecules. Rayleigh and Raman scattering

Imagine a gas sample illuminated with a monochromatic electromagnetic radiation of frequency ν . When $h\nu$ is not equal to any of the $\mathcal{E}'' - \mathcal{E}'$ energy level difference the

radiation can not be absorbed but it will be *scattered* by the gas molecules. There are two kinds of scattering.

The part scattered *elastically* has the same ν frequency as the illuminating radiation. This called *Reyleigh scattering* or *coherent scattering*.

Important 8.6.2. *Reyleigh scattering occurs every time when light or other electromagnetic radiation is scattered by particles much smaller than their wavelength. The particles may be individual atoms or molecules. It can occur when light travels through transparent solids and liquids and in gases.*

Rayleigh scattering causes the blue hue of the sky and the reddening of the Sun at sunset.

The other part of the scattered radiation has a frequency of $\nu' = \nu \pm \nu_v$, where ν_v is a frequency of the vibrational spectrum of the molecule. This is an inelastic scattering of the radiation and called *Raman scattering*

Important 8.6.3. *Raman scattering occurs when a molecule is excited by the non-resonant incoming radiation to some vibrational level which is above or below of the level corresponding to ν because of the selection rule $\Delta\nu = \pm 1$. When the molecules absorb energy, the energy of the scattered photons will be lower than that of the incoming photons. This is called Stokes Raman scattering. When the molecules lose energy in the emission the emitted photons will have larger energy than the incoming ones. This is called anti-Stokes Raman scattering.*

Raman scatterings allows e.g. measurement of rotation of homonuclear diatomic molecules.

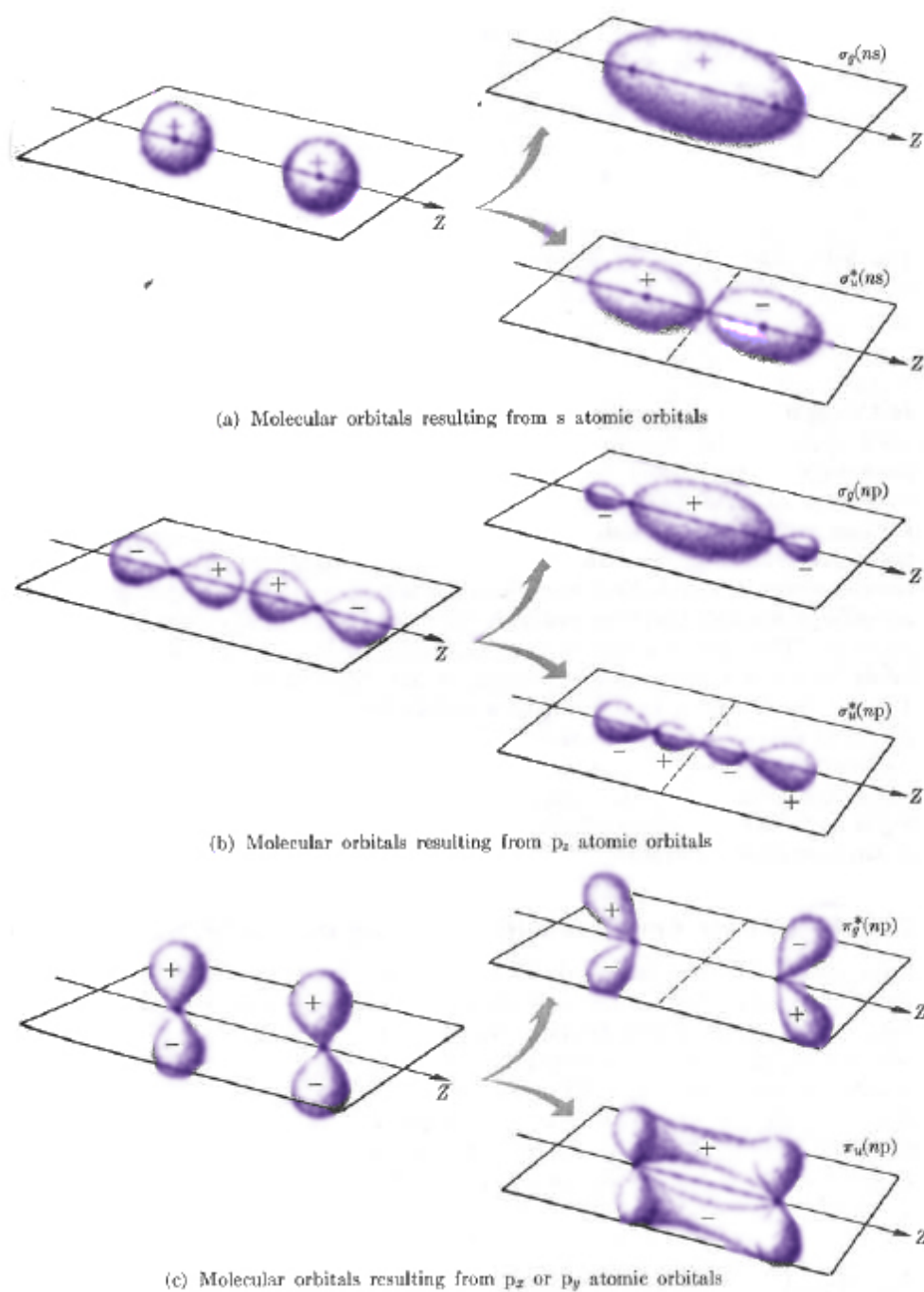


Figure 8.7: Angular distribution of molecular orbitals formed from atomic orbitals in diatomic molecules.

Molecule	Configuration								Dissociation energy, eV	Bond length, Å	Ground state
	$\sigma_g 1s$	$\sigma_u^* 1s$	$\sigma_g 2s$	$\sigma_u^* 2s$	$\pi_u 2p$	$\sigma_g 2p$	$\pi_u^* 2p$	$\sigma_u^* 2p$			
H_2^+	\uparrow								2.65	1.06	$^2\Sigma_g$
H_2	$\uparrow\downarrow$								4.48	0.74	$^1\Sigma_g$
He_2^+	$\uparrow\downarrow$	\uparrow							3.1	1.08	$^2\Sigma_u$
He_2	$\uparrow\downarrow$	$\uparrow\downarrow$							Not stable		$^1\Sigma_g$
Li_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$						1.03	2.67	$^1\Sigma_g$
Be_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$					Not stable		$^1\Sigma_g$
B_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow$				3.6	1.59	$^3\Sigma_g$
C_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$				3.6	1.31	$^1\Sigma_g$
N_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$	$\uparrow\downarrow$			7.37	1.09	$^1\Sigma_g$
O_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow$		5.08	1.21	$^3\Sigma_g$
F_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$		2.8	1.44	$^1\Sigma_g$
Ne_2	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$	$\uparrow\downarrow$	$\uparrow\uparrow\downarrow\downarrow$	$\uparrow\downarrow$	Not stable		$^1\Sigma_g$

Figure 8.8: Electronic configuration of some homonuclear diatomic molecules

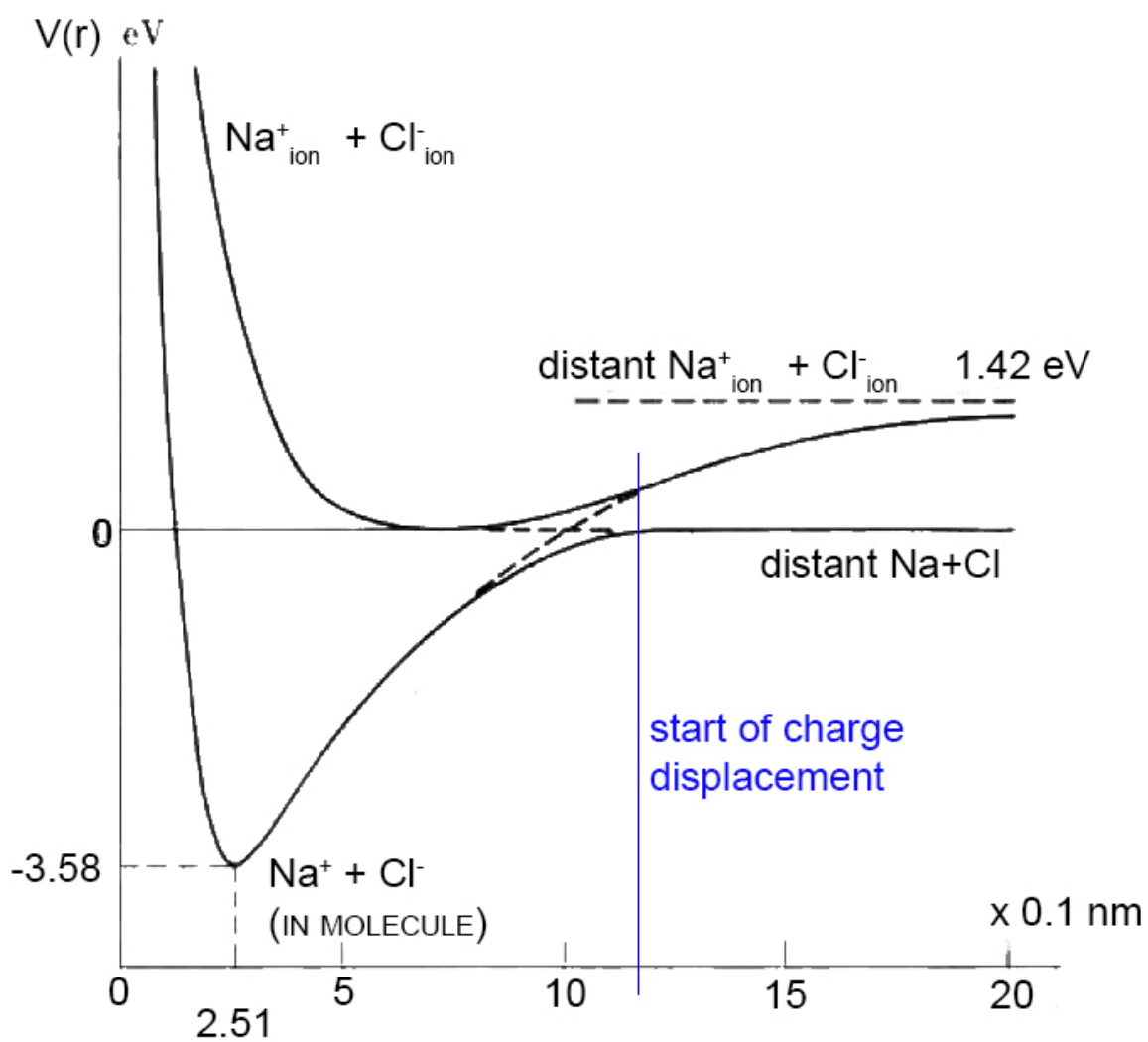


Figure 8.9: Potential energy curves vs distance for a NaCl molecule and $\text{Na}^+ + \text{Cl}^-$ ions.

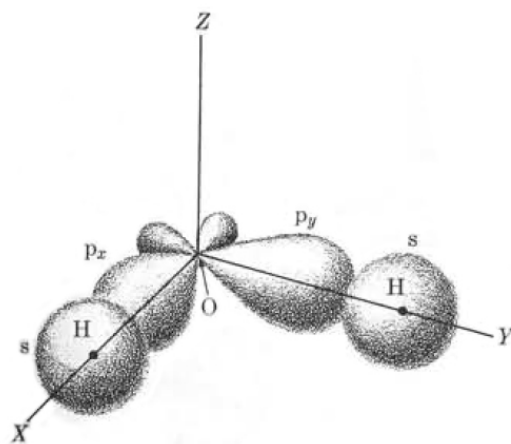


Figure 8.10: Angular part of the wave function in a water molecule in first approximation. The shape of the p orbitals is distorted by the presence of the H atom. The H-H repulsion is neglected.

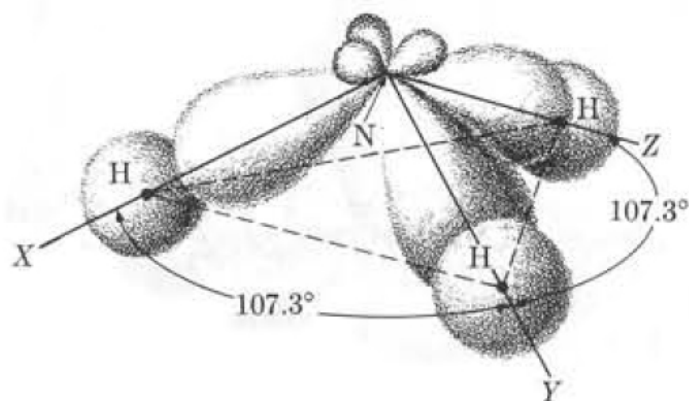


Figure 8.11: Angular part of the wave function in an NH_3 molecule.

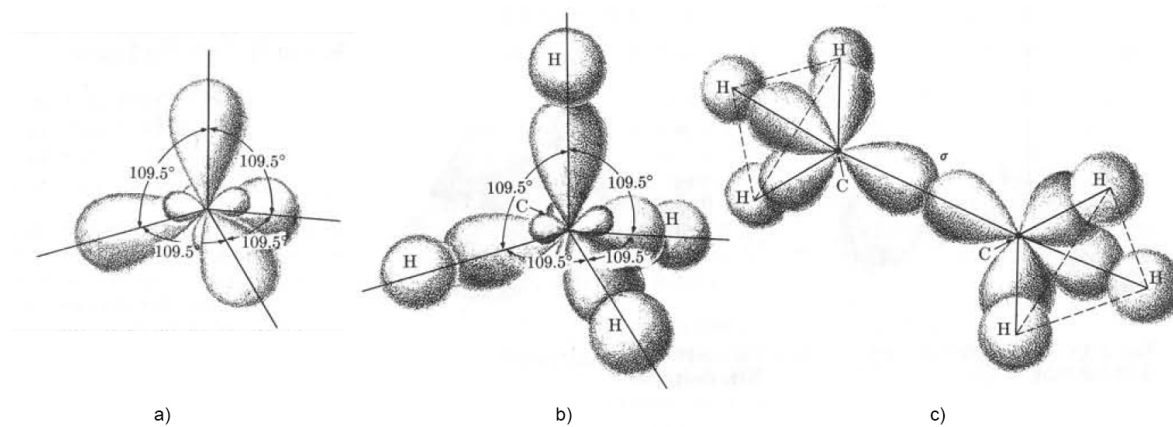


Figure 8.12: sp^3 hybridization. a) the 4 identical sp^3 electron orbitals in carbon, b) in methane, c) in ethane.

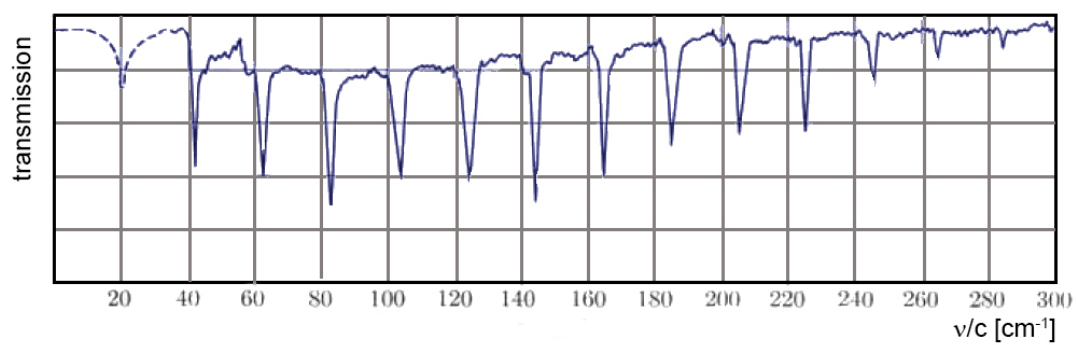


Figure 8.13: Rotational absorption spectrum of HCl in the gaseous phase.

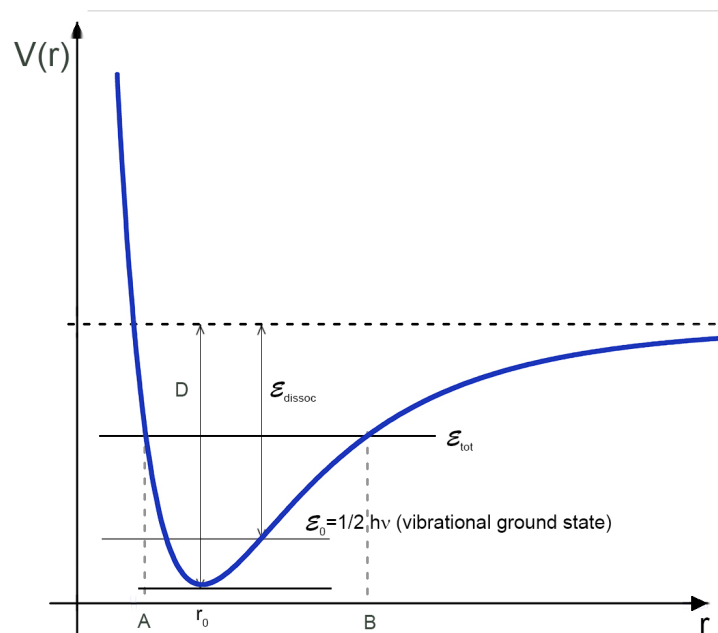


Figure 8.14: Schematic potential in a diatomic molecule.

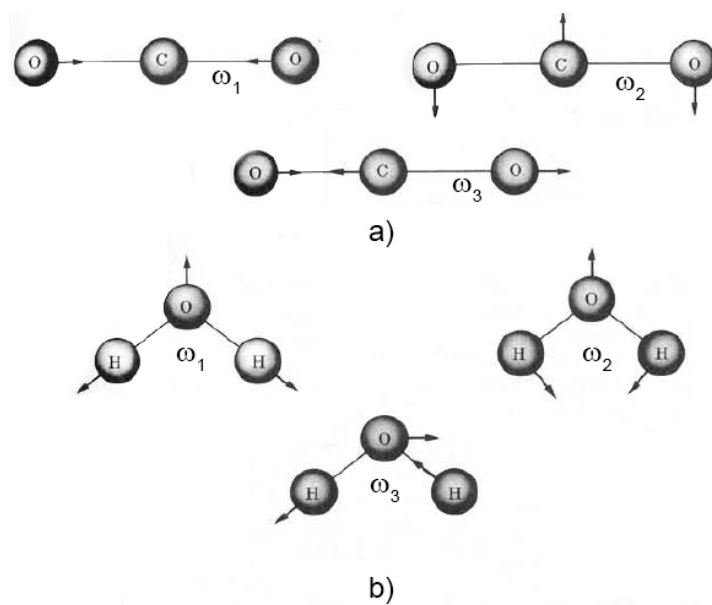


Figure 8.15: Normal vibrational modes in CO_2 and H_2O

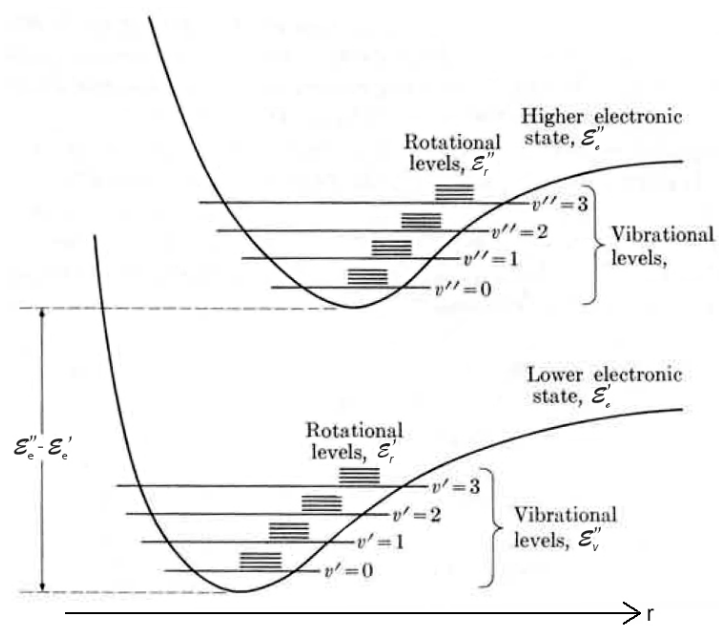


Figure 8.16: Vibrational and rotational energy levels associated with two electronic states.

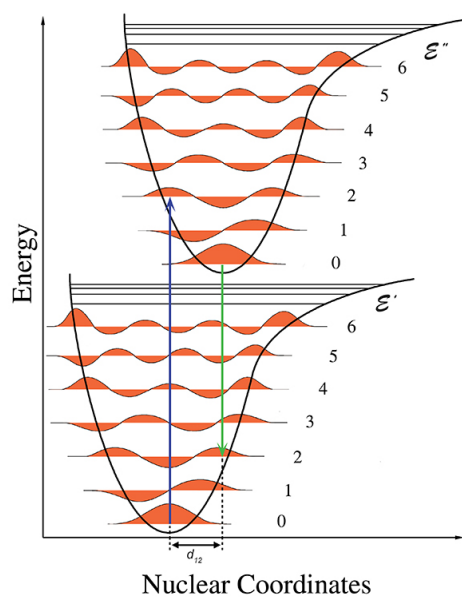


Figure 8.17: Franck-Condon energy diagram.

Chapter 9

Statistical physics.

9.1 Statistical equilibrium.

At standard temperature ($0^\circ C$) and pressure ($1atm$) 1 liter of any ideal gas contains $n = 6.022 \cdot 10^{23} / 22.41 l = 2.7 \cdot 10^{22}$ molecules. It is completely impossible to solve even the equations of motion of classical physics for so many molecules, let alone their Schrödinger equation. This is the area where statistical physics is used. We will use it to determine the possible states and their probabilities for N particle systems.

Let us suppose we have N identical (distinguishable or indistinguishable) particles with discrete or continuous energy levels $\mathcal{E}_1, \mathcal{E}_2, \dots$. If the number of particles on level \mathcal{E}_i is n_i , then $N = \sum_i n_i$ and the total energy $U \equiv \mathcal{E}_{tot} = \sum_i n_i \mathcal{E}_i$. The particles may or may not interact with each other. If they interact this interaction will modify the possible energies. To simplify things let us suppose that the interaction of the particles can be described by an average V_i^{aver} potential, i.e. \mathcal{E}_i will be replaced by $\mathcal{E}_i + V_i^{aver}$.

The actual state of the system then can be characterized by the set $\{n_i\}$ of n_i numbers: $\{n_i\} \equiv \{n_1, n_2, \dots\}$. This *particle distribution* or *partition* determines macroscopically observable physical properties of the system.

If there is an energy exchange between the particles (this may be in the form of collisions or caused by electromagnetic forces) the state of even closed systems (where $U = const$) may change in time because the interaction may modify the $\{n_i\}$ distribution.

For example in an equilibrium classical ideal gas the energy (or velocity) distribution of the particles is constant, therefore the pressure of the gas does not change macroscopically, but there are small fluctuations in the momentary pressure, which become observable only at very low gas densities.

Important 9.1.1. *If the change of the particle distribution $\{n_i\}$ is random and cause no macroscopically relevant changes, the system is in equilibrium. The corresponding macroscopically observable states are called macrostates. The possible configurations that result the same macrostate are called microstates.*

In many-particle systems equilibrium is statistical. In equilibrium the small fluctuations in the n_i values average out.

Usually every possible microstate is considered equally probable. The number of microstates for a given macrostate, called the statistical weight, however may differ, therefore the probability of different macrostates may be different. Non-equilibrium macrostates have a much smaller probability than states near the equilibrium. More than one macrostates may have the same or very nearly the same probability. The equilibrium distributions will be the most probable ones.

The $\{n_i\}$ distribution does not determine the configuration of the particles completely, because it just requires that n_i particles are on the \mathcal{E}_i level, but does not say which particles are among those. E.g. if the particles are distinguishable then the following two configurations are equivalent in $\{n_i\}$:

- particle “A” is at level “i” while particle “B” is at level “j” and
- “B” is at level “i” while particle “A” is at level “j”

In contrast with this there will be only one possibility for indistinguishable particles: one of it in the i -th, an other one in the j -th state.

9.2 Maxwell-Boltzmann distribution.

To describe the behavior of a classical gas we can use the following assumptions:

- the particles are distinguishable¹ (i.e. these are classical particles)
- the probability of the occupation of every energy level \mathcal{E}_i is equal
- one level may be occupied by any number of particles (no exclusion principle)
- the probability of a distribution (or *partition*) $\{n_i\}$ is proportional to the number of particle configurations that can realize it.
- thermal equilibrium corresponds to the maximum probability distribution

\mathcal{E}_1 is occupied by n_1 particles. This may happen in $\binom{N}{n_1}$ different ways². For each of these there are $\binom{N-n_1}{n_2}$ possibilities to fill \mathcal{E}_2 . For every possible configuration on \mathcal{E}_1

¹See section 6.7 for a short introduction between distinguishable and indistinguishable particles.

² $\binom{N}{m} \equiv \frac{N!}{m!(N-m)!}$. Also observe that if a level is unoccupied then $n_i = 0$, and $0! = 1$.

and \mathcal{E}_2 there are $\binom{N-n_1-n_2}{n_3}$ possibilities to fill \mathcal{E}_3 , etc. The total number of possible configurations therefore³

$$w = \binom{N}{n_1} \binom{N-n_1}{n_2} \binom{N-n_1-n_2}{n_3} \dots = \frac{N!}{n_1! n_2! n_3! \dots} \quad (9.2.1)$$

When the probabilities of occupation are different for different \mathcal{E}_i s, i.g. an energy level may be compatible with more different angular momentum states than the others, then it is more likely to be occupied, this formula should be modified to become⁴

$$w = \binom{N}{n_1} g_1^{n_1} \binom{N-n_1}{n_2} g_2^{n_2} \binom{N-n_1-n_2}{n_3} g_3^{n_3} \dots = \frac{N! g_1^{n_1} g_2^{n_2} g_3^{n_3} \dots}{n_1! n_2! n_3! \dots} = N! \prod_i \frac{g_i^{n_i}}{n_i!} \quad (9.2.2)$$

where the $g_i (\geq 1)$ numbers give the degeneracy of each energy level.

We calculated so far the number of possible ways the given partition can occur. If we are interested in the *probability* of this partition⁵ then we must divide it with the number of all possible particle permutations, i.e. with $N!$. So the probability of a given partition in the general case:

$$\mathcal{P}(\{n_i\}) = \prod_i \frac{g_i^{n_i}}{n_i!} \quad (9.2.3)$$

The next step is to calculate the maximum of this probability. However we are looking for a maximum with the following additional conditions: the total number of particles $N_{tot} = \sum_i n_i$ is constant and the total energy $\mathcal{E}_{tot} = \sum_i \mathcal{E}_i n_i$ is also constant. This is called a *conditional maximum* problem. The result⁶:

³If we write the product of binomial coefficients using factorials we can observe that the denominator of a factor in this product is the same as the numerator of the next factor, so these cancels each other out. This leads to (9.2.1).

We should have arrived to the same formula noting N particles may be ordered $N!$ times (permutations), and because we are not interested in the order of the particles on a given level this number must be divided by the product of all $n_i!$.

⁴If level \mathcal{E}_i has g_i possible sub-states and we have n_i particles on this level then each particle may be put into any of these g_i sub-states with the same probability, therefore the number of possible sub-states is $g_i^{n_i}$.

⁵probability of an event (e.g. of a partition) \approx number of ways it can happen divided by the total number of possible outcomes.

⁶Mathematical details are in an appendix: Appendix 22.13.

Important 9.2.1.

$$\mathcal{P}_{max,i} = \frac{1}{Z} g_i e^{-\beta \mathcal{E}_i}, \quad \text{where} \quad (9.2.4)$$

$$\beta = \frac{1}{k_B T} \quad (9.2.5)$$

$$Z(T) = \sum_i g_i e^{-\beta \mathcal{E}_i} \quad (9.2.6)$$

and $k_B = 1.38 \cdot 10^{-23} \text{ J/K}$ is the Boltzmann constant, T is the absolute temperature (in K) and Z is called the partition function.

The factor $e^{-\beta \mathcal{E}_i} = e^{-\frac{\mathcal{E}_i}{k_B T}}$ that gives the (unnormalised) relative probability of a state (i.e. the statistical weight) is called the Boltzmann factor.

Knowing the probability of a given macrostate we can easily calculate the *expectation value* of n_i ; i.e. the value we expect as the average of the results of many successive measurements of n_i at a given \mathcal{E}_i

$$\langle n_i \rangle = N \cdot \mathcal{P}_{max,i} = \frac{N}{Z} g_i e^{-\beta \mathcal{E}_i} \quad (9.2.7)$$

This is the *Maxwell-Boltzmann statistics*.

Important 9.2.2. For any energy dependent physical quantity $\mathcal{F}(\mathcal{E})$ the average value is the n_i weighted average of possible $\mathcal{F}(\mathcal{E}_i)$ values :

$$\langle \mathcal{F} \rangle \equiv \mathcal{F}_{aver} = \frac{1}{N} \sum_i n_i \mathcal{F}(\mathcal{E}_i) = \frac{1}{Z} \sum_i g_i \mathcal{F}(\mathcal{E}_i) e^{-\beta \mathcal{E}_i} \quad (9.2.8)$$

To calculate the total (internal) energy U of a system we can set $\mathcal{F}(\mathcal{E}_i) = \mathcal{E}_i$:

$$U = \frac{N}{Z} \sum_i g_i \mathcal{E}_i e^{-\beta \mathcal{E}_i} \quad (9.2.9)$$

Note that in this case the sum is the negative derivative of the (9.2.6) partition function with respect to β :

$$\sum_i g_i \mathcal{E}_i e^{-\beta \mathcal{E}_i} = -\frac{d}{d\beta} \sum_i g_i e^{-\beta \mathcal{E}_i} = -\frac{dZ}{d\beta},$$

therefore

$$U = \frac{N}{Z} \frac{dZ}{d\beta} = -N \frac{d}{d\beta} \ln Z \quad (9.2.10)$$

Furthermore $\frac{d}{d\beta} = \frac{d}{dT} \frac{dT}{d\beta} = -k_B T^2 \frac{d}{dT}$, so

$$U = N k_B T^2 \frac{d}{dT} \ln Z \quad (9.2.11)$$

The average energy of a particle is

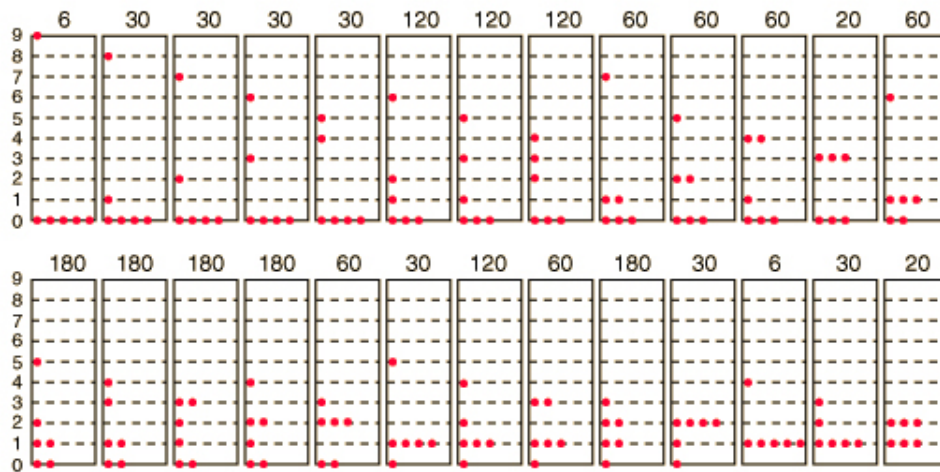
$$\mathcal{E}_{aver} = k_B T^2 \frac{d}{dT} \ln Z \quad (9.2.12)$$

therefore the temperature of a system is determined by the average energy per particle and the structure of the energy levels of the system described in Z .

Example 9.1. *In a system with equidistant energy levels how many ways can you distribute 9 units of energy among 6 identical, distinguishable particles? The energy of the ground state ($i=0$) is 0, and the levels are one unit of energy distant from each other.*
Solution In this case the observable different *macrostates* give the number of particles on every level, while the *microstates* are the possible ways to achieve a given macrostate.

Because we must distribute 9 units of energy among the particles and the energy of the ground state is 0, we have to use 10 energy levels.

The number of the macrostates are few so they can easily be counted in this case and the result is 26. The figure shows all macrostates with a total energy of 9 units, together with the number of the microstates that correspond to the same macrostate. The first macrostate in the first row have $\frac{6!}{6!} = 6$ microstates, the second one $\frac{6!}{4! 1! 1!} = 30$, while the first one in the second row have $\frac{6!}{2! 2! 1! 1!} = 180$, etc. The total number of microstates is 2002.



Example 9.2. Graph the distribution for Problem 9.1 and compare it with the Maxwell-Boltzmann distribution function!

Solution We have to graph the n_i vs \mathcal{E} discrete function of Problem 9.1. The average occupation numbers for the levels are

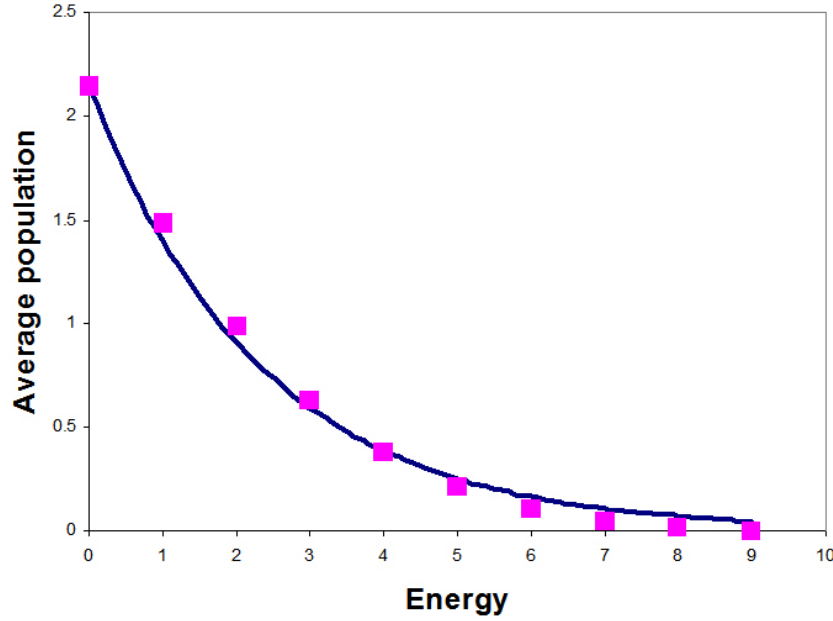
$$\langle n_i \rangle = \frac{\sum_{n=1}^{26} w_i(n) n_i(n)}{\sum_n^{26} w_i}$$

where the summation goes from 1 to the number of all possible macrostates, and $w_i(n)$ is the number of the microstates that results in the n -th macrostate. The denominator is the total number of microstates, which is 2002 as we have shown previously in Problem 9.1. So for instance for $i = 0$

$$n_0 = \frac{6 \cdot 5 + 4 \cdot 30 \cdot 4 + (3 \cdot 120 + 3 \cdot 60 + 20) \cdot 3 + (2 \cdot 60 + 4 \cdot 180) \cdot 2}{2002} + \frac{(30 + 120 + 60 + 180 + 30) \cdot 1 + (30 + 6 + 30 + 20) \cdot 0}{2002} = 2.143$$

The average occupation numbers or *average population* of the levels:

Energy level	0	1	2	3	4	5
$\langle n_i \rangle$	2.143	1.484	0.989	0.629	0.378	0.210
Energy level	6	7	8	9		
$\langle n_i \rangle$	0.104	0.045	0.015	0.003		



while in the figure you can see the results compared to that of the continuous Maxwell-Boltzmann distribution function.

As you can see the distribution for even as few as 6 particles closely approximates the Maxwell-Boltzmann distribution function.

9.2.1 Application of the Maxwell-Boltzmann statistics to the ideal gas

Even though we are talking about a classical ideal gas molecular excitations are quantum mechanical processes with transitions between discrete energy levels. The “classical” nature here refers to the molecules themselves, which at not extremely large pressures and at not extremely low temperatures are distinguishable particles.

According to (9.2.4) the equilibrium ratio of the (average) number of particles on energy levels \mathcal{E}_i and \mathcal{E}_j is

$$\frac{n_i}{n_j} = \frac{\mathcal{P}_i}{\mathcal{P}_j} = \frac{g_i}{g_j} e^{-\frac{\mathcal{E}_j - \mathcal{E}_i}{k_B T}} \quad (9.2.13)$$

which depends exponentially on $\Delta \mathcal{E} \equiv \mathcal{E}_j - \mathcal{E}_i$. Table 9.1 summarizes this:

The kinetic energy levels of gas molecules in a container, whose size is very large compared to the size of the molecules are so very close to each other at normal temperatures and pressures, that we cannot observe the discreteness of the levels and consider

	$\Delta \mathcal{E} [\text{eV}]$	100 K	300 K	1000 K
rotational	10^{-4}	0.989	0.996	0.999
vibrational	$5 \cdot 10^{-2}$	$3 \cdot 10^{-3}$	0.150	0.560
electronic	3	$13 \cdot 10^{-164}$	$8 \cdot 10^{-49}$	$8 \cdot 10^{-16}$

Table 9.1: Excitation energies ($\Delta \mathcal{E}$) and excitation probabilities of rotational, vibrational and electronic transitions in molecules of an ideal gas at different temperatures.

the kinetic energy as a continuous quantity.

For this case let us introduce a continuous $f_{MB}(E)$ function, which gives the probability of occupation, so that the number Δn of particles in a $\Delta \mathcal{E}$ energy interval around \mathcal{E} is $N \cdot f_{MB}(\mathcal{E}) g(\mathcal{E}) \Delta E = \Delta n$, or equivalently: $f_{MB}(\mathcal{E}) = \frac{1}{N g(\mathcal{E})} \frac{dn(E)}{dE}$.

Important 9.2.3.

$$f_{MB}(\mathcal{E}) = \frac{1}{Z} e^{-\frac{\mathcal{E}}{k_B T}} \quad (9.2.14)$$

$$Z = \int_0^{\infty} g(E) e^{-\frac{\mathcal{E}}{k_B T}} d\mathcal{E} \quad (9.2.15)$$

is the Maxwell-Boltzmann distribution function.

where $g(\mathcal{E}) d\mathcal{E}$ is the number of possible states in the energy interval $d\mathcal{E}$ around \mathcal{E} .
The average value of any energy dependent physical quantity $\mathcal{F}(E)$ for one particle is:

$$\langle \mathcal{F} \rangle = \frac{1}{Z} \int_0^{\infty} \mathcal{F}(\mathcal{E}) \cdot g(E) \cdot f_{MB}(\mathcal{E}) d\mathcal{E} \quad (9.2.16)$$

Again using $\mathcal{F} = \mathcal{E}$ we find that (9.2.12) is still valid:

$$\mathcal{E}_{aver} \equiv \langle \mathcal{E} \rangle = k_B T^2 \frac{d}{dT} \ln Z$$

For a classical ideal gas enclosed in a large container of volume V the degeneracy of the energy states according to (3.5.20) (where we called it the density of states):

$$g(\mathcal{E}) = \frac{4\pi V \sqrt{2m^3}}{h^3} \sqrt{\mathcal{E}}$$

The partition function in (9.2.15):

$$Z = \frac{4\pi V \sqrt{2m^3}}{h^3} \int_0^\infty \sqrt{\mathcal{E}} e^{-\frac{\mathcal{E}}{k_B T}} d\mathcal{E} = \frac{V (2\pi m k_B T)^{\frac{3}{2}}}{h^3} \quad (9.2.17)$$

Therefore the internal energy of the gas is

$$U = N \mathcal{E}_{aver} = N k_B T^2 \frac{d}{dT} \ln \left[\frac{V (2\pi m k_B T)^{\frac{3}{2}}}{h^3} \right] = \frac{3}{2} N k_B T \quad (9.2.18)$$

$$\frac{dn}{d\mathcal{E}} = \frac{2\pi N}{(\pi k_B T)^{3/2}} \sqrt{E} e^{-\frac{E}{k_B T}} \quad (9.2.19)$$

For classical monatomic ideal gases the energy is only kinetic, that is independent of the mass of the molecule, can be expressed with the velocity of the molecules $\mathcal{E} = \frac{1}{2} m v^2$. Substituting into the Maxwell-Boltzmann distribution formula gives the number of particles in a unit energy interval as a function of the magnitude of the velocity:

$$\begin{aligned} dN(\mathcal{E}) &= g(\mathcal{E}) f_{MB}(E) d\mathcal{E} = dN(v) = g(\mathcal{E}(v)) f_{MB}(v) dv \\ f_{MB}(v) &= f_{MB}(E(v)) \frac{d\mathcal{E}}{dv} \\ f_{MB}(v) &= \frac{2\pi N}{(\pi k_B T)^{3/2}} \sqrt{\frac{1}{2} m v^2} e^{-\frac{m v^2}{2 k_B T}} (m v) \\ f_{MB}(v) &= N \frac{\sqrt{2} \pi (m)^{3/2}}{(\pi k_B T)^{3/2}} v^2 e^{-\frac{m v^2}{2 k_B T}} \end{aligned} \quad (9.2.20)$$

This *Maxwell-Boltzmann velocity distribution* function is shown in Fig. 9.1.

9.3 Quantum statistics.

If the particles are microscopic, i.e. the uncertainty principle is not negligible, their wave functions spread out and may overlap. In this case they cannot be distinguished even in principle. i.e. they are indistinguishable and classical statistics, therefore the Maxwell-Boltzmann distribution cannot be used. There are two types of indistinguishable particles *bosons* (e.g. photons) and *fermions*⁷:

⁷The names *boson* and *fermion* were coined by the English theoretical physicists *Paul Adrian Maurice Dirac* (e.g. electrons). The first one to commemorate the contribution of Indian physicist *Satyendra Nath Bose* to the statistical theory of these particles, while the second one came from the surname of the Italian theoretical physicist *Enrico Fermi*.

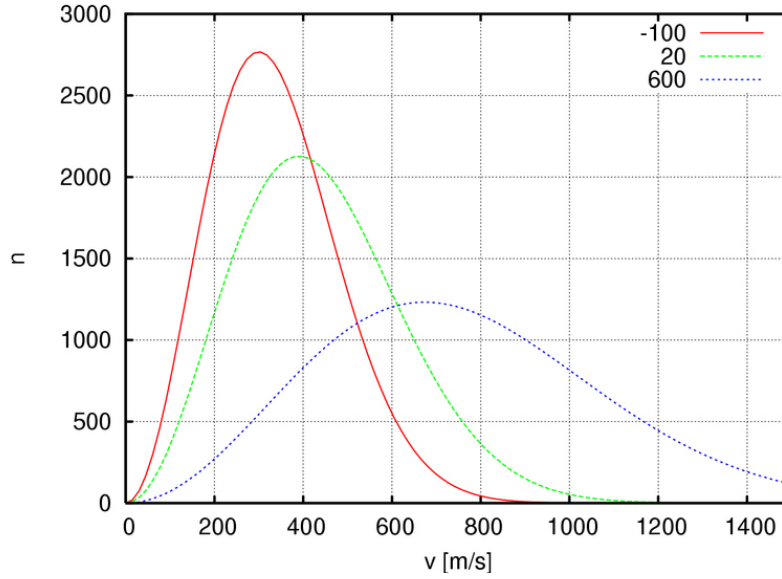


Figure 9.1: Velocity distribution of an ideal gas of 10^6 oxygen molecules at temperatures -100°C , 0°C and 600°C . The area below the curves are equal.

Important 9.3.1. • *For fermions the exclusion principle holds. Only a single fermion can occupy a non degenerate state. All half-integer spin particles (e.g. the electron) are fermions. The wave function of a system of fermions is anti-symmetric for the exchange of the coordinates of any two of the fermions:*

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k, \dots) = -\psi(\mathbf{r}_1, \mathbf{r}_k, \dots, \mathbf{r}_2, \dots)$$

- *Bosons are particles for which the exclusion principle is invalid. Any number of bosons can occupy the same non degenerate quantum state (described with the same quantum numbers). Bosons are integer spin particles. The wave function of a system of bosons is symmetric for the exchange of the coordinates of any two of the constituent bosons:*

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k, \dots) = \psi(\mathbf{r}_1, \mathbf{r}_k, \dots, \mathbf{r}_2, \dots)$$

Fermions and bosons show widely different behavior. This is most readily seen at $T = 0\text{ K}$, where with no thermal excitations a system will be in its lowest possible energy state.

Because of the exclusion principle every *fermion* will be in the lowest lying energy level not already occupied by an other fermion, therefore fermions will occupy all possible energy states (or energy levels) up to a certain energy called the *Fermi-energy* and denoted by \mathcal{E}_F and all states above \mathcal{E}_F will be empty.

Because there is no exclusion principle for *bosons* in the ground state at absolute zero all of them will be in the same lowest energy state. As the temperature goes above zero any or all of them can be excited to any level with $\mathcal{E} \geq k_B T$. The spread of the bosons to the energy levels will naturally depend on the temperature⁸.

9.4 Fermi-Dirac distribution.

Because of the exclusion principle if the degeneracy of the i -th energy level \mathcal{E}_i is g_i , then only $n_i < g_i$ fermions can occupy it. The number of microstates for distinguishable n_i particles on this level would be therefore

$$w_i^{dist} = g_i (g_i - 1) (g_i - 2) \dots (g_i - (n_i - 1)) = \frac{g_i!}{(g_i - n_i)!}$$

Because the particles are on the same level \mathcal{E}_i indistinguishable this should be divided by the number of possible permutations of the n_i particles $n_i!$:

$$w_i = \frac{g_i!}{n_i! (g_i - n_i)!}$$

The total number of the different configurations is

$$w = \prod_i w_i = \prod_i \frac{g_i!}{n_i! (g_i - n_i)!} \quad (9.4.1)$$

This is proportional to the probability of a given configuration of the system as it was in the case of the Maxwell-Boltzmann statistics, therefore the observable equilibrium state can be determined by calculating the conditional maximum⁹ of w . The result is the *Fermi-Dirac distribution*:

$$\begin{aligned} n_i &= \frac{g_i}{e^{\alpha + \beta \mathcal{E}_i} + 1} & \text{with} & \\ \beta &= \frac{1}{k_B T} \\ \alpha &= -\frac{\mathcal{E}_F}{k_B T} \end{aligned} \quad (9.4.2)$$

⁸According to *particle physics elementary particles* of corpuscular matter are fermions, while bosons are quanta of the interactions (forces) that act between fermions. (This is of course not true for non-elementary particles, where bosons and fermions may be formed from a combinations of elementary fermions.) For instance photons are bosons with a spin of 1 and they are responsible for the electromagnetic interaction between fermions.

⁹In Appendix 22.13 we have shown the method of the solution for the Maxwell-Boltzmann distribution. It follows similar steps for fermions.

$$n_i = \frac{g_i}{e^{\frac{\mathcal{E}_i - \mathcal{E}_F}{k_B T}} + 1} \quad (9.4.3)$$

The corresponding $\Theta_F = \frac{\mathcal{E}_F}{k_B}$ temperature is the *Fermi-temperature*.

As with the Maxwell-Boltzmann statistics for a system with very many close energy levels we can use a continuous function, the *Fermi-Dirac distribution function* instead:

Important 9.4.1. *The Fermi-Dirac distribution function gives the probability that the states in a $\Delta \mathcal{E}$ interval around \mathcal{E} are occupied.*

$$f_{FD}(\mathcal{E}) := \frac{1}{e^{\frac{\mathcal{E} - \mathcal{E}_F}{k_B T}} + 1} \quad (9.4.4)$$

The number of fermions then

$$N = \int_0^\infty g(E) f_{FD}(E) d\mathcal{E} \quad (9.4.5)$$

and the average value of any energy dependent physical quantity $\mathcal{F}(E)$ for one particle is:

$$\langle \mathcal{F} \rangle = \frac{\int_0^\infty \mathcal{F}(\mathcal{E}) \cdot g(E) \cdot f_{FD}(\mathcal{E}) d\mathcal{E}}{\int_0^\infty g(E) \cdot f_{FD}(\mathcal{E}) d\mathcal{E}} \quad (9.4.6)$$

Fig. 9.2 shows the Fermi-Dirac distribution function at different temperatures. At $T = 0 \text{ K}$ the function is a step function. We explained this in the previous section by a physical argument. Now we determine this mathematically by taking the limit of (9.4.3) at $T = 0 \text{ K}$:

$$\lim_{T \rightarrow 0} e^{(\mathcal{E}_i - \mathcal{E}_F)/k_B T} = \begin{cases} 0 & \mathcal{E}_i < \mathcal{E}_F \\ \infty & \mathcal{E}_i > \mathcal{E}_F \end{cases} \quad (9.4.7)$$

Consequently at $T = 0 \text{ K}$

$$\left. \frac{n_i}{g_i} \right|_{T=0 \text{ K}} = \begin{cases} 1 & \mathcal{E}_i < \mathcal{E}_F \\ 0 & \mathcal{E}_i > \mathcal{E}_F \end{cases} \quad (9.4.8)$$

At non zero temperatures if $\mathcal{E}/k_B T \gg 1$, then both \mathcal{E}_F in the exponent and the 1 in the denominator can be neglected and the distribution function becomes

$$f_{FD}(\mathcal{E}) \approx e^{-\mathcal{E}/k_B T}$$

which is approximately the same as the Maxwell-Boltzmann distribution function.

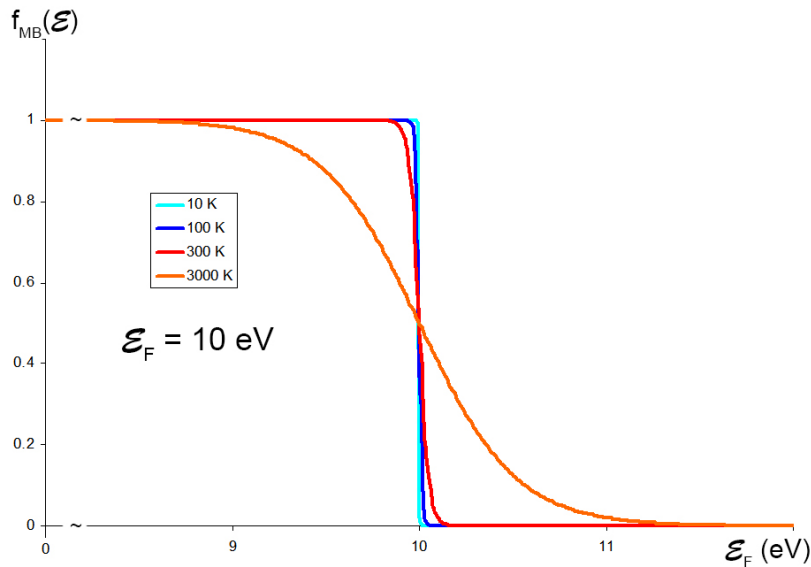
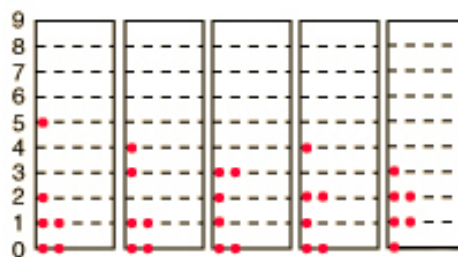


Figure 9.2: The Fermi-Dirac distribution function at different temperatures. Notice even the room temperature curve is almost a step function.

Example 9.3. In a system with equidistant energy levels how many ways can you distribute 9 units of energy among 6 fermions? The energy of the ground state ($i=0$) is 0, and the levels are one unit of energy distant from each other. Calculate and graph the distribution and compare it both with the Fermi-Dirac distribution function and with the Maxwell-Boltzmann distribution and distribution function! **Solution** Like in Problem 9.1 the observable different *macrostates* give the number of particles on every level, while the *microstates* are the possible ways to achieve a given macrostate.

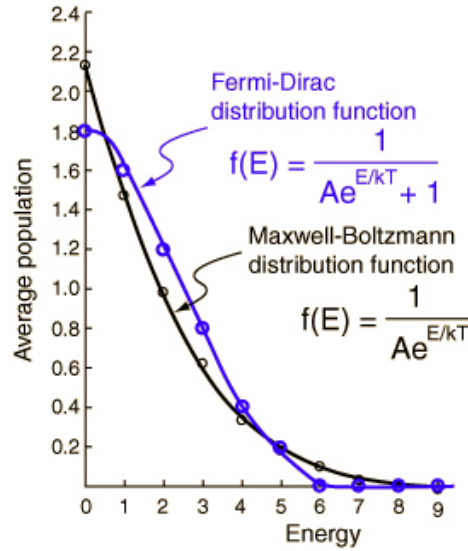
Because we must distribute 9 units of energy among the particles and the energy of the ground state is 0, we have to use 10 energy levels.

Because fermions are indistinguishable, obey the Pauli exclusion principle and have a half-integer spin there may be maximum 2 particles of opposite spins in each state:



Whereas there were 26 possible configurations for distinguishable particles (see Problem 9.1), these are reduced to the 5 states which have no more than two particles in each state. The average occupation numbers or *average population* of the levels are easier to calculate in this case. In the table we compared these numbers with the ones we got for the Maxwell-Boltzmann distribution.

Energy level	$\langle n_i^{FD} \rangle$	$\langle n_i^{MB} \rangle$
0	1.8	2.143
1	1.6	1.484
2	1.2	0.989
3	0.8	0.629
4	0.4	0.378
5	0.2	0.210
6	0.0	0.105
7	0.0	0.045
8	0.0	0.015
9	0.0	0.003



In the figure we used A for the factor e^α we got from our conditional maximum calculation. For the Maxwell-Boltzmann distribution $A \equiv Z$, for the Fermi-Dirac distribution $A \equiv e^{-\mathcal{E}_F/k_B T}$.

Low energy states are less probable with Fermi-Dirac statistics than with the Maxwell-Boltzmann statistics while mid-range energies are more probable. This difference is dramatic for large number of particles and for low temperatures as you will see later.

We can calculate the internal energy of a system of fermions using (9.4.6) and (9.4.5):

$$U = N \langle \mathcal{E} \rangle = \int_0^\infty \mathcal{E} g(\mathcal{E}) f_{FD}(\mathcal{E}) d\mathcal{E} = \int_0^\infty \frac{g(\mathcal{E}) \mathcal{E}}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} d\mathcal{E} \quad (9.4.9)$$

At 0 K this integral is simple as f_{FD} is 1 below \mathcal{E}_F and 0 above it. The density of states from (3.5.20)

$$g(\mathcal{E}) = 2 \cdot \frac{4\pi V \sqrt{2m^3}}{h^3} \sqrt{\mathcal{E}}, \quad (9.4.10)$$

where we used that fermions are half-integer spin particles, so any state can be occupied by 2 electrons; so $g(\mathcal{E})$ of (3.5.20) must be multiplied by 2. Therefore

$$U = \frac{8\pi V \sqrt{2m^3}}{h^3} \int_0^{\mathcal{E}_F} \mathcal{E}^{3/2} d\mathcal{E} = \frac{16\pi V \sqrt{2m^3}}{5h^3} \mathcal{E}_F^{5/2} \quad (9.4.11)$$

Compare this with the (9.2.18) internal energy of an ideal gas, which is 0 at $T = 0\text{ K}$.

Example 9.4. *As an example we will use the Fermi-Dirac distribution for the problem of the conduction electrons in metals, the so called electron gas model. In this model we assume that electrons can move freely inside a metal and behave like an ideal gas, i.e. all interaction between the electrons occurs only in collisions, and the Coulomb repulsion is considered zero¹⁰.*

The number of electrons between \mathcal{E} and $\mathcal{E} + \Delta\mathcal{E}$ is

$$dn(\mathcal{E}, \Delta\mathcal{E}) = g(\mathcal{E}) f_{FD}(\mathcal{E}) \Delta\mathcal{E} = \frac{8\pi V \sqrt{2m^3}}{h^3} \frac{\sqrt{\mathcal{E}}}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T + 1}}$$

The total number of electrons $N = \int dn = \int \frac{dn}{d\mathcal{E}} d\mathcal{E}$. Which is easy to compute at $T = 0\text{ K}$:

$$N = \frac{8\pi V \sqrt{2m^3}}{h^3} \int_0^{\mathcal{E}_F} \sqrt{\mathcal{E}} d\mathcal{E} = \frac{8\pi V \sqrt{2m^3}}{h^3} \mathcal{E}_F^{3/2}$$

from which the \mathcal{E}_F Fermi-energy at $T = 0\text{ K}$

$$E_F = \frac{h^2}{8m} \left(\frac{3N}{\pi V} \right)^{2/3}, \quad (9.4.12)$$

is a function of the density of electrons $\frac{N}{V}$. Combining this with (9.4.11) yields:

$$U = \frac{3}{5} N \mathcal{E}_F \quad (9.4.13)$$

The shape of the $dn/d\mathcal{E} = g(\mathcal{E}) f_{FD}(\mathcal{E})$ curve can be seen on Fig. 9.3 at different temperatures.

¹⁰Although this seems an invalid assumption we will see in Chapter 15 why this model is good for conduction electrons.

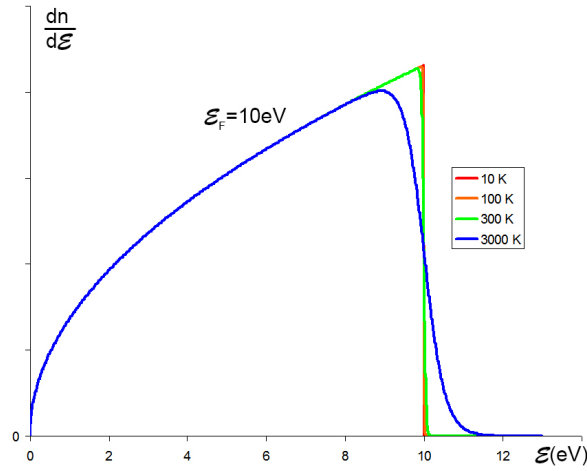


Figure 9.3: $\frac{dn}{d\varepsilon}$ curve for a free electron gas.

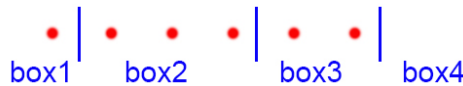
9.5 Bose-Einstein distribution.

Bosons may be elementary particles like photons¹¹ or composite particles, e.g ${}^4\text{He}$ atoms. For bosons the exclusion principle does not hold. Common feature is that the spin of bosons is integer.

At 0 K all of them are in the same (lowest lying) quantum state. Interestingly, at low but not zero temperatures still an unlimited number of bosons will "condense" into the lowest energy state, therefore quantum effects may become apparent on a macroscopic scale. This gives rise to the special state of matter, the so called *Bose Einstein Condensate* (see Appendix 22.14 for an example) .

Bosons have their own statistics called *Bose-Einstein statistics*

Let the degeneracy of state \mathcal{E} be g_i and let there be n_i bosons in this state. Then the number of microstates on this level will be equal to all of the possible ways n_i identical particles can be divided into g_i identical "boxes". We can visualize it by representing particles with a dot and "boxes" with two vertical lines. The next figure shows a partition of an energy level with 6 particles and 4 "boxes", the 4th of which is empty:



The number of lines needed to represent the boxes is 1 less than the number of boxes as seen on the figure. Therefore the number of possible partitions for this level is all

¹¹ Or the famous *Higgs boson* the Large Hadronic Collider (LHC) were constructed to detect.

possible permutations of $n_i + g_i - 1$ objects, from which n_i and $g_i - 1$ are identical:

$$w_i = \frac{(n_i + g_i - 1)!}{n_i! (g_i - 1)!} = \binom{n_i + g_i - 1}{n_i} \quad (9.5.1)$$

This is called the *number of combinations with repetition* in mathematics. The total number of all different configurations therefore is

$$w = \prod_i w_i = \prod_i \frac{(n_i + g_i - 1)!}{n_i! (g_i - 1)!} \quad (9.5.2)$$

With the method of analogous to that in Appendix 22.13 we find that in the maximum probability (i.e. in the equilibrium) state:

$$n_i = \frac{g_i}{e^{\alpha + \beta \varepsilon_i} - 1} \quad (9.5.3)$$

$$\beta = \frac{1}{k_B T} \quad (9.5.4)$$

and α is again determined from the condition

$$N = \sum_i n_i \quad (9.5.5)$$

In contrast with the Fermi-Dirac distribution we must assume that $\alpha > 0$, because $n_i \geq 0$ and there is a (-1) in the denominator. In the limit of very dense energy levels, i.e. continuous energies we can use the Bose-Einstein probability distribution function:

Important 9.5.1. *The Bose-Einstein distribution function.*

$$f_{BE}(E) := \frac{1}{A e^{\varepsilon/k_B T} - 1}, \quad \text{where} \quad (9.5.6)$$

$$A = e^\alpha \quad (9.5.7)$$

gives the probability that the states in a $\Delta \varepsilon$ interval around ε are occupied. The total number of bosons then

$$N = \int_0^\infty g(E) f_{BE}(E) d\varepsilon \quad (9.5.8)$$

and the average value of any energy dependent physical quantity $\mathcal{F}(E)$ for one particle is:

$$\langle \mathcal{F} \rangle = \frac{\int_0^\infty \mathcal{F}(\varepsilon) \cdot g(E) \cdot f_{BE}(\varepsilon) d\varepsilon}{\int_0^\infty g(E) \cdot f_{BE}(\varepsilon) d\varepsilon} \quad (9.5.9)$$

It may help to memorize all three distribution functions if we write them in similar forms¹²:

$$f_{MB}(E) = \frac{1}{e^{\alpha+\beta E}}, \quad \alpha = \frac{1}{Z} \quad (9.5.10a)$$

$$f_{FD}(E) = \frac{1}{e^{\alpha+\beta E} + 1}, \quad \alpha = -\frac{\mathcal{E}_F}{k_B T} \quad (9.5.10b)$$

$$f_{BE}(E) = \frac{1}{e^{\alpha+\beta E} - 1}, \quad \alpha = \ln A \quad (9.5.10c)$$

These are compared in Fig. 9.4.

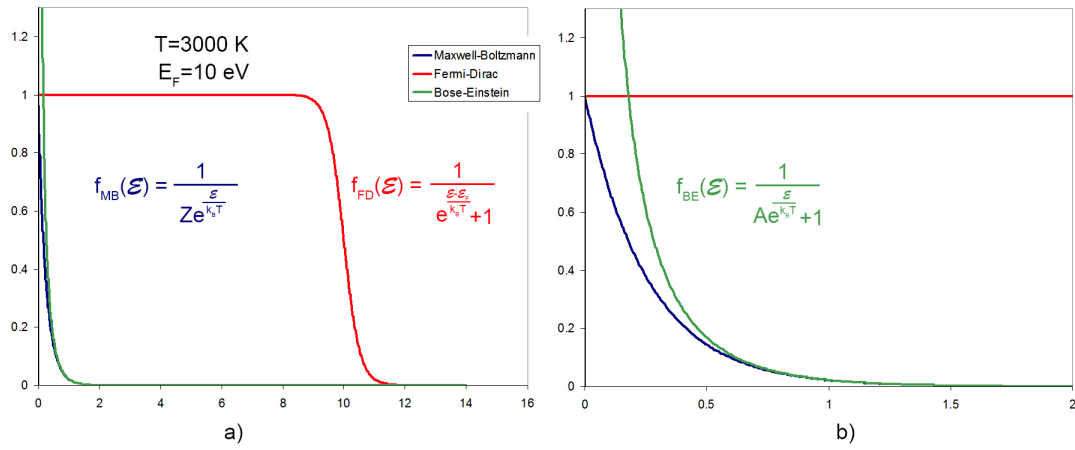


Figure 9.4: Comparison of the three distribution functions for a system with very large number of particles.

Example 9.5. *In a system with equidistant energy levels how many ways can you distribute 9 units of energy among 6 bosons? The energy of the ground state ($i=0$) is 0, and the levels are one unit of energy distant from each other. Calculate and graph the distribution and compare it both with the Maxwell-Boltzmann and Fermi-Dirac distribution! Solution Like in Problem 9.1 the observable different *macrostates* give the number of particles on every level, while the *microstates* are the possible ways to achieve a given macrostate.*

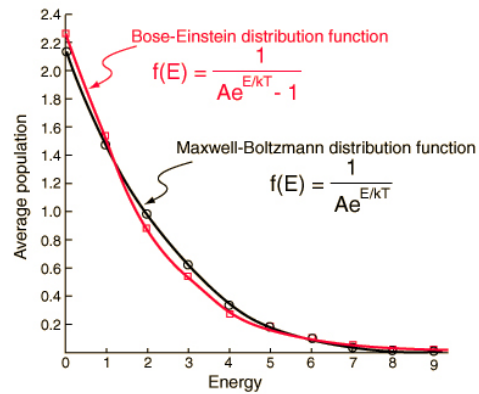
We must distribute 9 units of energy among the particles and the energy of the ground state is 0, we have to use 10 energy levels.

Because any number of bosons can be in the same state, similar to the classical distinguishable particles of the Maxwell-Boltzmann distribution, the

¹²For photons $\alpha = 0$, due to the fact that even in thermal equilibrium the number of photons is not constant, but fluctuate.

total number of macrostates is again 26. (See the corresponding figure at Problem 9.1.) But in this case it is the same as the number of microstates, because bosons are indistinguishable, and the exchange of two particles does not lead to a different microstate. In the next table we compared the average occupation numbers or *average population* of the levels with the ones we got for the other two distributions, but only graph the Bose-Einstein and Maxwell-Boltzmann curves.

Energy level	$\langle n_i^{BE} \rangle$	$\langle n_i^{FD} \rangle$	$\langle n_i^{MB} \rangle$
0	2.269	1.8	2.143
1	1.538	1.6	1.484
2	0.885	1.2	0.989
3	0.538	0.8	0.629
4	0.269	0.4	0.378
5	0.192	0.2	0.210
6	0.115	0.0	0.105
7	0.077	0.0	0.045
8	0.038	0.0	0.015
9	0.038	0.0	0.003



Chapter 10

Interaction of light and matter.

10.1 Photon gas

One of the basic phenomena that led to the development of quantum mechanics was the black-body radiation. (See Chapter 2). The electromagnetic waves inside a cavity are in a dynamic equilibrium with the walls of the cavity: the rate of absorption and emission are equal.

The photons in the cavity are elementary particles with $h\nu$ energy, $h\nu/c = h/\lambda$ momentum and spin 1. These do not interact with each other, and any number of them can be present with the same energy at any time. Therefore photons are bosons and we may say that the cavity is filled with a “gas” of *photons*.

However the number of the photons in the cavity is not constant, photons are continuously absorbed and emitted by the walls. So the value of α in (9.5.10c) must be set to 0 (in this case $A = 1$). If the cavity is large, the spectrum of the possible \mathcal{E}_i energies (i.e. the possible frequencies) can be considered continuous. So for photons in a thermal equilibrium we can use the continuous Bose-Einstein distribution function.

$$f_{BE}(E) := \frac{1}{e^{\mathcal{E}/k_B T} - 1} \quad \text{and} \quad (10.1.1)$$

$$dn(\mathcal{E}, d\mathcal{E}) = \frac{g(E)}{e^{\mathcal{E}/k_B T} - 1} d\mathcal{E} \quad (10.1.2)$$

We can easily get the density of states per unit frequency for photons from (3.5.19) using the $k = 2\pi/\lambda = 2\pi\nu/c$ relation¹:

$$g(\nu) = g(k(\nu)) \frac{dk}{d\nu} = \frac{V}{2\pi^2} \frac{(2\pi)^3}{c^3} \nu^2 = \frac{4\pi V}{c^3} \nu^2 \quad (10.1.3)$$

¹As we promised near Equation 3.5.19

We must find the density of states for photons. We start from (3.5.20),

$$g(\mathcal{E}) = \frac{4\pi V \sqrt{2m_e^3}}{h^3} \sqrt{\mathcal{E}},$$

we calculated based on the wave-like nature of the electrons in a potential box. We look for the density of states by frequency of the photons. For the photon $\mathcal{E} = h\nu = cp \rightarrow p = \frac{\mathcal{E}}{c} = \frac{h\nu}{c} \mathcal{E}_e = p_e^2/2m_e$. With $\mathcal{E}_{ph} = h\nu = cp_{ph}$ using $p_{ph} = h\nu/c$:

Background 10.1.1. • We replace the density of states for the energy $g(\mathcal{E}_e)$ with the density of states for the momentum $g(p_e)$. Using $\mathcal{E}_e = \frac{p_e^2}{2m_e}$ introduce :

$$\begin{aligned} dN(\mathcal{E}, d\mathcal{E}) &= g(\mathcal{E}_e) d\mathcal{E} = g(p_e) dp_e \\ g(p_e) &= g(\mathcal{E}_e(p_e)) \frac{d\mathcal{E}_e}{dp_e} = g(\mathcal{E}_e(p_e)) \frac{p_e}{m_e} = \\ &= \frac{4\pi V}{h^3} m_e \sqrt{2m_e} \sqrt{\mathcal{E}_e} \frac{p_e}{m_e} = \frac{4\pi V}{h^3} p \sqrt{2m_e} \sqrt{\mathcal{E}_e} \end{aligned}$$

• and substitute p_e into E_e :

$$\begin{aligned} g(p_e) &= \frac{4\pi V}{h^3} p_e \sqrt{2m_e} \sqrt{\frac{p_e^2}{2m_e}} = \\ &= \frac{4\pi V}{h^3} p_e^2 \end{aligned}$$

• then replace p_e with $p \equiv p_{ph}$ and introduce the frequency dependent density of state for the photon using the relation $p = \frac{h}{\lambda} = \frac{h\nu}{c}$ by

$$\begin{aligned} g(\nu) d\nu &= g(p) dp \\ g(\nu) &= g(p(\nu)) \frac{dp}{d\nu} = g(p(\nu)) \frac{h}{c} \end{aligned}$$

• finally substitute ν into p

$$\begin{aligned} g(\nu) &= \frac{4\pi V}{h^3} \frac{h^2 \nu^2}{c^2} \frac{h}{c} = \\ &= \frac{4\pi V}{c^3} \nu^2 \end{aligned}$$

Electromagnetic waves are transverse waves with two independent polarizations, which means that for every ν frequency there are 2 photon states available. The photon density of states therefore is

$$g(\nu) = \frac{8\pi V}{c^3} \nu^2 \quad (10.1.4)$$

With (10.1.4) (10.1.2) becomes

$$dn(\nu, d\nu) = \frac{8\pi V}{c^3} \nu^2 \frac{1}{e^{h\nu/k_B T} - 1} d\nu \quad (10.1.5)$$

The average photon energy at ν in a cavity with volume V is $h\nu \cdot dn(\nu, d\nu)$ and the average *energy density* $u(\nu)$ at frequency ν is

$$u(\nu) = \frac{h\nu}{V} \frac{dn(\mathcal{E}, d\mathcal{E})}{d\nu} = \frac{8\pi h}{c^3} \frac{\nu^3}{e^{h\nu/k_B T} - 1} \quad (10.1.6)$$

The connection between the energy density and the *spectral radiance*² $\varepsilon(\nu, T)$ is

$$u = \frac{4\pi}{c} \varepsilon$$

which leads to the famous formula of the black-body radiation:

$$\varepsilon(\nu, T) = \frac{c}{4\pi} u = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(k_B T)} - 1} \quad (10.1.7)$$

10.2 Interaction of light and matter

In the previous section we dealt with the electromagnetic radiation only and arrived to the Planck formula by using the Bose-Einstein distribution function. But to really explain the black-body radiation we have to examine how the equilibrium between the radiation and the atoms in the wall of the black-body cavity interact. The model we discuss was first developed by Einstein.

Suppose for simplicity that the walls of the cavity consist of the atoms that have only two discrete energy levels \mathcal{E}_1 and $\mathcal{E}_2 > \mathcal{E}_1$. The frequency of the emitted or absorbed photon is $\nu = \Delta\mathcal{E}/h = (\mathcal{E}_2 - \mathcal{E}_1)/h$. Let $u(\nu)$ denote the electromagnetic energy density, and $W(A \rightarrow B)$ the probability of the transition per unit time (called *probability coefficients*) and per unit energy density from level A to level B ! The possible processes, and their probabilities then are (c.f. Section 4.3.1)

²C.f. Section 2.1

Process	Transition probability per unit time	
	probability coefficients	total
absorption	$B_{12} \equiv W(1 \rightarrow 2)$	$B_{12} u(\nu)$
induced emission	$B_{21} \equiv W(2 \rightarrow 1)$	$B_{21} u(\nu)$
spontaneous emission	A_{12}	A_{12} - independent of $u(\nu)$

The number of transitions per unit time $\mathcal{N}_{1 \rightarrow 2}$ and $\mathcal{N}_{2 \rightarrow 1}$ will be proportional to the number of atoms in the corresponding states (N_1 and N_2):

$$\begin{aligned}\mathcal{N}_{1 \rightarrow 2} &= N_2 [A_{21} + B_{21} u(\nu)] \quad \text{emission} \\ \mathcal{N}_{2 \rightarrow 1} &= N_1 B_{12} u(\nu) \quad \text{absorption}\end{aligned}$$

The rate of change of the number of excited atoms then

$$\frac{dN_2}{dt} = \underbrace{N_1 B_{12} u(\nu)}_{\text{absorption}} - \underbrace{N_2 [A_{21} + B_{21} u(\nu)]}_{\text{emission}}$$

In equilibrium the number of excited atoms is constant: $\left(\frac{dN_2}{dt} = 0\right)$, therefore

$$\frac{N_2}{N_1} = \frac{B_{12} u(\nu)}{[A_{21} + B_{21} u(\nu)]} \quad (10.2.1)$$

On the other hand atoms are (in principle) distinguishable particles therefore the Maxwell-Boltzmann statistics can be used in thermal equilibrium, i.e.

$$\frac{N_2}{N_1} = e^{-\frac{(\varepsilon_2 - \varepsilon_1)}{k_B T}} = e^{-h\nu/k_B T} \quad (10.2.2)$$

From (10.2.1) and (10.2.2)

$$B_{12} u(\nu) e^{h\nu/k_B T} = [A_{21} + B_{21} u(\nu)], \quad (10.2.3)$$

from which we can express the electromagnetic energy density:

$$u(\nu) = \frac{A_{21}}{B_{12} e^{h\nu/k_B T} - B_{21}} = \frac{\frac{A_{21}}{B_{12}}}{e^{h\nu/k_B T} - \frac{B_{21}}{B_{12}}} \quad (10.2.4)$$

The common name for the A_{12} , B_{12} and B_{21} probability coefficients is *Einstein coefficients*. Comparing this with (10.1.6) we find³ that

$$\begin{aligned}B_{21} &= B_{12} \quad \text{and} \\ A_{21} &= \frac{8\pi\nu}{c^3} B_{21}\end{aligned}$$

³The Planck formula is valid for all frequencies, while these are only valid for the given two energy levels. If an atom has multiple levels, then similar formulas are valid for any two of them, but still the possible frequency spectrum would not be continuous. But, as we will see in solid state physics, the possible energy levels in a solid made of a large number of atoms are contained in quasi-continuous *energy bands*, therefore the frequency spectrum will also be quasi-continuous.

The probability of the absorption is equal to the probability of the induced emission, and the ratio of the probabilities of the spontaneous and induced emissions per unit time is

$$\frac{A_{21}}{B_{21}} = e^{\frac{h\nu}{k_B T}} - 1$$

As a consequence

Important 10.2.1. • For light and higher frequency electromagnetic radiation $\frac{h\nu}{k_B T} \gg 1$ the induced emission is negligible in thermal equilibrium.

- In the region of microwaves and below $\frac{h\nu}{k_B T} \ll 1$ and the induced emission is dominant.

Photons emitted in the induced emission process will have the same frequency and phase as the absorbed photons, therefore the induced emission creates coherent electromagnetic radiation. While photons emitted in the spontaneous emission process are incoherent.

10.3 Laser operation.

10.3.1 Optical amplification

The fact that electromagnetic radiation (including visible light) from induced (stimulated) emission is coherent with the absorbed radiation it can be used to create coherent radiation sources. Depending on the frequency range such a device is called either a *laser* (*Light Amplification by Stimulated Emission of Radiation*) or a *maser* (*Microwave Amplification by Stimulated Emission of Radiation*⁴).

As we saw for frequencies in the visible range and above the induced emission is usually negligible. The ratio of the number of transitions per unit time with reordering (10.2.1) and using that $B_{12} = B_{21}$:

$$\frac{\mathcal{N}_{2 \rightarrow 1}}{\mathcal{N}_{1 \rightarrow 2}} = \frac{N_2}{N_1} \frac{[A_{21} + B_{21} u(\nu)]}{B_{12} u(\nu)} = \frac{N_2}{N_1} \left(1 + \frac{A_{21}}{B_{21} u(\nu)} \right) \quad (10.3.1)$$

In thermal equilibrium $N_2 < N_1$, but in an open medium (the *gain medium*) that is not thermal equilibrium using special non-thermal external excitation it is possible to

⁴In modern usage "light" broadly denotes electromagnetic radiation of any frequency, not only visible light, hence we can talk about infrared laser, ultraviolet laser, X-ray laser, and so on. Because the microwave predecessor of the laser, the maser, was developed first, devices of this sort operating at microwave and radio frequencies are referred to as "masers" rather than "microwave lasers" or "radio lasers".

artificially reverse this so that the population of the higher level will be the larger one, i.e. $N_2 > N_1$. Such excitation is called “*pumping*”. This is called *population inversion*. A small perturbation of such a metastable system can start the de-excitation process (*spontaneous emission*) whose end result will be the equilibrium state. This perturbation can be caused by a light of suitable frequency. Because of the population inversion the probability of induced emission (proportional to $\mathcal{N}_{2 \rightarrow 1}$) is much larger than the probability of absorption and as we saw the emission will be coherent with the perturbation. This process is an *optical amplification*: a small intensity light enters the gain medium and a higher intensity coherent light leaves it. The gain medium therefore is itself an *optical amplifier*.

10.3.2 Laser operation

A laser which produces light by itself is technically an *optical oscillator* rather than an optical amplifier as suggested by the acronym. When an *optical amplifier* is placed inside an *optical resonator*⁵, one obtains a *laser oscillator*. This optical resonator (see Fig. 10.1) usually consist of two parallel mirrors and the coherent light beam of the optical amplifier travels to and fro between these in both direction. This way a if the amplification (gain) is larger than the resonator losses (caused by absorption and diffraction) the power of the light bouncing between the two mirrors increases exponentially. As more and more atoms will go back to the lower energy state the gain will decrease, until the losses overcome the gain. The level of gain equal to the losses is the laser *threshold*.

Simple 2 level systems cannot be used in lasers, because the pumping source not only provides energy for excitations, but at the same time it creates the induced emission itself, effectively negating the pumping effect. At least 3 levels are required as seen in Fig. 10.2 a. The pumping creates a popular inversion on level \mathcal{E}_3 with short occupation life time, from which the electrons almost immediately decay onto the metastable level \mathcal{E}_2 . In a 3 level system a perturbation of frequency $(\mathcal{E}_2 - E_1)/h$ starts the laser process. 3 level lasers work only in *pulsed operation* mode, because pumping is a linear process, therefore it cannot compensate for the exponential process of the laser emission⁶.

In a 4 level system (Fig. 10.2 b) the laser process occurs between the metastable level \mathcal{E}_3 and the short lifetime level \mathcal{E}_2 . This is the model of a continuous operation mode laser, because in this case the ground level may have a linear filling up rate from electrons that participated in the laser process therefore the linear pumping process can maintain a steady (dynamic equilibrium) state.

⁵The optical resonator is sometimes referred to as an “optical cavity”, but this is a misnomer: lasers use open resonators as opposed to the literal cavity that would be employed at microwave frequencies in a maser.

⁶Naturally any laser can be used in pulse operation mode by either switching it on and off or by using pulsed pumping, but 3 level lasers cannot work in continuous operation mode.

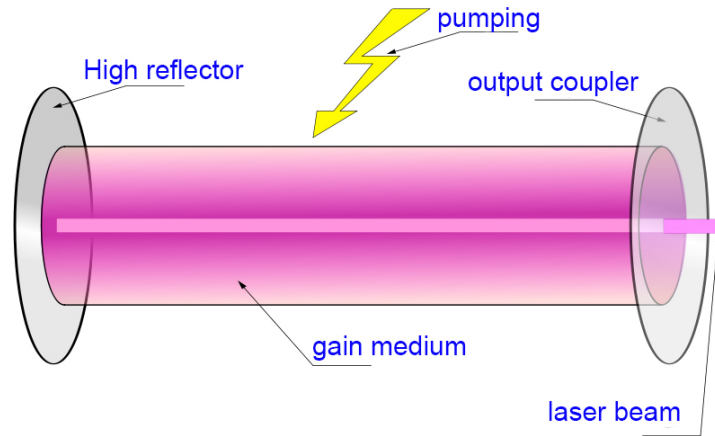


Figure 10.1: Components of a typical laser. The "high mirror" is a perfect mirror, while the "output coupler" (OC) is only partially reflecting, the reflectivity required depends on the gain medium. E.g. for He-Ne lasers the reflectivity must be at least 99%, while nitrogen lasers have extremely high gain and do not require any OC at all.

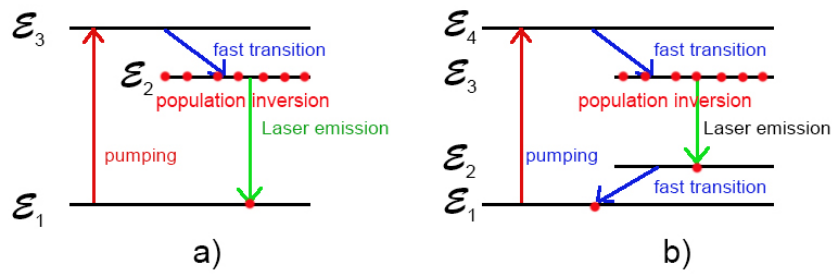


Figure 10.2: Simplest laser level structures. a) 3 level laser, b) 4 level laser

10.3.3 Types of lasers

There are very different types of lasers, possessing properties like parallelism, monochromatic radiation, high average or peak power, very short pulse length. In size they vary from microscopic semiconductor lasers to the building sizes system used for laser research.

Gas lasers

the gain medium is a gas (e.g. CO_2) or gas mixture (e.g. He-Ne). He-Ne lasers are able to operate at a number of different wavelengths, however the vast majority of gas lasers are engineered *to lase* at 633 nm; these relatively low cost but highly coherent lasers are extremely common in optical research and educational laboratories. Commercial carbon dioxide (CO_2) lasers can emit many hundreds of watts in a single spatial mode which can be concentrated into a tiny spot. This emission is in the thermal infrared at $10.6\ \mu m$; such lasers are regularly used in industry for cutting and welding. The efficiency of a CO_2 laser is unusually high: over 10%.

Chemical lasers

these lasers are powered by a chemical reaction permitting a large amount of energy to be released quickly. Such very high power lasers are especially of interest to the military, however continuous wave chemical lasers at very high power levels, fed by streams of gasses, have been developed and have some industrial applications.

Solid-state lasers

the gain medium is a crystalline or glass rod containing impurity ions (the accepted terminology is that they are *doped* with this impurities, usually called *dopants*) that provide the required energy states. In fact the first working laser was a ruby laser, made from ruby (chromium-doped corundum). The population inversion is actually maintained in the "dopant", such as chromium or neodymium. These materials are pumped optically using a shorter wavelength than the lasing wavelength, often from a flashtube or from another laser. These lasers are also commonly frequency doubled, tripled or quadrupled, in so-called "*diode pumped solid state*" or DPSS lasers. Under second, third, or fourth harmonic generation these produce 532 nm (green, visible), 355 nm and 266 nm (UV) beams. This is the technology behind the bright laser pointers particularly at green (532 nm) and other short visible wavelengths. Some doped crystal (e.g. Ti:GaS) have very broad amplification spectral range that allows generation of very short pulses up the range of femtoseconds ($10^{15} sec$).

Semiconductor lasers

are diodes which are electrically pumped. Recombination of electrons and holes created by the applied current introduces optical gain. Reflection from the ends of the crystal form an optical resonator, although the resonator can be external to the semiconductor in some designs. Commercial laser diodes emit at wavelengths

from 375 nm to 3500 nm. Low to medium power laser diodes are used in laser pointers, laser printers and CD/DVD players. Laser diodes are also frequently used to optically pump other lasers with high efficiency. The highest power industrial laser diodes, with power up to 10 kW (70dBm) are used in industry for cutting and welding. They serve also as light sources for fiber optic communication systems that are the technical foundation of the internet.

Fiber lasers

These are solid-state lasers or laser amplifiers where the light is guided due to the total internal reflection in a single mode optical fiber. Pump light can be used more efficiently by creating a fiber disk laser, or a stack of such lasers. Fiber lasers have a fundamental limit in that the intensity of the light in the fiber cannot be so high that optical nonlinearities induced by the local electric field strength can become dominant and prevent laser operation and /or lead to the material destruction of the fiber. This effect is called *photodarkening*.

Free electron lasers'

or FELs, generate coherent, high power radiation, that is widely tunable, currently ranging in wavelength from microwaves, through terahertz radiation and infrared, to the visible spectrum, to soft X-rays. They have the widest frequency range of any laser type. While FEL beams share the same optical traits as other lasers, such as coherent radiation, FEL operation is quite different. Unlike gas, liquid, or solid-state lasers, which rely on bound atomic or molecular states, FELs use a modulated relativistic electron beam as the lasing medium, hence the term free electron.

Solid State Physics

Chapter 11

Fundamentals

11.1 Categorization of Solids

Solids are composed of atoms or molecules at fixed relative positions. Bonding between atoms ensures that the form and volume of solids remain the same in a large temperature and pressure range for a long period of time. (There are exceptions: glass behaves like a fluid over a long period – hundreds of years – of time.)

Important 11.1.1. *The number of atoms in 1cm^3 of a solid is about 10^{24} .*

There are two kinds of atomic ordering in solids:

Short Range ordering:

First- or second-nearest neighbors of an atom are arranged in the same structure. At distances that are many atoms away, however, the positions of the atoms are uncorrelated.

Long Range Ordering:

Once the positions of an atom and its neighbors are known at one point, the place of each atom is known precisely throughout the material.

Solids that have short-range order but lack long-range order are called amorphous, while a solid is crystalline if it has long-range order.

Many solids are crystalline in nature, that is, the atoms are arranged in a regular three-dimensional periodic pattern. There is a wide variety of crystal structures formed by different elements and by different combinations of elements.

In the following we will concentrate on crystalline solids.

Another possible categorization is by electric conductivity:

Conductors (metals) and insulators

Materials are conductors if the atomic orbitals significantly overlap, otherwise they are insulators. As we will see later on the available electronic energies form *energy bands* in

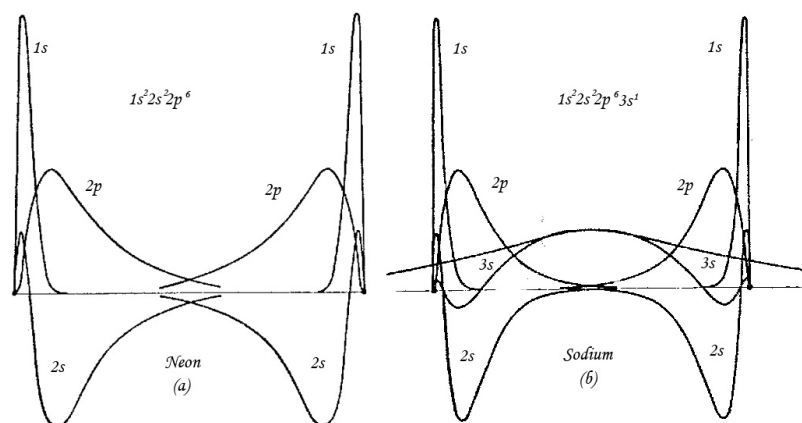


Figure 11.1: Atomic orbitals in an insulator (Ne) and a metal (Na)

every material. These are possibly separated by a *band gap* in which there are no energy levels available for electrons.

11.2 Bonding in crystals

Although the forces that hold a solid together are electrostatic in nature, the bonding can only be explained using quantum mechanics. In solids the bonding strength is about the same as in molecules.

There are several categories of solids depending on the bond type.

We will discuss the following bond types:

Bond type	Examples
covalent bond	H_2 , CO_2 , diamond
molecular bond	P_4 , Cl_2 , hydrocarbons ($H_n C_m$), fullerenes
ionic bond	$NaCl$, MgO , $CaCO_3$
hydrogen bond	ethanol ($C_2 H_6 O$), diethyl ether ($C_4 H_{10} O$), water
metallic bond	Al , Fe

11.2.1 Covalent Crystals

Some electrons are not sharply localized around the nuclei and their spatial distribution is not uniform, there are preferred directions with high electron density (these are called in chemistry as bonds)

Properties of covalent bonded crystals:

- they have a rigid electronic structure
- they are hard materials
- they are bad heat and electric conductors (there are no free electrons available)
- they have high frequency lattice vibrations (with excitation energies in the infrared (IR) range)
- they have a large electronic band gap therefore they are transparent for visible light

Example: diamond. the C atoms have 4 valence electrons per atom in sp^3 hybrid orbitals which form localized bonding electron pairs. The band gap is 5.5. eV

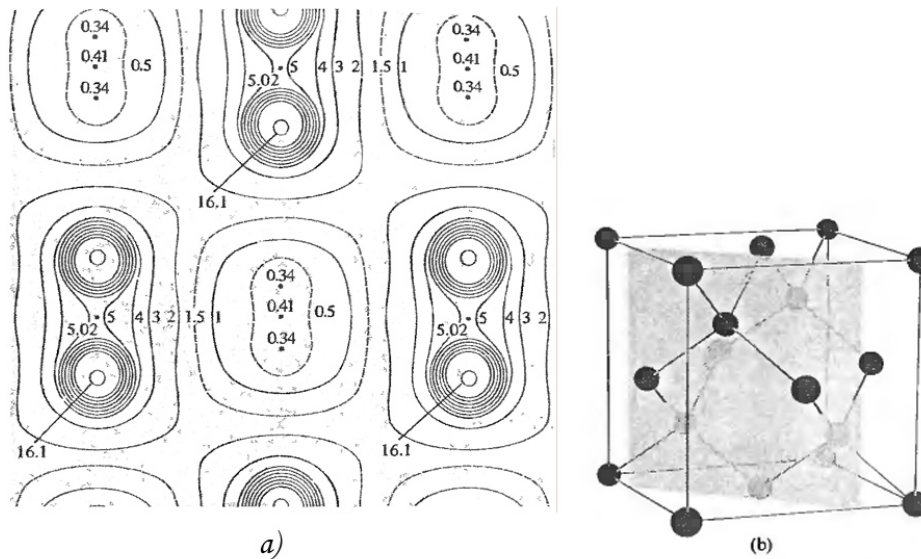


Figure 11.2: a) Electronic charge distribution by line density in diamond (numbers: electrons per cubic angstrom) on the plane displayed in b). Electron density is very high where the plane intersects a bond.

11.2.2 Ionic Crystals

The positively and negatively charged ions are held together by electrostatic forces. The ions themselves are almost impenetrable as a consequence of the Pauli principle. The attraction between the oppositely charged ions tries to pull them together. When the electronic charge distributions (i.e. the wave functions of the valence electrons) would start to overlap, which would violate the exclusion principle, the electron configuration must change to prohibit it. This can be described as an appearance of an excess charge on the next free energy level. Because the ions have stable closed shells this requires much energy.

Properties of ionic crystals:

- there are no free electrons therefore they are bad heat and electric conductors
- they are rigid and brittle
- the bonds are strong \Rightarrow high melting point
- the lattice vibration frequencies lies in IR (ionic bonds are weaker than covalent bonds)
- they have closed shells \Rightarrow they are diamagnetic (C.f. Section 19.1)

Purely ionic bonds cannot exist, as the proximity of the entities involved in the bond allows some degree of sharing electron density between them. Therefore, all ionic bonds have some covalent character. Thus, an ionic bond is considered a bond where the ionic character is greater than the covalent character. Ionic crystals may be categorized into sub-groups:

Example 11.1. Alkali Halides (I-VII ionic crystals¹)

+ ion - Li^+, Na^+, K^+, Rb^+ or Cs^+

- ion - F^-, Cl^-, Br^- or I^-

They usually crystallize in sodium chloride structure, except $CsCl$, $CsBr$ and CsI which are most stable in the cesium chloride structure.

¹When discussing periodic table groups, semiconductor physicists always use the older notation, instead of the current IUPAC group notation. For example, the carbon group is called "Group IV", not "Group 14"

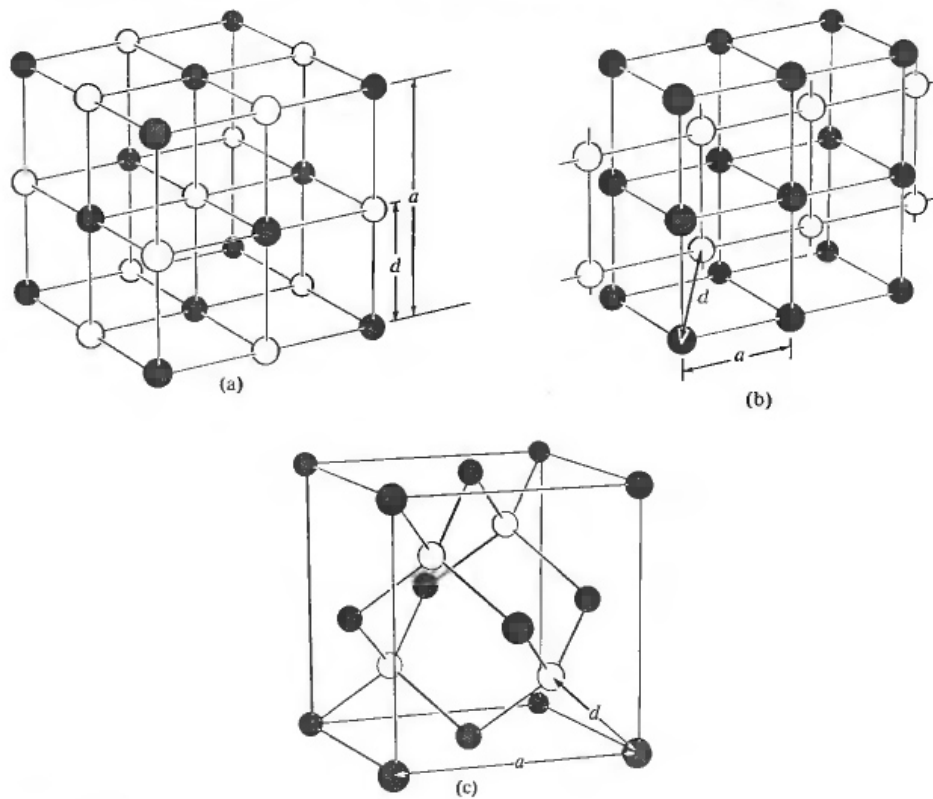


Figure 11.3: a) sodium chloride, b) cesium chloride, c) zincblende structure. a side of conventional cubic cell, d nearest neighbor distance (sodium chloride $d = a/2$, cesium chloride $d = \sqrt{3}a/2$, zincblende $d = \sqrt{3}a/4$)

II-VI ionic crystals

Double ionized elements from columns II and VI. + ion - Be, Mg, Ca, Sr, Ba

- ion - O, S, Se, Te

Usually sodium chloride structure, except BeS, BeSe and BeTe which are most stable in the zincblende structure.

III-V mixed ionic and covalent crystals

+ ion - Al, Ga, In

- ion - P, As, Sb

Usually zincblende structure.

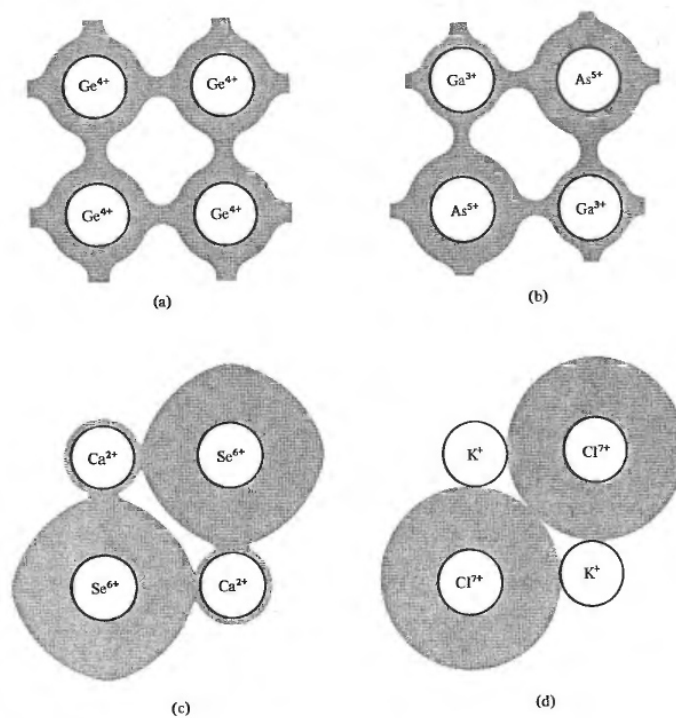


Figure 11.4: Schematic representation of the continuity from perfect covalent to perfect ionic crystals. a) perfectly covalent germanium, b) covalent gallium arsenide, c) ionic calcium selenide, d) perfectly ionic potassium chloride

11.2.3 Hydrogen bond crystals

Hydrogen is unique in 3 important ways:

- The H^+ ion (proton) is small (diam. 10^{-13} cm) about 10^{-5} times smaller than any other ion core. \Rightarrow it may sit on any other ion
- H is but 1 electron shy from stable helium configuration, which is the only one that has just 2 electrons on the outer shell \Rightarrow it cannot form covalent bonds
- The first ionization potential is high (-13.6 eV C.f. Li: 5.39 eV, Na: 5.14 eV, K: 4.34 eV) \Rightarrow does not behave as an alkali metal ion

Example: water ice

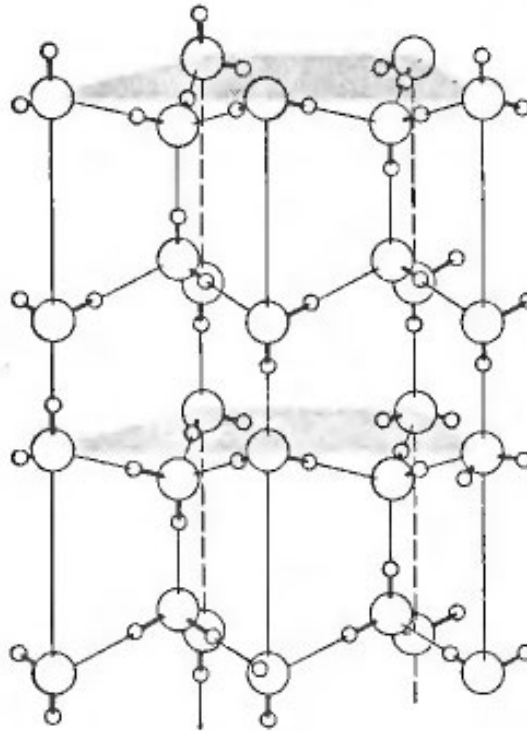


Figure 11.5: Crystal structure of a selected ice phase. The large circles are oxygen, the small one are the bonding protons.

11.2.4 Molecular crystals

This type of crystals form of atoms with no electric dipole moment \Rightarrow no electrons of uncompensated spin.

Properties:

- extreme weak bonding
- bad heat conductors
- bad electric conductors
- low melting point
- low boiling point
- easily compressible and deformable

Best examples are the elements in column VIII of the periodic table. Noble gases (except He) crystallize in this system. The solid is held together by very weak forces called *van der Waals forces* whose origin is explained qualitatively in Appendix 23.1.

Examples: elements in group V, VI and VII has both covalent and molecular character (exceptions: metallic polonium, semi-metals antimony and bismuth)

11.2.5 Metals

For metals (conductors) the covalent bonds between atoms expand to cover the whole crystal: electron density is appreciable throughout the interstitial regions² forming the so called electron gas. The atoms have relatively low ionization energies and lose their valence electrons because of the perturbation of other atoms as the crystal is formed.

Properties of metals:

- ionic cores are small and are surrounded by almost free electrons
- good heat and electric conductors
- excitation of electrons is easy in every frequency range \Rightarrow they are opaque and have high reflectivity

Visualization of the bond types

11.3 Crystal structures, unit cells and lattices.

Crystals are materials with long range ordering, i.e. there are periodic equivalence points in the material from which the surroundings of any atom look like the same.

The periodicity of crystals have important consequences concerning mechanical, electromagnetic and thermal properties of solids.

There are many ordinary solids we encounter in everyday life in which there exists a surprising degree of crystallinity. For example, a bar of soap, a chocolate bar, candles, sugar or salt grains, even bones in the human body, are an aggregate of crystallites of sizes between 0.5 and 50 μm .

In these examples, what determines the properties of the material is not so much the structure of individual crystallites but their relative orientation and the structure of boundaries between them. Even in this case, however, the nature of a boundary between two crystallites is ultimately dictated by the structure of the crystal grains on either side of it.

²Regions between the nuclei

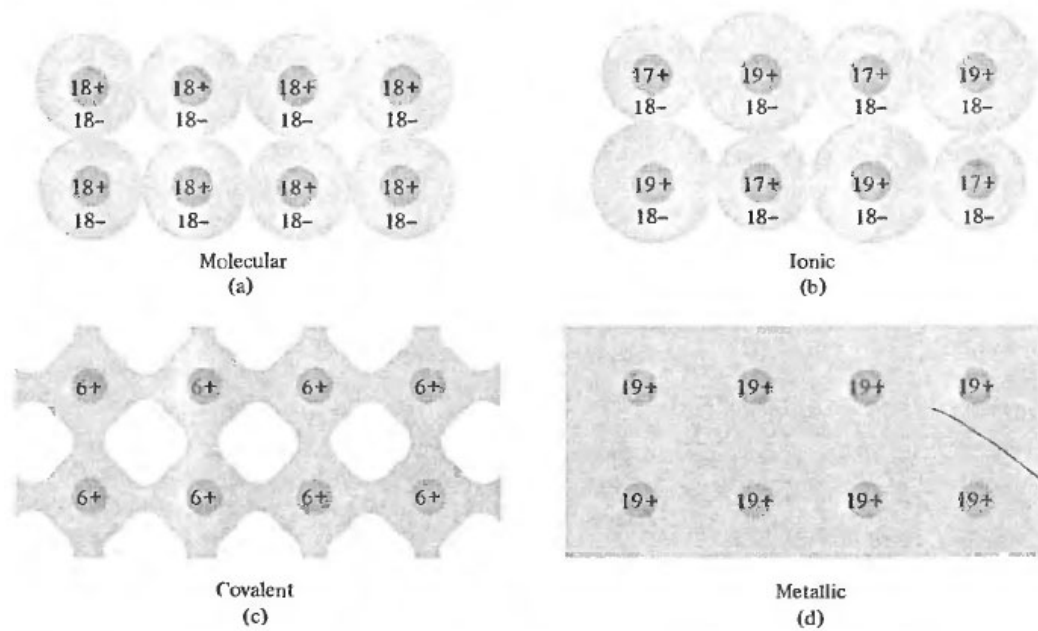


Figure 11.6: Highly schematic 2D comparison of different bond types: The small circles represent the positively charged nuclei, the shaded parts where electron density is appreciable (*but not uniform*) a) molecular bond - e.g. argon, b) ionic bond - e.g. potassium chloride, c) covalent bond - e.g. carbon, d) metallic bond- e.g. potassium

Important 11.3.1. *Because of the long range ordering every crystal must have translational symmetry.*

Let \mathbf{r}_1 and \mathbf{r}_2 two equivalent points separated by the vector \mathbf{R}

$$\mathbf{r}_2 = \mathbf{r}_1 + \mathbf{R}$$

then if we select a 3^{rd} point \mathbf{r}_3 so, that

$$\mathbf{r}_3 = \mathbf{r}_2 + \mathbf{R}$$

then point \mathbf{r}_3 will be equivalent with both \mathbf{r}_2 and \mathbf{r}_1 . This definition however requires mathematically the crystal to be perfect and infinite.

In real crystals this symmetry is broken at the surface and by any imperfections present in the solid. Because the interstitial (interatomic) distances for macroscopic crystals are much smaller than the size of the crystal relatively very few atoms are at or near the surface³. Inside the solid in first approximation we disregard these surface atoms and consider the crystal as an infinite one.

³Bulk atom concentration is about 10^{24} atoms/cm³, while at the surface the atom concentration is about 10^{15} atoms/cm²

A crystal may contain just a single kind of atom (monatomic crystal), but it may contain groups of atoms or even molecules from any number of any kind of atoms.

Important 11.3.2. *To simplify the description we may separate the physical structure of the crystal to a system of geometrical points, the so called point lattice and the basis containing atoms or molecules, which are located at every one of these geometrical points.*

Because of the translational symmetry there exist regions (volumes of space) of the crystal called *cells* containing one or more atoms from which the whole infinite crystal may be built using only the translational symmetry. These volumes may be represented by 1 point in space (\mathbf{r}_o) and by 3 non-coplanar vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 so that the origin \mathbf{r}_R of any other such cell may be represented by

$$\mathbf{r}_R = \mathbf{r} + \mathbf{R}, \text{ where} \quad (11.3.1)$$

$$\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3 \quad (11.3.2)$$

The set of points determined by (11.3.2) using all positive and negative numbers for n_1, n_2, n_3 forms the *point lattice*, and the vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 are the *primitive vectors*.

Important 11.3.3. *A point lattice that satisfies condition (11.3.2) is called a Bravais lattice (Auguste Bravais - 1845).*

The primitive vectors are said to *generate* or *span* the lattice.

Important 11.3.4. *The volume of space that when translated through all of the point lattice vectors just fills the complete space without overlap or without leaving voids is called a primitive cell or a primitive unit cell.*

These are not unique as can be seen in Fig. 11.7.

The selection of the primitive vectors is also not unique (Fig. 11.8).

Important 11.3.5. *The coordination number is the number of nearest neighbors (the number of lattice points which are closest to a selected point) in a lattice. The lattice constant (or lattice parameter is the distance between the origin of the neighboring unit cells.*

Putting the basis into each lattice points we arrive to the complete crystal structure⁴.

⁴As seen in Fig. 11.9 the whole plane (space) may be filled with any shape.

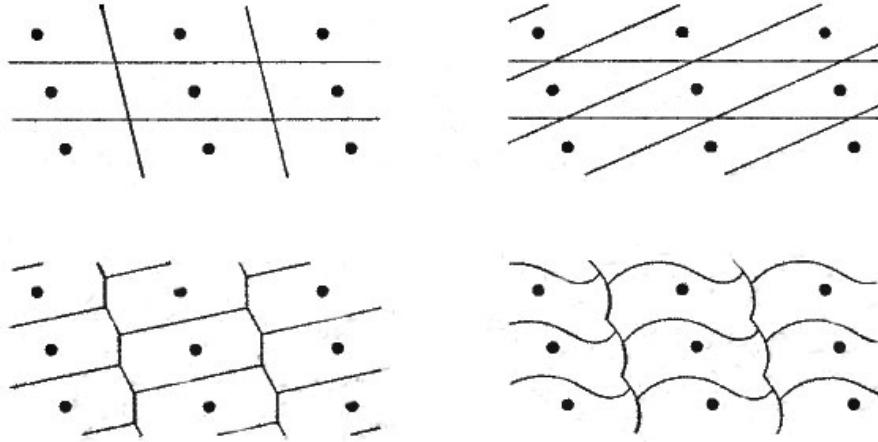


Figure 11.7: Various selections of primitive cells for a 2D Bravais lattice

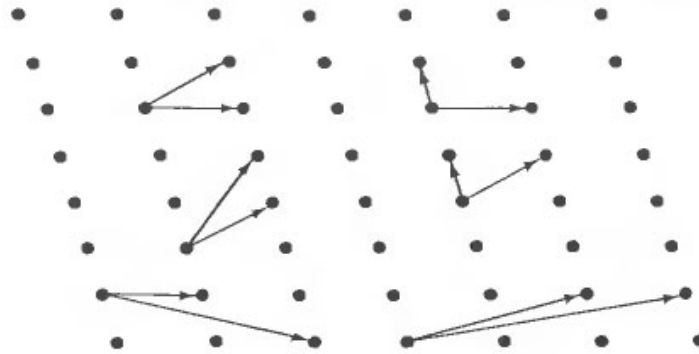


Figure 11.8: Various selections of primitive vectors for a 2D Bravais lattice

Example 11.2. *An infinite number of points are aligned along a line so that the distance between the n -th and $(n+1)$ -th points is $d_n = 7.8 [(n \bmod 2) \cdot 0.21 + (1 - n \bmod 2) \cdot 0.17] \text{ nm}$, where $a \bmod b$ is the remainder of the integer division a/b . May these points describe a one dimensional linear “crystal” and if they do then what do the neighboring points correspond to or if they not why not?* **Solution** If n is an even number then $n \bmod 2$ is 0, if it is an odd number then it is 1. From the definition above the distances between consecutive points are $d_1 = 7.8 \cdot 1.7 = 1.33 \text{ nm}$ for even n/s and $d_2 = 7.8 \cdot 2.1 = 1.64 \text{ nm}$ for odd n . These distances are repeated, which means this structure has a translational symmetry, i.e. it may correspond to a crystal. These points may refer to atoms in a diatomic crystal with a two

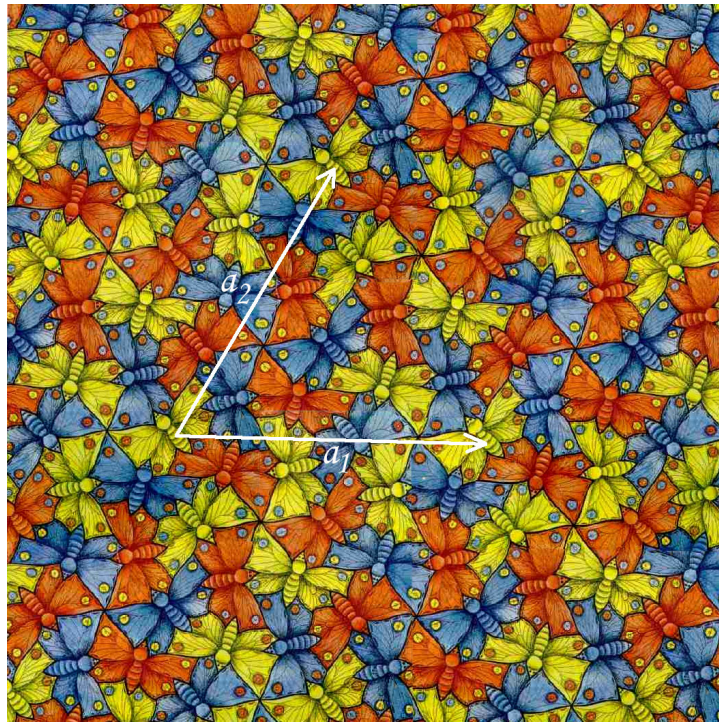


Figure 11.9: A lattice with a basis. (picture by Escher, Maurits). Imagine that the atoms of the basis are the dots on the butterflies's wings, and the point lattice is the one determined by the vectors \mathbf{a}_1 and \mathbf{a}_2 . (Can you find a smaller basis for this "crystal"?)

atom basis and a base vector of $d_1 + d_2 = 2.97nm$

The same crystal may be described with several different selection of point lattice and basis (Fig. 11.10). A cell of a point lattice may contain any number of points.

Important 11.3.6. Primitive cells *contain either a single point inside the cell's volume, or – if the cell shares several points on its surface with neighboring cells – the sum of the points divided by the number of cells that share them must be 1 (see Problem 11.4).* All cells containing more than one point are called unit cells.

Conventional unit cells *are cells that have all of the symmetries of the crystal.*

11.4 Symmetries. Bravais lattices.

In order to describe 3 dimensional periodic structures it is useful to consider the symmetries of these arrangements.

Important 11.4.1. *All crystals must have translational symmetry.*

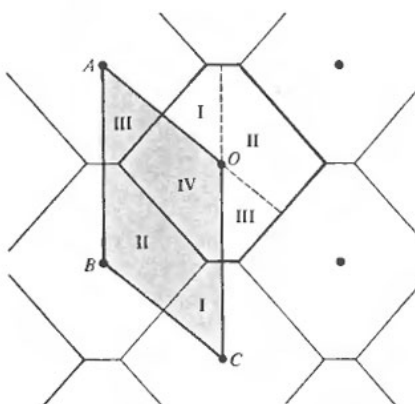


Figure 11.10: Two possible primitive cells for a 2D Bravais lattice

But translational symmetry is not the only symmetry possible. There are many other type of symmetries that can coexist with translational symmetry⁵. These include:

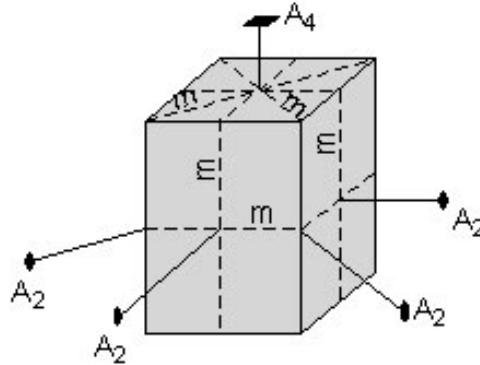
- rotation around an axis by degrees 60, 90, 120, 180.
 - These are called: *6-fold*, *4-fold*, *3-fold*, *2-fold rotations* respectively.
 - Other (e.g. 5-fold) rotations are not compatible with translational symmetry therefore cannot be present *in a crystal*⁶, because the crystallographic plane may not be covered with polygons corresponding to those symmetries (e.g. pentagons) without leaving voids or without overlap⁷.
- *mirror symmetry*
 - reflection across a plane
- *inversion* through a point (center of symmetry)
- *improper rotation* or *rotoinversion*
 - Combinations of a rotation with inversion through a point (may also be described as a rotation about an axis and a reflection in a plane perpendicular to the axis - or rotation and inversion in a point). Objects that have rotoinversion symmetry have an element of symmetry called a *rotoinversion axis*.

⁵There are 32 possible combinations of the symmetry elements which are consistent with the translational symmetry. These 32 combinations define the 32 *crystal classes*. Each crystal must belong to one of these. These classes may be grouped into 7 *crystal systems*, which are used for Bravais lattices (see below). Combining the 32 crystal classes with the 14 Bravais lattices all of the 230 possible *symmetry groups* (*space groups*) may be generated. As an example: Si is in group 225.

⁶Although we may talk about the trivial case of 1-fold rotation (rotation by 360°) too.

⁷It is still an unsolved mathematical problem whether it is possible to find a set of shapes with five-fold symmetry that together will tile the plane. This is the *five-fold tiling problem*.

Example 11.3. *Enumerate the symmetries the following “crystal” has.*



Solution

This “crystal” has the following symmetry elements:

- 1 - 4-fold rotation axis (A_4)
- 4 - 2-fold rotation axes (A_2), 2 cutting the faces and 2 cutting the edges
- 5 mirror planes (m), 2 cutting across the faces, 2 cutting through the edges, and one cutting horizontally through the center.
- There is a center of symmetry (i).

Some of these symmetries may be found in architecture (e.g. on many mosaics of Alhambra, Spain) and in the work of the Dutch graphics artist Maurits Cornelis Escher.

The name *Bravais lattice* may mean:

- The infinite set of discrete points with an arrangement *and* orientation that appears exactly the same, from whichever of the points the array is viewed⁸.
- The set of position vectors in (11.3.2)
- The set of translations determined by the position vectors

There exist 14 Bravais lattices corresponding to the possible combinations of symmetries. These are depicted in Fig.11.11.

But not all lattices are Bravais lattices, see Fig.11.13.

In Appendix 23.2 we present some example Bravais lattices.

⁸For this reason the vertices of a 2D honeycomb structure seen in Fig.11.13 do not form a 2D Bravais lattice. But the same structure with a diatomic basis does.

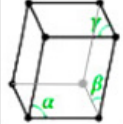
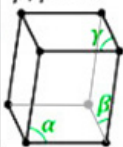
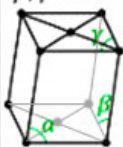
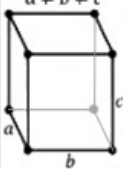
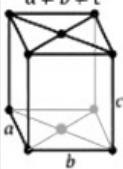
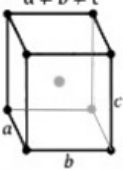
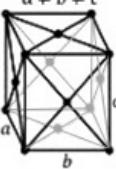
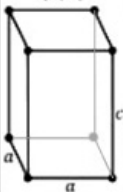
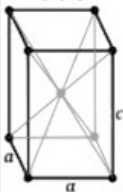
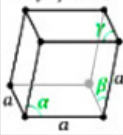
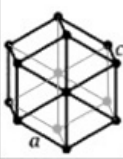
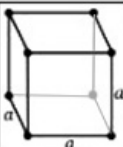
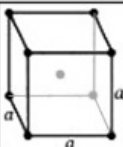
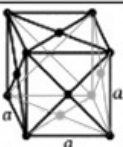
The 7 lattice systems	The 14 Bravais Lattices			
triclinic (parallelepiped)	$\alpha, \beta, \gamma \neq 90^\circ$ 			
monoclinic (right prism with parallelogram base; here seen from above)	simple	base-centered		
	$\alpha \neq 90^\circ$ $\beta, \gamma = 90^\circ$ 	$\alpha \neq 90^\circ$ $\beta, \gamma = 90^\circ$ 		
orthorhombic (cuboid)	simple	base-centered	body-centered	face-centered
	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 	$a \neq b \neq c$ 
tetragonal (square cuboid)	simple	body-centered		
	$a \neq c$ 	$a \neq c$ 		
rhombohedral (trigonal trapezohedron)	$\alpha = \beta = \gamma \neq 90^\circ$ 			
hexagonal (centered regular hexagon)				
cubic (isometric; cube)	simple	body-centered	face-centered	
				

Figure 11.11: The 14 Bravais lattices

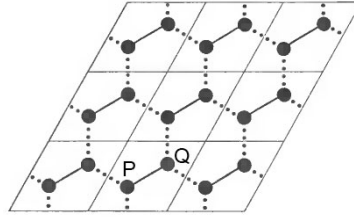


Figure 11.12: The hexagons do not form a Bravais lattice, because the orientation of the points is not the same when viewed from P or Q. An alternative description with a 2 point basis (heavy solid lines) however is a Bravais lattice.

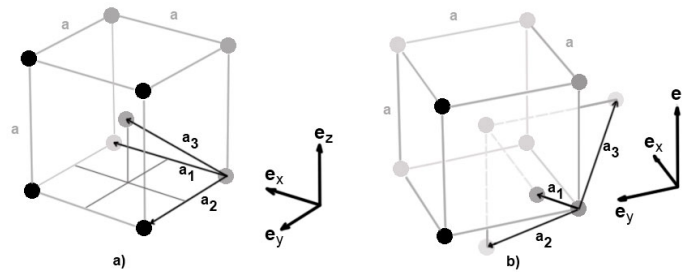
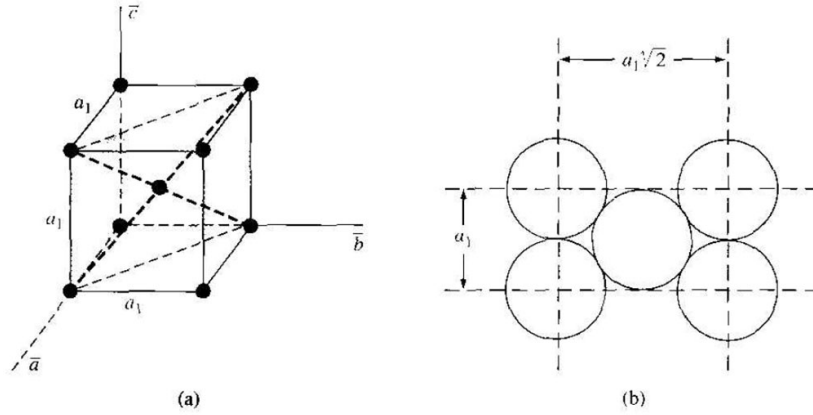


Figure 11.13: Two different primitive vector sets for the bcc lattice. a) $\mathbf{a}_1 = a\mathbf{e}_x$, $\mathbf{a}_2 = a\mathbf{e}_y$, $\mathbf{a}_3 = a\mathbf{e}_z$ b) $\mathbf{a}_1 = \frac{a}{2}(\mathbf{e}_y + \mathbf{e}_z - \mathbf{e}_x)$, $\mathbf{a}_2 = \frac{a}{2}(\mathbf{e}_z + \mathbf{e}_x - \mathbf{e}_y)$, $\mathbf{a}_3 = \frac{a}{2}(\mathbf{e}_x + \mathbf{e}_y - \mathbf{e}_z)$

Example 11.4. Calculate the surface density of atoms in a bcc crystal if the lattice constant is $a = 0.5 \text{ nm}$ ($= 5 \text{ \AA}$) and the surface plane cuts the cells diagonally and it is perpendicular to the plane of the \mathbf{a}_1 and \mathbf{a}_2 vectors⁹. **Solution**

⁹i.e. it is a (110) plane - see Miller indices below



The plane in question goes through 4 corner atoms and the middle atom of the cell. Only 1/4th of each of the cross sections of the corner atoms belong to our cell, while the cross section of the middle atom is completely inside it. Therefore the number of atoms on this plane is 2. The area of the plane is $a \cdot a \sqrt{2}$, so the density of atoms is

$$\frac{2}{a^2 \sqrt{2}} = \frac{\sqrt{2}}{(0.5 \cdot 10^{-9})^2} = 5.66 \cdot 10^{18} \frac{\text{atoms}}{\text{m}^2}$$

11.5 The Wigner-Seitz cell

An important kind of unit cell is the Wigner-Seitz (WS) primitive cell, which is the region around any lattice point that is closer to that point than to any other lattice point. For 3D lattices the shape of Wigner-Seitz cell is much more complicated. The

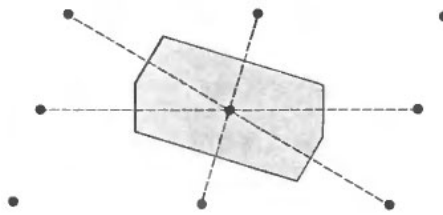


Figure 11.14: Wigner-Seitz cell for a 2D Bravais lattice

Wigner-Seitz cell will be very important in the study of the band structure of solids (see ref Brillouin-zone) where it will be called the *Brillouin zone*.

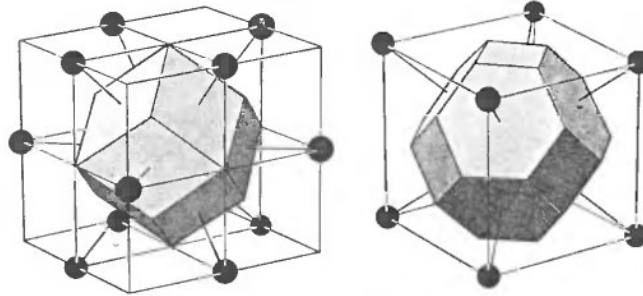


Figure 11.15: Wigner-Seitz cell for a fcc (left) and bcc (right) Bravais lattices. The Wigner-Seitz cell for the fcc lattice is not the conventional cubic cell but one in which lattice points are at the center of the cube and at the center of the 12 edges. Each of the faces is perpendicular to the line joining the central point on the center of an edge. The WS cell for the bcc lattice is the conventional cell with hexagonal and square faces. The hexagons are regular.

11.6 Non-ideal crystals. Crystal defects

In real crystals the crystal structure may deviate from the ideal periodic one because of

- lattice vibrations
- point defects
- line defects
- planar defects
- bulk defects
- finite crystals

These may change the mechanical, electrical and optical properties of solids.

11.6.1 Point Defects

Point defects are defects that occur only at or around a single lattice point. They are not extended in space in any dimension. Strict limits for how small a point defect is, are generally not defined explicitly, but typically these defects involve at most a few extra or missing atoms.

Kinds of point defects:

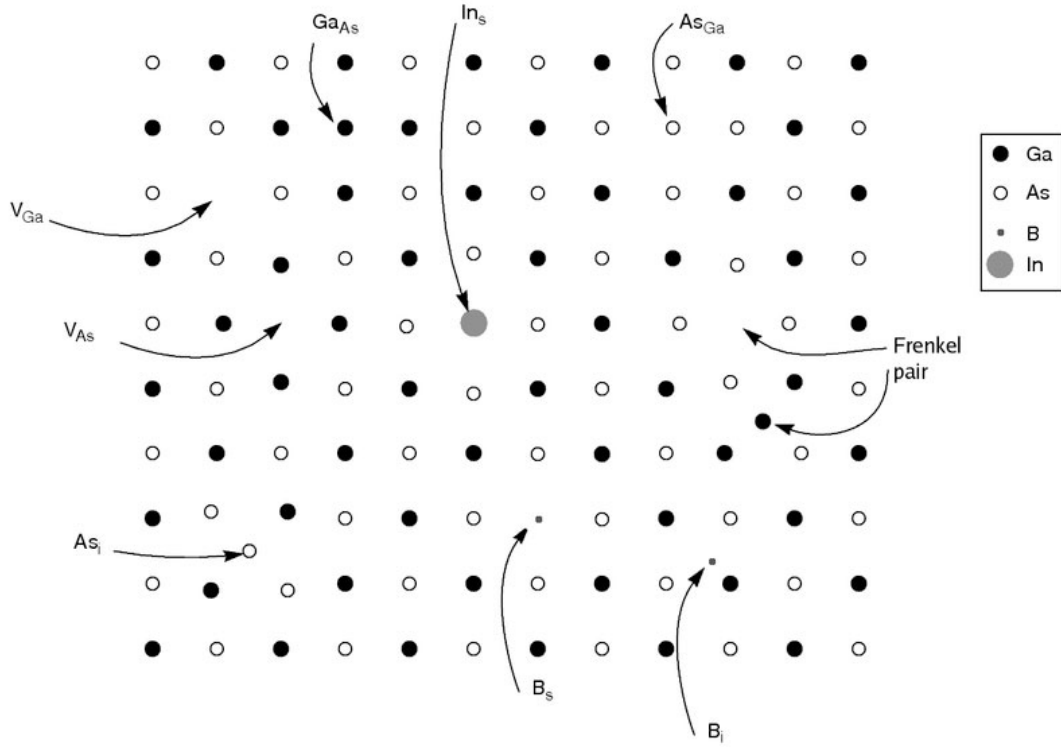


Figure 11.16: Point defects in a GaAs crystal. V_{As} and V_{Ga} denotes As and Ga vacancies, index i refer to interstitial atoms, index s to foreign substitutional atoms of which In atoms are larger and B atoms are smaller than Ga or As atoms, and As_{Ga} and Ga_{As} denote substitutional Ga and As atoms respectively.

- *Vacancy or Schottky* defects are lattice sites which would be occupied in a perfect crystal, but are vacant. Statistical physics and thermodynamics requires that in equilibrium all crystals have them¹⁰. At $T = 300\text{K}$ the ratio of the number of vacancies to the total number of atoms is about 10^{-17} , at $T = 1000\text{K}$ about 10^{-5})
- *Interstitial defects* are atoms that occupy a site in the crystal structure at which there is usually not an atom. They are generally high energy configurations. Small atoms in some crystals can occupy interstices without high energy, such as hydrogen in palladium (used in storage cells).
- *Frenkel pair* or *Frenkel defect*. A nearby pair of a vacancy and an interstitial caused when an ion moves into an interstitial site and creates a vacancy.

¹⁰The equilibrium concentrations can be calculated by either minimizing the Gibbs free energy or maximizing the entropy.

- *Substitutional atoms* when a foreign (impurity) atom occupies a lattice position.
- *Antisite defects*. Occur in an ordered alloy or compound when atoms of different type exchange positions.
- *Topological defects* Regions in a crystal where the normal chemical bonding environment is topologically different from the surroundings.

11.6.2 Line defects (*Dislocations*)

Kinds of line defects are the different dislocations. The magnitude and direction of the lattice distortion in a dislocation can be described by the so called *Burgers vector*. Imagine a perfect crystal and a closed path in it, then introduce a dislocation in the area surrounded by this path. The dislocation will break this path. The vector between connecting the two ends of this severed path is the Burgers vector.

- *Edge dislocation*: a plane of atoms terminate in the crystal. For edge dislocations the Burgers vector and the dislocation line are at right angles to each other.

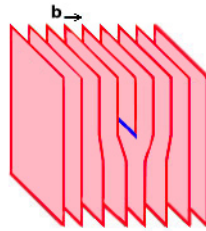


Figure 11.17: Edge dislocation. The blue line denotes the dislocation line. \mathbf{b} is the Burgers vector.

- *Screw dislocations*: the Burgers vector is parallel to the dislocation line
- *Mixed dislocations*: dislocations with the characteristics of both edge and screw dislocations, in this case the line direction and Burgers vector are neither perpendicular nor parallel

Dislocation can be created by plastic deformations. They cause mechanical stress and may move in the crystal until they interact with other dislocations. When many dislocations meet the crystal becomes brittle. ¹¹ Dislocations need not be pure line or

¹¹That is the reason why a copper wire can be broken by bending it this way and that at the same point. Bending creates dislocations. After a time the number of dislocation of the point of bending becomes so large that the wire will break.

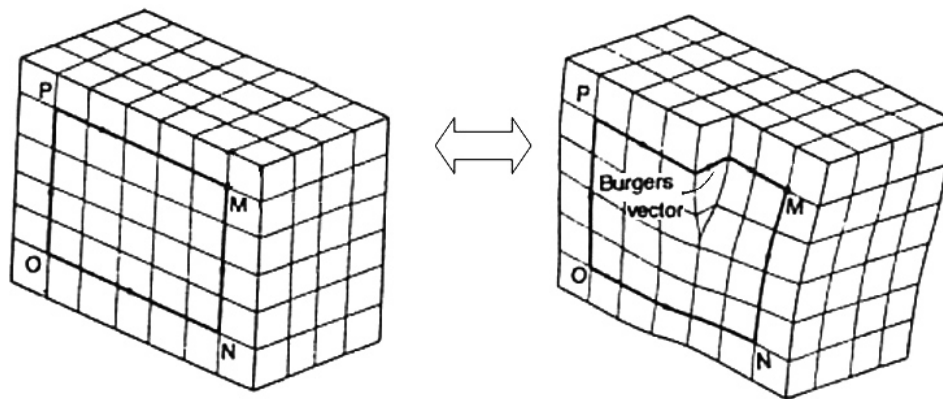


Figure 11.18: Screw dislocation. Left: ideal crystal, right: screw dislocation.

pure screw dislocations, they may also be *mixed* dislocations exhibiting the properties of both of these.

Dislocations can be observed using transmission electron microscopy, field ion microscopy and atom probe techniques.

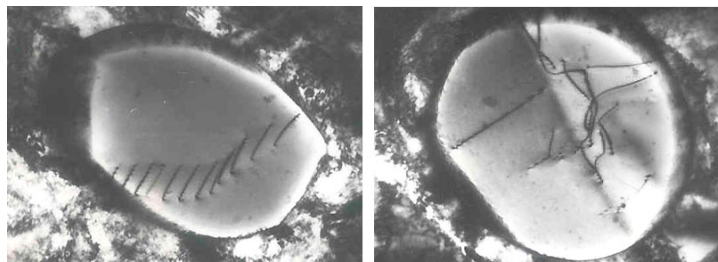


Figure 11.19: Transmission Electron Micrograph of dislocations

11.6.3 Planar defects

Kinds of planar defects are:

- *Grain boundaries* usually occurs when two crystals begin growing separately and then meet.
- *Antiphase boundaries* occur in ordered alloys: in this case, the crystallographic direction remains the same, but each side of the boundary has an opposite phase: For example, if the ordering is usually ABABABAB, an antiphase boundary takes the form of ABABBABA.

- *Stacking faults* are one or two layer interruption in the stacking sequence, for example, if the sequence ABCABABCAB were found in an fcc structure.

11.6.4 Bulk defects

Defects in the volume of the crystal. They may be:

- *Voids* small regions where there are no atoms, and can be thought of as clusters of vacancies.
- *Precipitates* Impurities can cluster together to form small regions of a different phase.

11.6.5 Effect of defects on the properties of crystals

Crystal defects affect many properties of solids. They may modify the mechanical, electrical, thermal, magnetic and chemical properties. This can be used to our advantage, but may also create problems. Whether the advantages or disadvantages dominate is usually determined by the type and concentration of these defects. The best examples when they work at our advantage are found in semiconductor devices (see Section 16.2.1), where the introduction of very few foreign atoms in otherwise very pure crystals hugely modifies their electrical properties.

Wrought iron¹² has special properties not found in other ferrous metals, mostly because of the slag inclusions it contains. A freshly fractured metal contains these in a surface concentration of about 40 000 per cm^2 . The slags contain most of the impurities present in the material, therefore wrought iron is purer than plain carbon steel.

¹²Wrought iron are no longer produced commercially. The last forge which produced it was closed in 1973.

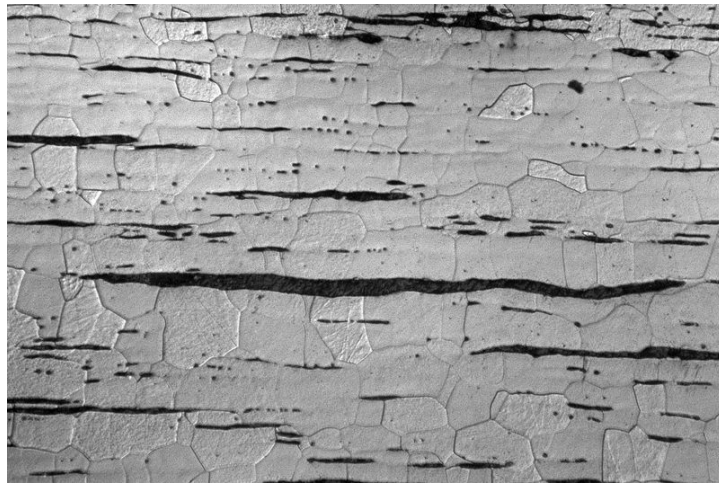


Figure 11.20: Optical Micrograph of wrought iron showing dark slag inclusions in ferrite.

Chapter 12

Determination of crystal structures by X-ray diffraction

12.1 Reciprocal lattice. Miller indices.

The structure of crystals can be studied using electromagnetic waves (X-rays) with wavelengths in the nanometer range comparable with the lattice constants. Diffraction and interference restricts the wavelengths and wave vectors of electromagnetic waves that can propagate through the periodic structure. The possible discrete wave vectors determine another periodic structure called the *reciprocal lattice*. Let us consider a plane wave $e^{i(\omega t + \mathbf{k} \cdot \mathbf{r})}$ with wave vector \mathbf{k} in any direction that propagates through a Bravais lattice with lattice constant \mathbf{R} . We can select a set of those \mathbf{K} vectors that correspond to plane waves with the periodicity of the lattice, i.e. for which

$$e^{i\mathbf{K} \cdot (\mathbf{r} + \mathbf{R})} = e^{i\mathbf{K} \cdot \mathbf{r}} \quad \forall \mathbf{r} \quad (12.1.1)$$

from which the equation of the possible \mathbf{K} vectors is

$$e^{i\mathbf{K} \cdot \mathbf{R}} = 1 \quad \forall \mathbf{K} \quad (12.1.2)$$

or equivalently:

$$\mathbf{K} \cdot \mathbf{R} = 2\pi \quad (12.1.3)$$

The fact that reciprocal vectors and wave vectors are equivalent is very important for crystallography as well as for the theory of conductivity.

12.1.1 The reciprocal lattice.

The possible values of \mathbf{K} can be considered as points of a „*k-space*” with axes k_1, k_2, k_3 , where they determine another Bravais lattice, the so called *reciprocal lattice* of the given

Bravais lattice. The original Bravais lattice is called the *direct lattice*. That the set of \mathbf{K} vectors is itself a Bravais lattice can be seen from (12.1.2) or (12.1.3) because the letters \mathbf{K} and \mathbf{R} may be interchanged in the formula. This also proves that the reciprocal lattice of the reciprocal lattice is the direct lattice.

Let us denote the three primitive lattice vectors of the reciprocal lattice with \mathbf{b}_1 , \mathbf{b}_2 and \mathbf{b}_3 respectively. As these are vectors of the reciprocal lattice (12.1.3) must be fulfilled. If each of these is related to the base vectors \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 of the direct lattice by

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij} \quad (12.1.4)$$

then they can be expressed with the following algebraic formulas¹ :

$$\begin{aligned} \mathbf{b}_1 &= 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \\ \mathbf{b}_2 &= 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1)} \\ \mathbf{b}_3 &= 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)} \end{aligned} \quad (12.1.5)$$

It is easy to show² that the inverse of these formulas are

$$\begin{aligned} \mathbf{a}_1 &= 2\pi \frac{\mathbf{b}_2 \times \mathbf{b}_3}{\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3)} \\ \mathbf{a}_2 &= 2\pi \frac{\mathbf{b}_3 \times \mathbf{b}_1}{\mathbf{b}_2 \cdot (\mathbf{b}_3 \times \mathbf{b}_1)} \\ \mathbf{a}_3 &= 2\pi \frac{\mathbf{b}_1 \times \mathbf{b}_2}{\mathbf{b}_3 \cdot (\mathbf{b}_1 \times \mathbf{b}_2)} \end{aligned} \quad (12.1.6)$$

From definition (12.1.5) it follows that the length of the l -th reciprocal base vector is

$$b_l = |\mathbf{b}_l| = \frac{2\pi}{a_l} \quad (12.1.7)$$

Important 12.1.1. *It is generally true that the same relation holds between corresponding lengths of the direct and reciprocal lattice:*

$$|\mathbf{g}_{hkl}| = \frac{2\pi}{|\mathbf{r}_{ijm}|} \quad \text{or} \quad |\mathbf{r}_{ijm}| = \frac{2\pi}{|\mathbf{g}_{hkl}|} \quad (12.1.8)$$

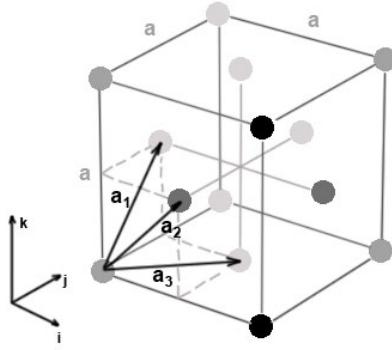
where $\mathbf{r}_{ijm} = i\mathbf{a}_1 + j\mathbf{a}_2 + m\mathbf{a}_3$ and $\mathbf{g}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ are vectors in the direct and the reciprocal lattice respectively related by using formulas (12.1.5).

¹To use symmetric formulas we used the property of the *triple scalar product* that $\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = \mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1) = \mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)$. For this reason the triple scalar product sometimes written as $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$.

²For this we need the rule of the vector triple product: $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}$

Example 12.1. *Prove that the reciprocal lattice of an fcc lattice is a bcc lattice!*

Solution Start with the following selection of primitive fcc lattice vectors:



Then the 3 primitive vectors in (12.1.5) are

$$\begin{aligned} \mathbf{a}_1 &= \frac{a}{2}(\mathbf{i} + \mathbf{k}) \\ \mathbf{a}_2 &= \frac{a}{2}(\mathbf{i} + \mathbf{j}) \\ \mathbf{a}_3 &= \frac{a}{2}(\mathbf{j} + \mathbf{k}) \end{aligned} \quad (12.1.9)$$

Determine first the denominator in (12.1.5), which is the volume of the primitive cell:

$$\begin{aligned} \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) &= \frac{a}{2}(\mathbf{i} + \mathbf{k}) \cdot \left(\frac{a}{2}(\mathbf{i} + \mathbf{j}) \times \frac{a}{2}(\mathbf{j} + \mathbf{k}) \right) \\ &= \frac{a^3}{8}(\mathbf{i} + \mathbf{k}) \cdot ((\mathbf{i} + \mathbf{j}) \times (\mathbf{j} + \mathbf{k})) \\ &= \frac{a^3}{8}[\mathbf{i} \cdot (\mathbf{i} \times \mathbf{j}) + \mathbf{i} \cdot (\mathbf{i} \times \mathbf{k}) \\ &\quad + \mathbf{i} \cdot (\mathbf{j} \times \mathbf{k}) + \mathbf{k} \cdot (\mathbf{i} \times \mathbf{j}) \\ &\quad + \mathbf{k} \cdot (\mathbf{i} \times \mathbf{k}) + \mathbf{k} \cdot (\mathbf{j} \times \mathbf{k})] \\ &= \frac{a^3}{8}(\mathbf{i} \cdot (\mathbf{j} \times \mathbf{k}) + \mathbf{k} \cdot (\mathbf{i} \times \mathbf{j})) \end{aligned}$$

here we used that \mathbf{i} , \mathbf{j} and \mathbf{k} are perpendicular to each other. Furthermore

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \mathbf{j} \times \mathbf{k} = \mathbf{i} \quad \text{and} \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

Therefore

$$\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = \frac{a^3}{4}$$

Now work with the numerators using the same formulas for \mathbf{i} , \mathbf{j} and \mathbf{k} :

$$\mathbf{a}_2 \times \mathbf{a}_3 = \frac{a^2}{4}(\mathbf{k} - \mathbf{j} + \mathbf{i}) \mathbf{a}_3 \times \mathbf{a}_1 = \frac{a^2}{4}(\mathbf{i} - \mathbf{k} + \mathbf{j}) \mathbf{a}_1 \times \mathbf{a}_2 = \frac{a^2}{4}(\mathbf{j} - \mathbf{i} + \mathbf{k})$$

Which gives us the reciprocal base vectors:

$$\begin{aligned} \mathbf{b}_1 &= \frac{2\pi}{a} \left(\frac{a^2}{4}(\mathbf{k} - \mathbf{j} + \mathbf{i}) \right) \\ \mathbf{b}_2 &= \frac{2\pi}{a} (\mathbf{i} - \mathbf{k} + \mathbf{j}) \\ \mathbf{b}_3 &= \frac{2\pi}{a} (\mathbf{j} - \mathbf{i} + \mathbf{k}) \end{aligned} \tag{12.1.10}$$

Compare (12.1.10) with vectors in Fig. 11.13 to see that we, in fact got the primitive vectors of a bcc lattice. The only difference is that the length is now. $\frac{2\pi}{a}$.

The volume of the primitive cell in a reciprocal lattice is $\frac{(2\pi)^3}{V}$, where V is the volume of the primitive cell of the original lattice. Because the reciprocal lattice of a reciprocal lattice is the original (direct) lattice, we also proved that the reciprocal lattice of a bcc lattice is an fcc lattice.

Brillouin zone

Because the reciprocal lattice is an ordinary (point) lattice itself all of the different kind of unit cells can be construed with the reciprocal lattice. There is a single difference though:

Important 12.1.2. *The Wigner-Seitz cell of a reciprocal lattice is called the first Brillouin zone.*

12.1.2 Miller indices

Important 12.1.3. *Lattice planes are imaginary planes in the crystal that contain at least 3 non co-linear points. All such planes contain an infinite number of lattice points.*

To characterize planes and directions in crystal (Bravais) lattices we use a notation system called *Miller indices*.

Let \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3 be the base vectors in the direct lattice, while \mathbf{b}_1 , \mathbf{b}_2 and \mathbf{b}_3 denote the corresponding base vectors of the reciprocal lattice. Any direction or lattice plane can be described by these vectors. Because crystals are not continuous media meaningful vectors and lattice planes require the use of only integer numbers as coefficients. These integer numbers are usually denoted by the letters h, k, l .

The following notations are used in crystallography:

For lattice planes:

(hkl) - is called the Miller index of the family of parallel lattice planes perpendicular to the direction given by the reciprocal vector³

$$\mathbf{g}_{hkl} = h \mathbf{b}_1 + k \mathbf{b}_2 + l \mathbf{b}_3.$$

By convention, negative integers are written with a bar above them, e.g. $(12\bar{3})$ for $h = 1, k = 2$ and $l = -3$, unless they are larger than 9, but such indices are rare.

The integers are usually written in lowest terms, i.e. their greatest common divisor should be 1. This means that the corresponding reciprocal vector is the shortest one between neighboring planes. From (12.1.8) the distance d_{hkl} between adjacent lattice planes is

$$d_{hkl} = \frac{2\pi}{|\mathbf{g}_{hkl}|} = \frac{1}{\sqrt{\frac{h^2}{a_1^2} + \frac{k^2}{a_2^2} + \frac{l^2}{a_3^2}}} \quad (12.1.11)$$

Examples:

Miller index (100) represents planes orthogonal to direction \mathbf{b}_1 ; index (010) represents planes orthogonal to direction \mathbf{b}_2 , and index (001) represents planes orthogonal to \mathbf{b}_3 .

$\{hkl\}$ - denotes all planes which are equivalent in the crystal due to its symmetries. E.g. in simple cubic crystals planes (100) , (010) and (001) are equivalent and $\{100\}$ means all and any of these.

For directions:

$[hkl]$ - is the Miller index of a direction in the *direct* lattice

Examples:

$[100]$ is a direction parallel to \mathbf{a}_1 , $[111]$ is parallel to $\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3$.

$\langle hkl \rangle$ - denotes all equivalent directions in the *direct* lattice. E.g. in cubic crystals $\langle hkl \rangle$ denotes either of the equivalent $[100]$, $[100]$ and $[100]$ directions.

There are two ways to calculate the Miller indices of a plane for a given crystal:

- via a point (vector) of the reciprocal lattice, or
- as the inverse intercepts along the lattice vectors in the direct lattice

Let us denote the three lattice vectors of the direct lattice that define the unit cell with $(\mathbf{a}_1, \mathbf{a}_2, \text{ and } \mathbf{a}_3)$, and the 3 primitive lattice vectors of the reciprocal lattice with $(\mathbf{b}_1, \mathbf{b}_2, \text{ and } \mathbf{b}_3)$.

³In this case h, k, l are the coordinates of the wave vector of a wave with the same periodicity as the selected lattice planes.

Method 1

The 3 integers h, k, l determine a direction (vector) in the reciprocal lattice:

$$\mathbf{g}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$$

which is the direction a plane wave with the same periodicity as the lattice planes that lie perpendicular to this direction travels. \mathbf{g}_{hkl} is a vector of k-space. The requirement of lowest terms means that it is the shortest reciprocal lattice vector in the given direction. The planes of indices (hkl) are perpendicular to this vector, and their distance is the length of \mathbf{g}_{hkl} .

Method 2

The planes of constant phase of a plane wave traveling in the direction given by \mathbf{g}_{hkl} intersects the three $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ direct lattice vectors at the three points $(\mathbf{a}_1/h, \mathbf{a}_2/k, \mathbf{a}_3/l)$ or some multiple thereof.

Derivation 12.1.1. *1cm The equation of the lattice planes is:*

$$\mathbf{K} \cdot \mathbf{r} = 2\pi A = \text{const}$$

where \mathbf{r} is any vector to a point in the plane and the factor 2π is selected to simplify the final formulas. Let us take the plane that intersects the axes of the coordinate system of the three primitive vectors at $x_1 \cdot \mathbf{a}_1, x_2 \cdot \mathbf{a}_2$ and $x_3 \cdot \mathbf{a}_3$ respectively, i.e. x_1, x_2 and x_3 are the coordinates of the intersections in the basis of the lattice vectors. From the equation

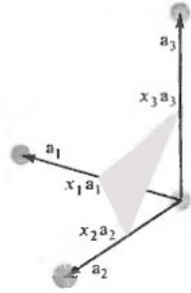


Figure 12.1: Intersection of a lattice plane with primitive vectors.

above:

$$\mathbf{K} \cdot (x_1 \cdot \mathbf{a}_1) = \mathbf{K} \cdot (x_2 \cdot \mathbf{a}_2) = \mathbf{K} \cdot (x_3 \cdot \mathbf{a}_3) = 2\pi A$$

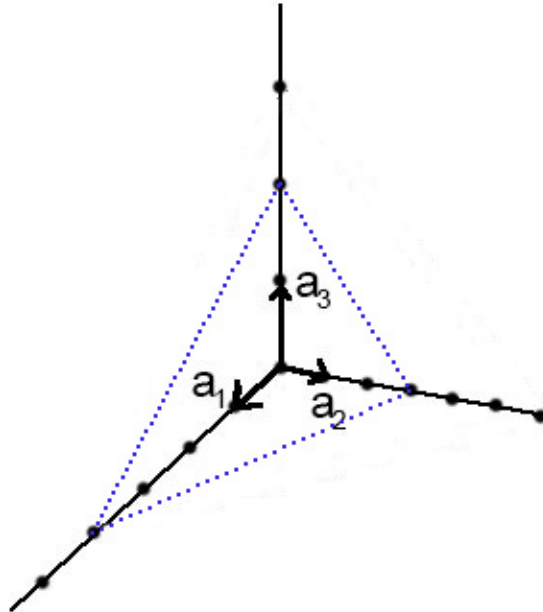
But $\mathbf{K} \cdot \mathbf{a}_1 = 2\pi h$, $\mathbf{K} \cdot \mathbf{a}_2 = 2\pi k$ and $\mathbf{K} \cdot \mathbf{a}_3 = 2\pi l$, therefore

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

That is, the Miller indices are proportional to the inverses of the intercepts of the plane, in the basis of the lattice vectors. Therefore the method of obtaining the Miller indices is that first we determine the coordinates of the 3 intersections (x_1, x_2, x_3) then find a multiplier which when applied to them gives the smallest positive integer (h, k, l) numbers.

If one of the indices is zero, it means that the planes do not intersect that axis (the intercept is "at infinity").

Example 12.2. Determine the Miller indices for the plane on the figure!



Solution The intersections with the three axes are at $4a_1$, $3a_2$ and $2a_3$. Then the inverse intercepts in lattice vector units are:

$$\frac{1}{4}, \frac{1}{3}, \frac{1}{2}$$

To get integer numbers we have to calculate the lowest common denominator of this fraction, which is 12. Multiplying each fraction with 12 gives the three Miller indices: (346)

Example 12.3. In the previous example let the length of all the three base vectors of the direct lattice $a = 5\text{nm}$. Determine the distance of the lattice planes (346) . **Solution**

The three basis vectors are of the same length and they are perpendicular to each other. Therefore the three reciprocal base vectors will also be of the same length and perpendicular to each other. By substitution into (12.1.5) this length is $b = \frac{2\pi}{a} = 1.26\text{nm}^{-1}$. The length of the reciprocal vector perpendicular to the (346) planes is

$$|b_{346}| = \sqrt{3^2 + 4^2 + 6^2} \cdot \frac{2\pi}{a} (= \sqrt{61} \cdot 1.26 = 9.81\text{nm}^{-1})$$

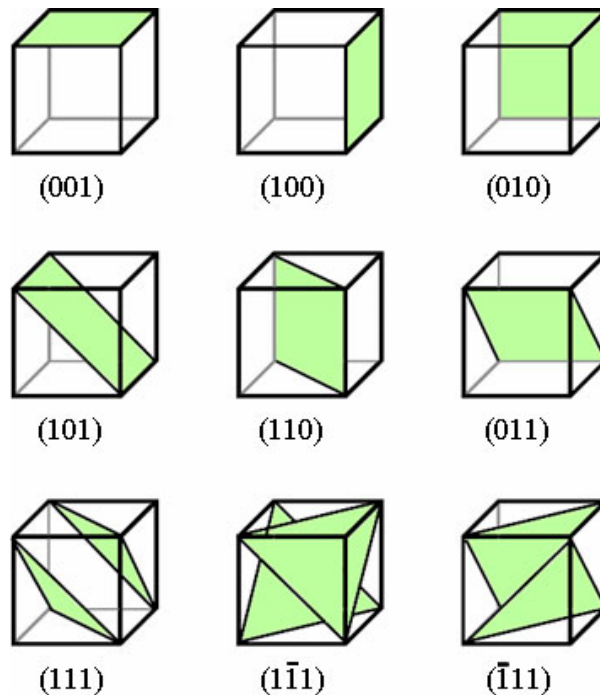
The distance of the planes (346) therefore is

$$\underline{\underline{d_{346} = \frac{a}{\sqrt{61}} = 0.64\text{nm}}}$$

Example 12.4. The 3 base vectors of a crystal are orthogonal and 2.3, 3.4 and 4.5 nm long. What is the distance of the lattice planes (211) ? **Solution** The distance can be calculated from (12.1.11):

$$d_{211} = \frac{1}{\sqrt{2^2/2.3^2 + 1/3.4^2 + 1/4.6^2}} = 1.06\text{nm} \quad (12.1.12)$$

Example 12.5. Draw all 9 lattice planes and determine the Miller indices in a simple cubic Bravais lattice. **Solution**



(Where is the origin of the 3 lattice vectors in the cubes?)

12.2 Determination of crystal lattices by X- ray diffraction. Bragg and Laue formulas

Atomic distances are in the range of $\sim 10^{-10}m$. To study them we need an electromagnetic wave with the same wavelength. The corresponding photon energy is $\mathcal{E} = \frac{hc}{\lambda} \sim 10^4$ eV and this is the characteristic X-ray wavelength and the reason X-rays are used in crystallography.

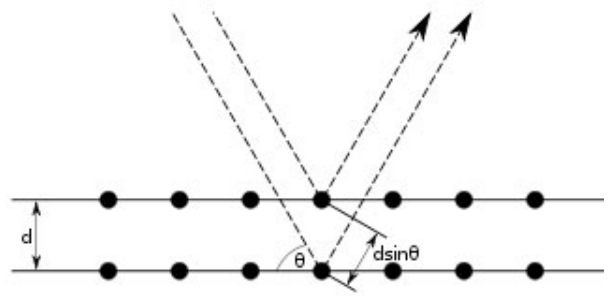
What will we see? Because X-rays are invisible (and harmful) for us the crystal structures are either determined from some “image” of the crystal recorded on film or on digital media, or from X-ray intensities collected by a detector. When a crystal is irradiated by X-rays of suitable wavelengths it will diffract them. So our first task must be the description of X-ray diffraction.

12.2.1 Bragg diffraction formula

or *Bragg’s law*

The Braggs⁴ proposed in 1913 that X-rays are reflected back from lattice planes. Constructive interference occurs when :

$$2d \sin\Theta = n\lambda \quad (12.2.1)$$



6

Figure 12.2: A plane wave approaches a crystalline solid and is reflected back by the lattice planes. This wave is represented by two rays on the figure. The lower ray traverses an extra length of $2d\sin\Theta$. Constructive interference occurs when this length is equal to an integer multiple of the wavelength of the radiation: $2d\sin\Theta = n\lambda$. The *Bragg angle* Θ is half of the total angle the incident beam is deflected.

⁴William Lawrence Bragg and his father William Henry Bragg.

The advantage of this approach is its simplicity but this is also the reason why it is not enough: Bragg's law only gives the directions of the diffracted waves and not their amplitudes. Furthermore physical explanation is lacking as the lattice planes are not "real" planes and the diffraction in fact is caused by the atoms of the crystal.

Example 12.6. *Determine the possible diffraction angles for an X-ray of 10 keV from the (111) planes of a simple cubic lattice, if the lattice constant is $a = 5.3$.* **Solution**
The wavelength of the X-ray:

$$E = h\nu = \frac{hc}{\lambda} \Rightarrow \lambda = \frac{hc}{E} = 1.24 \cdot 10^{-10} \text{m}$$

To apply Bragg's law

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

we need to calculate the distance of the lattice planes of Miller indices (111), which are planes going through 3 non adjacent corner of the cube. The distance of two such planes that intersects a primitive cell (see the figure in the previous example) is 3rd of the body diagonal:

$$d = \frac{1}{3} \sqrt{3} a = 0.305 \text{nm}$$

The possible diffraction angles are determined by:

$$\sin\theta = n \frac{\lambda}{2d} = 0.20325 n$$

Here $n = 1, 2, 3, 4$, i.e. the angles are

$$11.72^\circ, 23.99^\circ, 35.57^\circ, \text{ and } 54.39^\circ$$

A more exact derivation of the diffraction equation was given by Max von Laue.

12.2.2 Laue equations

When the X-ray of wave vector \mathbf{k} interacts with the atoms in the crystal it excites them. During the transition to their ground state atoms themselves emit waves with the same frequency (and wave number of the same magnitude: $k' \equiv |\mathbf{k}'|$) as those of the incoming X-ray's ($k = |\mathbf{k}|$). The resulting diffracted wave will be the result of the interference of the incoming and emitted X-rays. Now consider two atoms separated by the lattice constant d . Both of the atoms will emit interfering waves.

For constructive interference the total path difference between the diffracted waves is an integer multiple of the wavelength (see Fig. 12.3). Let us denote the normal vectors

of the incoming and diffracted waves with $\mathbf{n} := \mathbf{k}/k$ and $\mathbf{n}' := \mathbf{k}'/k$ respectively. Then the condition for constructive interference can be written as

$$\mathbf{d} \cdot (\mathbf{n} - \mathbf{n}') = m \cdot \lambda, \quad m = 1, 2, \dots$$

Using the definition of $k \equiv |\mathbf{k}| = \frac{2\pi}{\lambda}$ and the fact that $|\mathbf{k}| = |\mathbf{k}'|$ both the λ and the normal vectors can be eliminated:

$$\mathbf{d} \cdot (\mathbf{k} - \mathbf{k}') = 2\pi m$$

But \mathbf{d} itself must be a Bravais lattice vector \mathbf{R} , because lattice sites are displaced from one another by Bravais lattice vectors:

$$\mathbf{R} \cdot (\mathbf{k} - \mathbf{k}') = 2\pi m \quad \Rightarrow \quad e^{i\mathbf{R} \cdot (\mathbf{k} - \mathbf{k}')} = 1$$

so $\mathbf{K} \equiv (\mathbf{k} - \mathbf{k}')$ must be a vector of the reciprocal lattice.

Comparing Figs 12.2 and 12.3 we see that the two descriptions concerning the direction of the diffracted wave are equivalent. The condition that the wave vector difference must be a vector of the reciprocal lattice combined with the relation between the reciprocal lattice vectors and the primitive vectors of the crystal leads to the Laue equations:

$$\begin{aligned} \mathbf{a}_1 \cdot \Delta\mathbf{k} &= 2\pi h \\ \mathbf{a}_2 \cdot \Delta\mathbf{k} &= 2\pi k \\ \mathbf{a}_3 \cdot \Delta\mathbf{k} &= 2\pi l \end{aligned} \tag{12.2.2}$$

For example the first of these equations says that $\Delta\mathbf{k}$ must lie on a cone around \mathbf{a}_1 . So to satisfy all of these conditions $\Delta\mathbf{k}$ must lie on the intersection of 3 cones: one around each base vectors.

The Geometrical Structure Factor

If we take into account that the primitive cell may contain not just a single atom but also a basis then rays scattered from the atoms of the basis may modify the amplitude of the scattered wave even when the Laue conditions are satisfied. Let us consider a monatomic lattice with a basis (e.g. diamond). If the vectors of the atoms of the basis are $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$ and there exists a Bragg peak associated with a change of the wave vector $\mathbf{K} = \mathbf{k}' - \mathbf{k}$ then the phase difference between the rays scattered at \mathbf{d}_j and \mathbf{d}_k will be $\mathbf{K} \cdot (\mathbf{d}_j - \mathbf{d}_k)$ and the amplitude of the two rays will differ by a factor of $e^{i\mathbf{K} \cdot (\mathbf{d}_j - \mathbf{d}_k)}$. Thus the amplitudes of the rays scattered by any scattering center \mathbf{d}_m will be proportional to $e^{\mathbf{K} \cdot \mathbf{d}_m}$, and the net amplitude will be proportional to

$$S_k = \sum_{m=1}^n e^{i\mathbf{K} \cdot \mathbf{d}_m}$$

S_k is called the *geometrical structure factor* and expresses the extent to which interference of the waves scattered from identical ions within the basis can diminish the intensity of the Bragg peak associated with the reciprocal lattice vector \mathbf{K} . The intensity contains a factor of $|S_k|^2$

Other factors⁵ also modify the intensity so the only case when the structure factor can be used with assurance is when it vanishes. This is the case when the Bragg peak disappears completely.

12.3 X-ray diffraction methods.

X-ray diffraction with constructive interference can only occur whenever Bragg's law is satisfied. With monochromatic (single wavelength) radiation, and an arbitrary orientation of a single crystal to the X-ray beam will not generally produce any diffracted beams. There would therefore be very little information in a single crystal diffraction pattern from using monochromatic radiation unless special conditions are met.

This problem can be overcome by continuously varying either λ or Θ or both over a range of values, to satisfy Bragg's law. Practically this is done by:

- using a range of X-ray wavelengths (i.e. white radiation), or
- by rotating the crystal or,
- by using a powder or polycrystalline specimen.

The detailed description of the three basic methods corresponding to these possibilities are in Appendix 23.3. Here we present the short summary of these only.

12.3.1 The Laue Method

The Laue method is mainly used to determine the orientation of large *single crystals* whose structure is known. White radiation of wavelengths between λ_{min} and λ_{max} and of a *fixed direction* is reflected from, or transmitted through, a *single crystal of fixed orientation*. A sheet film perpendicular to the incident beam records the diffraction points, which lie on curves. Each curve corresponds to a different wavelength.

12.3.2 The Rotating Crystal Method

In the rotating crystal method, a single crystal is mounted with an axis normal to a monochromatic x-ray beam. A cylindrical film is placed around it and the crystal is

⁵X-rays are scattered by the electron cloud of the atom. An *atomic form factor* of this must also be taken into account. E.g. X-rays are not very sensitive to light atoms, such as hydrogen and helium, and there is very little contrast between elements adjacent to each other in the periodic table.

rotated about the chosen axis. As the crystal rotates, sets of lattice planes will at some point make the correct Bragg angle with the monochromatic incident beam, and at that point a diffracted beam will be formed. The main application of the rotating crystal method is the determination of unknown crystal structures.

12.3.3 The Debye-Scherrer Powder method

The powder method is used to determine the value of the lattice parameters accurately. Lattice parameters are the magnitudes of the unit vectors a_1 , a_2 and a_3 which define the unit cell for the crystal. The crystal is finely grounded and the resulting powder containing thousands of crystallites are put into the monochromatic X-ray beam. A circle of film is used to record the diffraction pattern.

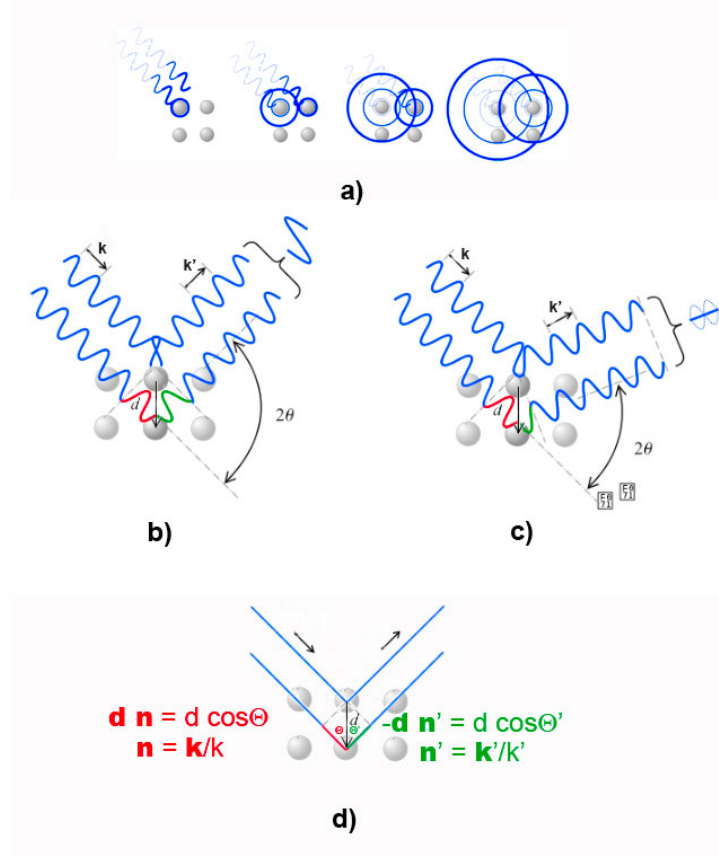


Figure 12.3: a) A plane electromagnetic wave approaches a crystalline solid and interacts with the atoms. This wave is represented by two rays on the figures. b) When the Bragg condition ($2d \sin \Theta = n \lambda$) is fulfilled constructive interference occurs. c) When the Bragg condition is not fulfilled the interference may be destructive. d) The path difference between incoming and outgoing waves from neighboring atoms may be expressed through the normal vectors ($\mathbf{n} = \mathbf{k}/k$, $\mathbf{n}' = \mathbf{k}'/k'$) and the displacement vector \mathbf{d} of the two atoms. The two angles between \mathbf{d} and the two waves may differ. The total path difference is: $\mathbf{d}(\mathbf{n} - \mathbf{n}')$

Chapter 13

Theory of lattice vibrations

Classical theory of lattice vibrations

In a real crystal the atoms are not fixed into rigid lattice sites, but are vibrating around the lattice points, their equilibrium positions. The understanding of lattice vibrations are important because of many reasons. They determine sound propagation through the crystal and the thermal properties of it too. They have an effect on electron propagation and even affect the light absorption and emission of the crystals.

In a crystal the vibration of every atoms is coupled with the vibration of the neighboring atoms. At the first sight this is an incredibly complicated problem, but we will see how this problem can be reduced to a much more benign one by using a simple, but powerful model. As a result the vibrations of the system may be described by harmonic waves with \mathbf{k} wave vectors and $\omega(\mathbf{k})$ frequencies.

In the study of the lattice vibrations we must answer the following questions:

1. Is it possible to use a classical physical model?
2. Which is the simplest solvable model that describes them?
3. What controls the $\omega(\mathbf{k})$ relation?
4. What is the amplitude of the vibrations and how does it depend upon the temperature?
5. What is the specific heat of a crystal and how does it depend upon the temperature?
6. If a classical physical model is possible how does it relate to the quantum mechanical model?

Luckily the answer to the first question is yes: a classical physical model exists that explains some of the properties of the lattice vibrations.

Our classical physical model uses the following assumptions:

- The equilibrium position of each ion is on a Bravais lattice site \mathbf{R}
- The typical excursions of each ion from its equilibrium position are small, therefore atomic forces are harmonic.¹

The simplest model of a crystal lattice is a 1D, monatomic linear chain with harmonic forces between the atoms.

13.1 Monatomic linear chain, phonons

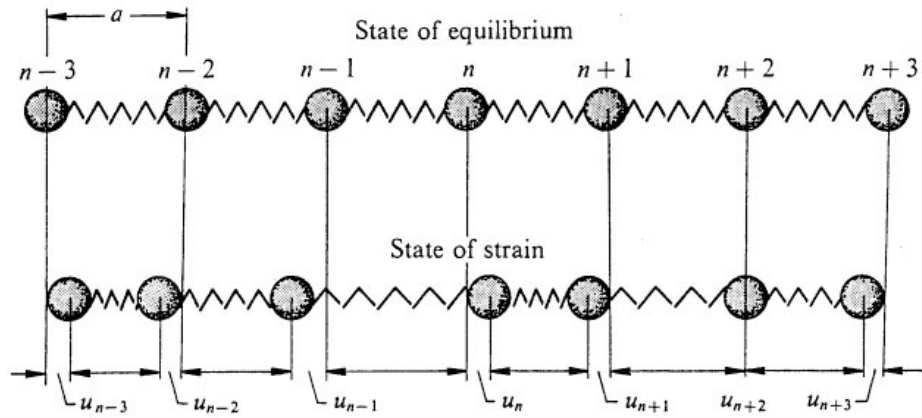


Figure 13.1: Section of a *Monatomic Linear Chain* In equilibrium the crystal displays translational symmetry. In a state of strain the n -th atom is displaced a distance u_n from its equilibrium position. If the restoring forces are harmonic only nearest neighbor interactions need to be considered

The equation of motion for the n -th atom inside an N atom linear chain:

$$M \frac{d^2 u_n}{dt^2} = \beta(u_{n+1} - u_n) - \beta(u_n - u_{n-1}) = \beta(u_{n+1} - 2u_n + u_{n-1}) \quad (13.1.1)$$

There is a problem with the atoms at the ends of the chain, which only have one neighbor. In other words we need a well behaved boundary condition.

The simplest and most convenient one is the *Born - von Karman periodic boundary condition*:

$$u_{N+1} = u_1$$

This boundary condition may be achieved in two equivalent ways:

¹This harmonic approximation is broken near the melting point.

- connect the last atom on the chain with the first using a massless rigid rod
- create a ring from the chain

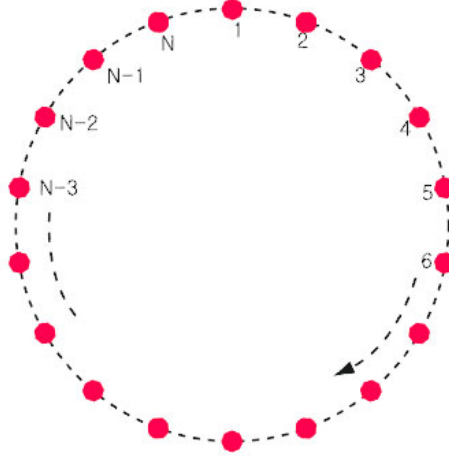


Figure 13.2: Born Karman periodic boundary condition for the monatomic linear chain model

Let us try to use a test solution of the form

$$u_n = u_o e^{\pm i(\omega t + k n a)} \quad (13.1.2)$$

The result of the calculations (see Appendix 23.4.1)

$$\omega(k) = 2 \sqrt{\frac{\beta}{M}} \sin \frac{1}{2} k a \quad (13.1.3)$$

The resulting $\omega(k)$ function is called the *dispersion relation* of the lattice vibrations, depicted in Fig. 13.3. This dispersion relation has a maximum at $k = \pm \frac{\pi}{a}$.

Every possible ω represents a special *vibrational mode* for the system. $\omega(k)$ is a periodic function of k .

The ansatz (or probe function) in (13.1.2) is a continuous function of k , but k can only take the discrete values (see (23.4.3)) determined by the periodic boundary conditions. Consequently the set of possible $\omega(k)$ values will also be discrete. A linear chain of N atoms thus may only have N different discrete vibrational modes, although the number of the possible k -s is infinite. (See Fig. 13.4.) Because for even a small solid the number of atoms N are in the order of 10^{23} , the difference between neighboring k values

$$\Delta k = \frac{2\pi}{N a} \quad (13.1.4)$$

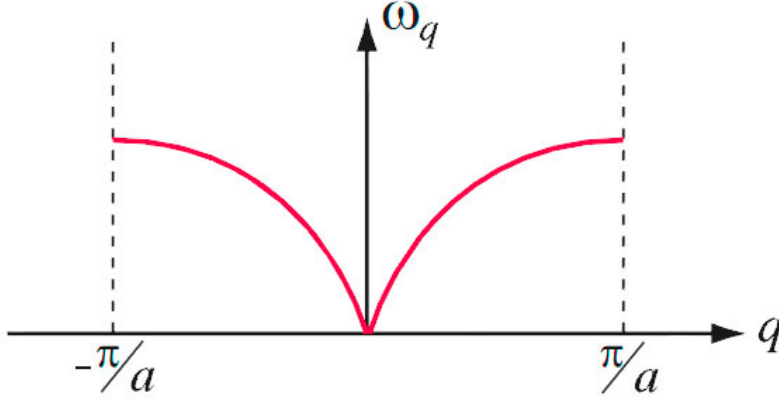


Figure 13.3: Dispersion relation for the monatomic linear chain

is too small to be measured or even observed. So we may treat k as a (*quasi*)continuous quantity.

If $|k| \ll \frac{\pi}{a}$ then $\omega = c \sqrt{\frac{\beta}{M}} k$ where $c \equiv \frac{\omega}{k}$ the phase velocity of sound in the medium. The group velocity of the “waves” describing the lattice vibrations from (13.1.3) is

$$v_g = \left(\frac{d\omega(k)}{dk} \right) = a \sqrt{\frac{\beta}{M}} \cos \frac{1}{2} ka$$

At $k = 0$ $v_g = c$. In many solids $c \simeq 10^3$ m/s and $a \simeq 10^{-10}$ m, so the maximum of ω is about 10^{12} Hz, which is in the IR range.

The dispersion is periodic in k : $\omega(k) = \omega(k + \frac{2\pi}{a})$, but we only sample the wave at atomic positions, so we cannot tell waves with k and $k + \frac{2\pi}{a}$ apart.

Conventionally, we only consider the wave vectors between $-\frac{\pi}{a}$ and $\frac{\pi}{a}$. This region corresponds to a primitive unit cell in reciprocal space whose boundaries are half way between neighboring points, i.e. they lie in the Wigner-Seitz cell of the reciprocal space. This is an important quantity and have its own name: we call it the (first) *Brillouin zone*.

The total energy of the vibrating linear chain is the sum of the kinetic and potential energies of all of the vibrating atoms, which depends on the difference $u_n(t) - u_{n+1}(t)$ and its derivative. We can calculate it using formula (23.4.6) from the Appendix 23.4.1, which states that $u_n(t) = \sum_k \chi_k(t) e^{ikna}$. The result of the calculation is:

$$\mathcal{E}_{tot} = \sum_k \left(\frac{1}{2M} p_k p_k^* + \frac{1}{2} M \omega^2(k) q_k q_k^* \right) \quad (13.1.5)$$

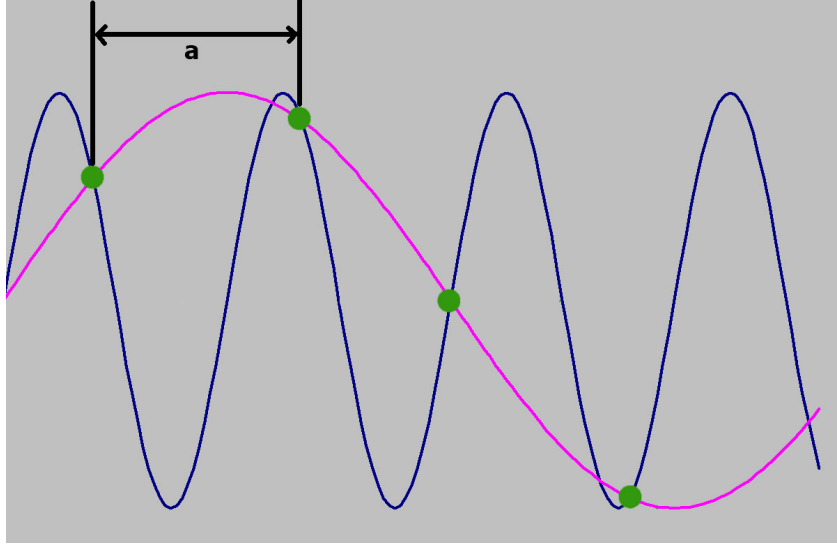


Figure 13.4: The wave represented by the blue curve conveys no additional information to that given by the purple one. Only wavelengths longer than $2a$ are needed to represent the motion of the atoms and the waves only have meaning at the lattice points.

where we introduced the $q_k(t)$ and $p_k(t)$ *normal coordinates* with:

$$q_k(t) := \frac{1}{\sqrt{N}} \sum_n u_n(t) e^{-ikna} \quad (13.1.6)$$

$$p_k(t) := \frac{1}{\sqrt{N}} \sum_n p_n(t) e^{-ikna} \quad (13.1.7)$$

i.e. the total energy may be written as the sum of energies of a system of independent linear harmonic oscillators with frequencies $\omega(k)$! We know from quantum mechanics that the energy of a linear harmonic oscillator may only change in quanta of $\hbar\omega$. So from a quantum mechanical point of view the possible *modes* of lattice vibrations can be characterized by their frequencies. The quanta of the lattice vibrations are called *phonons*.

13.2 Diatomic linear chain. Optical and acoustical branches of the dispersion relation

Let us consider a case where the linear lattice have a 2 atom basis of masses M_1 and M_2 whose distance in equilibrium is b (i.e. the lattice constant $a = 2b$), and let the interaction be the same between the different atoms! (13.5)

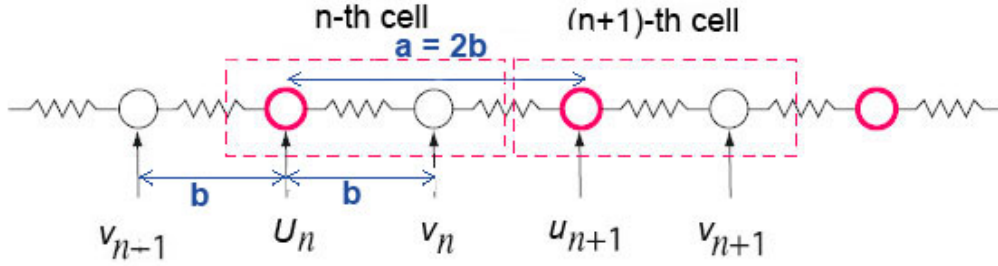


Figure 13.5: Linear chain with 2 atoms in a unit cell. The equilibrium distance of any two atoms are b , the lattice constant $a = 2b$.

In this case there are 2 sets of N coupled equations:

$$M_1 \frac{d^2 u_n}{dt^2} = \beta ((v_n - u_n) - (u_n - v_{n-1})) \quad (13.2.1)$$

$$M_2 \frac{d^2 v_n}{dt^2} = \beta ((u_{n+1} - v_n) - (v_n - u_n)) \quad (13.2.2)$$

Try the solutions in the form

$$u_n = u_k e^{i(\omega t + k n a)} \quad (13.2.3)$$

$$v_n = v_k e^{i(\omega t + k n a)} \quad (13.2.4)$$

The detailed derivation is in Appendix [23.4.2](#)

The solution:

$$\omega_{\pm}^2 = \frac{\beta}{M_1 M_2} \left(M_1 + M_2 \pm \sqrt{(M_1 + M_2)^2 - 4 M_1 M_2 \sin^2 k b} \right) \quad (\text{where } a = 2b) \quad (13.2.5)$$

The 2 solutions for ω are in Fig. [13.6](#). If $k = 0$ then $\sin^2 k b = 0$ and

$$\omega_-(0) = 0$$

$$\omega_+(0) = \sqrt{\frac{2\beta(M_1 + M_2)}{M_1 M_2}}$$

Let us suppose that $M_1 < M_2$! Then at $k = \frac{\pi}{2b}$ where $\sin^2 k b = 1$

$$\omega_-\left(\frac{\pi}{2a}\right) = \sqrt{\frac{2\beta}{M_2}}$$

$$\omega_+\left(\frac{\pi}{2a}\right) = \sqrt{\frac{2\beta}{M_1}}$$

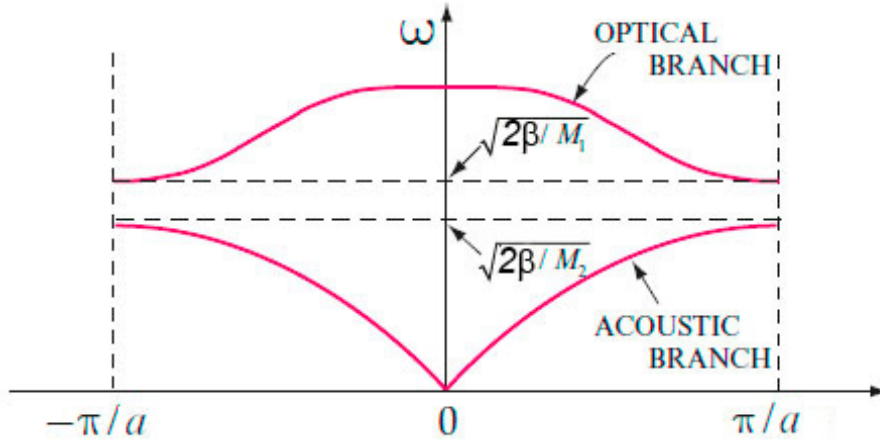


Figure 13.6: Dispersion relation of a linear chain with 2 atoms in a cell.

The curves on this dispersion relation could also be extended periodically outside the first Brillouin zone, because again k and $k + \frac{2\pi}{b}$ give the same ω .

The lower branch, which is similar to the one of the monatomic linear chain is called the *acoustic* branch, the upper one the *optical* branch. To understand the origin of the names let us visualize the phase of the motion of the two atoms in the unit cell of length $a = 2b$!

The ratio of the displacement of the 2 kinds of atoms u_k/v_k at the long wavelength ($k \approx 0$) limit can be calculated e.g. by substituting $\omega(0)$ into e.g. (13.2.1) which results in

$$\begin{aligned} \frac{u_k}{v_k} &> 0 && \text{acoustic branch} \\ \frac{u_k}{v_k} &< 0 && \text{optical branch} \end{aligned}$$

This means that in the optical branch the different kind of atoms of the linear chain vibrate with opposing phases, while in the acoustic branch the phases are the same. If this linear chain contains positively charged M_1 and negatively charged M_2 ions then it is easy to see that this kind of vibration can be caused by an incident electromagnetic wave of optical frequencies, which acts on the + and - charged atoms in opposite directions. This is why we call this branch the optical branch. On the other hand in the acoustic branch the atoms move in the same phase like they would in a mechanical (sound) wave.

At the limits of the Brillouin zone in the acoustic branch $\omega^2 = 2\beta/M_1$ and in the

optical branch $\omega^2 = 2\beta/M_2$ where $M_1 < M_2$. From here it follows that

$$\begin{aligned} \frac{u_k}{v_k} &= \infty & \text{at } \pm \frac{\pi}{a} & \text{ in the acoustic branch} \\ \frac{v_k}{u_k} &= \infty & \text{at } \pm \frac{\pi}{a} & \text{ in the optical branch} \end{aligned}$$

i.e. the two sub-lattices of atoms M_1 and M_2 act as if decoupled: one lattice remains at rest while the other lattice moves.

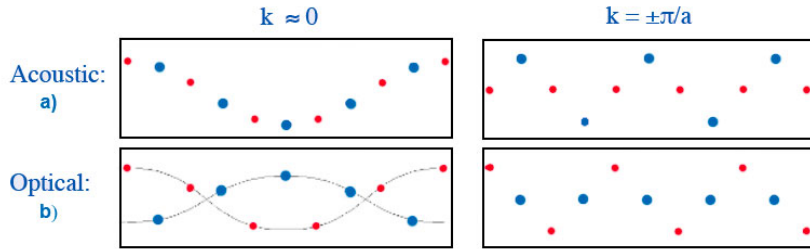


Figure 13.7: The long ($k \simeq 0$) and short ($k \simeq \pi/a$) wavelength limit acoustic (a) and optical (b) modes in a diatomic linear chain

Notice that when $M_1 = M_2$ there is no gap between the two branches of the dispersion relation. We expect that because in that case we have a monatomic chain. But the dispersion relation of a monatomic chain only has a single branch and we seemingly have two albeit touching ones. But wait! The diatomic chain ($M_1 \neq M_2$) had a lattice constant of $a = 2b$, while the lattice constant for the monatomic chain ($M = M_1 = M_2$) is b . I.e. the Brillouin zone for the diatomic chain ($[-\frac{2\pi}{b}, \frac{2\pi}{b}]$) is twice as large as that for the monatomic chain ($[-\pi/b, \pi/b]$) and the part of the band that we called the optical branch for the diatomic chain is equivalent to the part of the acoustic band between $[-\frac{2\pi}{b}, -\frac{\pi}{b}]$ and $[\frac{\pi}{b}, \frac{2\pi}{b}]$

13.3 Three dimensional lattices

In 3D the equations are much more complicated. There are harmonic forces between all of the atoms, so for instance if we have a Bravais lattice containing N points with an n atom basis the potential energy of the lattice vibrations, will depend on every displacement vectors of every atoms of the basis at every lattice site. Details are in Appendix 23.4.3.

In 3D k will become a vector (\mathbf{k}) and the dispersion relation will depend on the direction of \mathbf{k} as well. As a consequence there will be 3 independent acoustic branches: 1 longitudinal and 2 transverse ones, and if the crystal has a basis of n atoms then there

will be $3n - 3$ optical branches i.e. $n - 1$ longitudinal and $2n - 2$ transverse branches. In special high symmetry directions these branches become degenerate, i.e. some of them will overlap.

The energy of a transverse acoustic (TA) branch is usually lower than that of the longitudinal acoustic (LA) branches.

To display a 3D dispersion relation a special method is used, based on our experience in 1D.

We know the branches with $k > 0$ and $k < 0$ are symmetrical so it is enough information if we display one half of them only, and because they are periodic with the period of the Brillouin zone it is enough to confine the display to the first Brillouin zone. This makes it possible to display quite complicated 3D dispersion relations on a single 2D graphics. The method is the following:

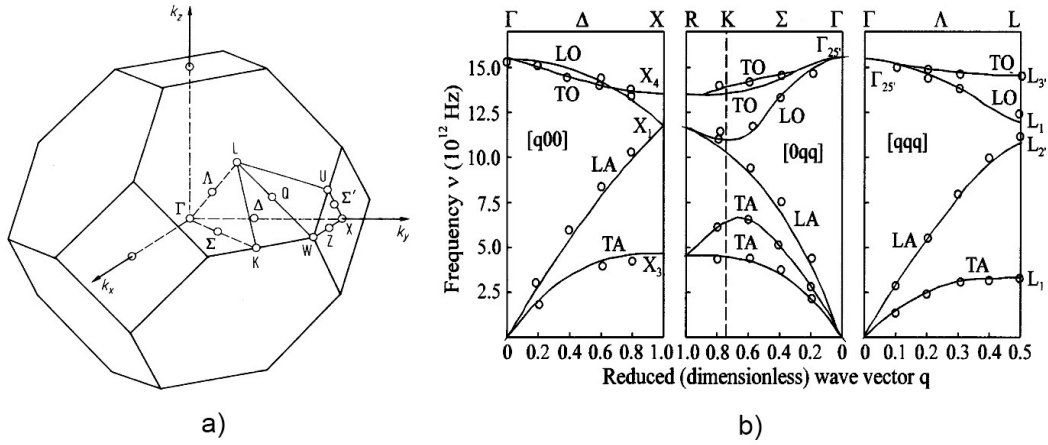


Figure 13.8: High symmetry points and the corresponding dispersion relations in Si. Γ is the origin (0,0,0), $X = (1/2, 0, 1/2)$, $L = (1/2, 1/2, 1/2)$, $W = (1/2, 1/4, 3/4)$, etc are high symmetry points (all coordinates are in units of π/a), Δ , Σ and Λ , etc are the lines connecting them. K , R , U , etc are some special points along the lines

Select high symmetry points in k space (Fig. 13.8). Create a set of straight lines connecting them². Start from one of the high symmetry points, calculate and draw the dispersion relation along these lines between the selected points in k space adjacent to each other as in Fig. 13.8

²E.g. along the Δ line \mathbf{k} is parallel with the 3rd axis of k -space and X is its endpoint at the Brillouin-zone boundary.

Quantum theory of lattice vibrations

In the previous sections of this chapter we used a simple classical model of the lattice vibrations of solids. We found collective vibrational modes (*normal modes*) that are linear harmonic oscillations. According to quantum mechanics (Chapter 3.5.6) the energy of such oscillations is quantized and the total energy is $\mathcal{E}_{tot} = \hbar\omega_s(k)(n + 1/2)$, where s is the branch index and n is the excitation number. As a convenience we can describe these quanta in a corpuscular manner as in quantum mechanics. There we described the electromagnetic radiation as a collection of radiation modes characterized by their wave numbers k and polarization states and found their energy quantized with a total energy of $\mathcal{E} = \hbar\omega(n + 1/2)$, where n is the number of quanta. These quanta are called photons. Analogously the quantum of energy of lattice vibrations of type s having a wave vector of \mathbf{k} is called a *phonon* and the number $n_{\mathbf{k},s}$ gives the number of phonons present in the crystal.

Although the language of phonons is more convenient than that of normal modes, the two nomenclatures are completely equivalent.

13.4 Specific heat of lattice vibrations

Because the number of ions (atoms) in a crystal is very large the discrete spectrum of the lattice vibrations may be considered continuous (see (13.1.4)). And because phonons in a suitable frequency range describe sound waves in the solid they may represent either longitudinal or transverse waves. The velocities of these can differ. But the lattice vibrations may be described by phonons therefore we may ask how many phonons are in a crystal in a given frequency range. In the continuous limit this can be calculated in a 3D box of volume V (where the phonons represent standing waves) using the phonon density of states. The number of phonon states of kind s with velocity $v_s(k)$ available in the frequency range $[\nu, \nu + d\nu]$ then³

$$dn_s(\nu, \nu + d\nu) = g_s(\nu)d\nu = \frac{4\pi V \nu^2}{v_s^3(\nu)}d\nu$$

We must use separate formulas for the 2 identical transverse ($s \equiv t$) and 1 longitudinal ($s \equiv l$) phonon modes:

³Using (3.5.19)

$$dn(\nu, \nu + d\nu) = g(\nu)d\nu = g(k(\nu))\frac{dk}{d\nu}d\nu = \frac{V}{8\pi^3} 4\pi k^2 dk$$

and the formula of $k(\nu) = 2\pi/\lambda = 2\pi \cdot \nu/v_s(k)$ we can replace k with ν :

$$g(\nu)d\nu = \frac{V}{8\pi^3} 4\pi \left(\frac{2\pi\nu}{v(\nu)} \right)^2 \frac{2\pi}{v(\nu)}d\nu = \frac{4\pi V \nu^2}{v_s^3(\nu)}d\nu$$

$$g_t(\nu) = 2 \frac{4\pi V \nu^2}{v_t^3(\nu)} \quad (13.4.1)$$

$$g_l(\nu) = \frac{4\pi V \nu^2}{v_l^3(\nu)} \quad (13.4.2)$$

$$g_{total}(\nu) = 4\pi V \nu^2 \left(\frac{1}{v_l^2(\nu)} + \frac{2}{v_t^2(\nu)} \right) \quad (13.4.3)$$

The crystal contains N atoms which give $3N$ degree of freedom for atomic movement, therefore the total number of independent modes is also $3N$:

$$\int_0^\infty g(\nu) d\nu = 3N$$

13.5 Debye model

Unfortunately the functional form of $v_s(\nu)$ is quite complicated so this integral can only be calculated numerically. But if we assume after Debye that the velocity is constant ($v_s(\nu) \approx v_s(0) \equiv v_s$) from 0 up to a cut-off frequency ν_D (*Debye frequency*) and 0 above it the integral can easily be calculated, and we know that the result must be equal to $3N$:

$$3N = 4\pi V \left(\frac{1}{v_l^2} + \frac{2}{v_t^2} \right) \int_0^{\nu_D} \nu^2 d\nu = 4\pi V \left(\frac{1}{v_l^2} + \frac{2}{v_t^2} \right) \frac{\nu_D^3}{3}$$

From this we can substitute back the unknown factor containing the velocities into $g(\nu)$ and get a simpler expression:

$$\left(\frac{1}{v_l^2} + \frac{2}{v_t^2} \right) = \frac{9N}{4\pi V \nu_D^3} \quad (13.5.1)$$

$$g(\nu) = \frac{9N}{\nu_D^3} \nu^2 \quad (13.5.2)$$

The number of phonons of the same frequency ν is unlimited: *phonons are bosons*. The number of phonons in the system with energies in range $[\mathcal{E}(\nu), \mathcal{E}(\nu + d\nu)]$ is :

$$dn = g(\nu) f_{BE}(\nu) d\nu = \frac{9N}{\nu_D^3} \frac{\nu^2}{e^{h\nu/k_B T} - 1} d\nu$$

In 1D energy $\mathcal{E}(\nu) = h\nu(n + 1/2)$. Then the average thermal energy of the system is:

$$U = \int_0^{\nu_D} \mathcal{E}(\nu) dn + const = \int_0^{\nu_D} \mathcal{E}(\nu) g(\nu) f_{BE}(\nu) d\nu + const$$

$$U = \frac{9 N h}{\nu_D^3} \int_0^{\nu_D} \frac{\nu^3}{e^{h\nu/k_B T} - 1} d\nu + const \quad (13.5.3)$$

The specific heat at constant volume then

$$c_V = \left(\frac{\partial U}{\partial T} \right)_{V=const} = \frac{9 N h}{\nu_D^3} \frac{\partial}{\partial T} \left(\int_0^{\nu_D} \frac{\nu^3}{e^{h\nu/k_B T} - 1} d\nu \right)_{V=const} \quad (13.5.4)$$

The integration over ν and the derivation by T are interchangeable operations

$$c_V = \frac{9 N h}{\nu_D^3} \int_0^{\nu_D} \frac{\partial}{\partial T} \left(\frac{\nu^3}{e^{h\nu/k_B T} - 1} \right)_{V=const} d\nu = \frac{9 N h^2}{\nu_D^3 T^2} \int_0^{\nu_D} \frac{\nu^4 e^{h\nu/k_B T}}{(e^{h\nu/k_B T} - 1)^2} d\nu$$

For 1 mol atoms⁴ ($N = L_A$) c_v becomes the molar heat capacity C_V of lattice vibrations. Using the universal gas constant⁵ $R = k_B N_A$, and introducing the *Debye temperature* Θ_D with the formula $k_B \cdot \Theta_D \equiv h \nu_D$, and a new variable y by

$$y = \frac{h \nu}{k_B T} \Rightarrow \nu = \frac{k_B T y}{h} \text{ and } d\nu = \frac{k_B T}{h} dy$$

the molar heat capacity becomes

$$C_V = 9R \left(\frac{T}{\Theta_D} \right)^3 \int_0^{\Theta_D/T} \frac{y^4 e^y}{(e^y - 1)^2} dy \quad (13.5.5)$$

At low temperatures ($T \ll \Theta_D$) the upper limit of the integral is may be approximated with ∞ which makes its value constant and

$$\lim_{T \rightarrow 0} C_V \propto T^3 \quad T \ll \Theta_D \quad (13.5.6)$$

The Debye temperature can be calculated from low temperature specific heat measurements.

Material	Ag	Au	diamond	Cu	Ge	Na	Ni	Pt
Θ_D (K)	225	165	1860	339	366	159	456	229

Table 13.1: Debye temperature of some materials

At high temperatures ($T \gg \Theta_D$) both e^y and $e^y - 1$ may be approximated by the leading terms of their Taylor series with which the integrand

$$\frac{y^4(1 + y + \frac{y^2}{2} + \dots)}{[(1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \dots) - 1]^2} = \frac{y^4(1 + y + \frac{y^2}{2} + \dots)}{y^2(1 + \frac{y}{2} + \frac{y^2}{6} + \dots)^2} \approx \frac{y^4}{y^2} = y^2$$

⁴ $L_A = 6.022 \cdot 10^{23} [1/mol]$ is the Avogadro constant

⁵ $R = 8.3144621 \frac{J}{mol K}$.

And so

$$C_V = 9R \left(\frac{T}{\Theta_D} \right)^3 \int_0^{\Theta_D/T} y^2 dy = 9R \left(\frac{T}{\Theta_D} \right)^3 \frac{1}{3} \left(\frac{\Theta_D}{T} \right)^3$$

$$C_V = 3R \quad T \gg \Theta_D \quad (13.5.7)$$

Which is the well known *Doulong-Petit* law of classical physics⁶.

13.6 Specific heat of metals

In metals the conduction electrons also contribute to the internal energy. So the specific heat capacity of metals contains terms from both the electrons and lattice vibrations. Detailed discussion of the electron specific heat requires the quantum mechanical treatment of the electron system and it is in Section 14.3.1. Here just a very simple train of thought is given. The electrons excited by thermal vibration at $T > 0 K$ have energies in the $k_B T$ vicinity of \mathcal{E}_F , shown shaded in Fig. 14.5, so both the number of the excited electrons and their energy around \mathcal{E}_F is proportional to $k_B T$. The total excitation energy therefore is

$$U \sim (k_B T)^2$$

which gives the electron heat capacity proportional to the temperature (c.f. (14.3.13)):

$$C_V^{(el)} = \frac{\partial u}{\partial T} \sim T = \frac{\pi^2 R k_B}{2 \mathcal{E}_F} T$$

and the ratio of the electronic and lattice specific heat capacities at $T \ll \Theta_D$ is then

$$\frac{\text{electron specific heat}}{\text{lattice specific heat}} \propto \frac{T}{T^3} = T^{-2} \quad (13.6.1)$$

It follows that the electron specific heat will become important only at very low temperatures where the lattice specific heat decreases more rapidly.

⁶ According to the equipartition theorem of classical physics there is $1/2 k_B T$ energy available for every single degree of freedom. A single harmonic oscillator has a degree of freedom of 2 which results in $k_B T$ energy per oscillator. N atomic oscillators have $2 \cdot 3 N$ degrees of freedom, so the total energy is $3 k_B T N$. For 1 mol this results in an internal energy of $3 R T$ and molar heat of $3 R$, independent of the kind of the material. This prediction of classical physics is clearly wrong.

Chapter 14

Electrical properties

Important 14.0.1. *Electrical conductivity of different materials has the largest range among all physical quantities spanning about 25 orders of magnitude.*

14.1 Conductors and insulators. Band theory of solids

The behavior of the electrons in a solid – just like in molecules – is different from what they exhibit in an individual atom especially when the solid is a (periodic) crystal. Let us build up a crystal from N individual atoms! When the atoms are so far from each other that they may be considered independent electrons on the same orbit in every atom have the same energies, making the energy levels for the whole system of independent atoms degenerate. When the atoms are brought together to form a crystal they will interact and as a result the degeneration will be broken. The interaction is stronger for higher lying levels (greater extent of the wave functions).

It follows from the Pauli principle that a selected atomic state in all of the N atoms (with quantum numbers n, l, m) may hold $2N$ electrons. Therefore 1 degenerate atomic energy level will yield $2N$ possible non-degenerate levels in the crystal.

For a given n there may exist s, p, d , etc shells that create the bands s, p, d , etc. The smaller are the inter-atomic distances the wider are the bands. This may lead to band overlap.

The electric conductivity of a material is determined by its band structure at the equilibrium distances of the atoms. Lower lying shells have smaller overlap (the electrons are localized) and are usually completely filled therefore the corresponding bands will have no overlap with partially or totally empty other bands, therefore do not take part of electric conduction.

Bands formed from valence electrons are the important ones.

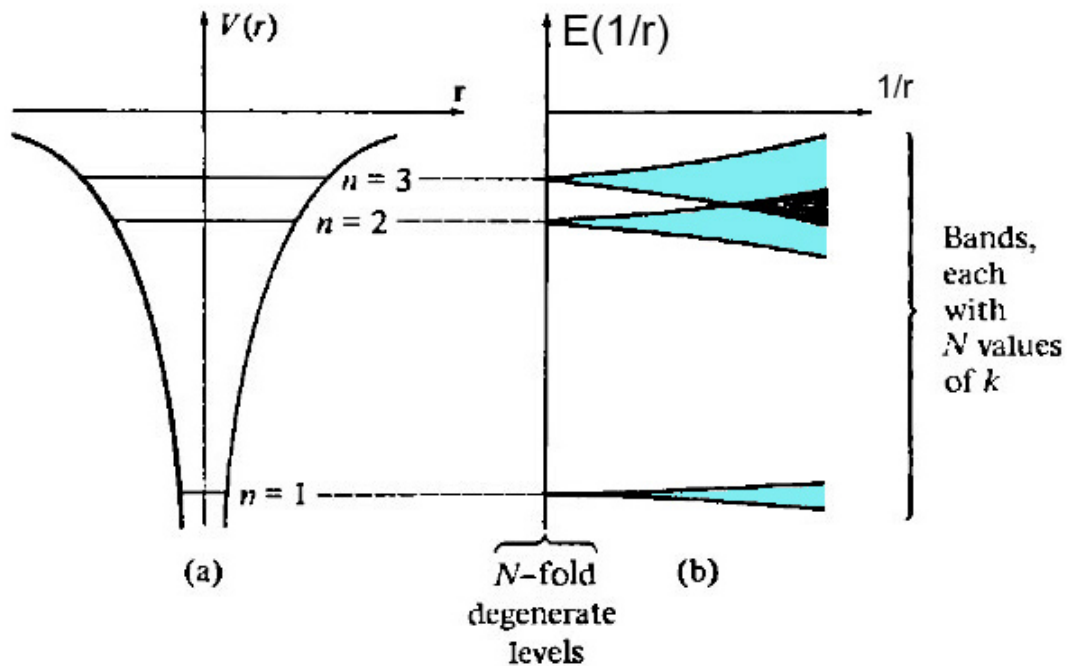


Figure 14.1: a) Schematic representation of non degenerate electronic levels in an atomic potential b) bands as a function of inter-atomic distance when N atoms are brought together. Notice that there may be forbidden energy regions between some of the bands, while other bands may overlap.

Metals

Metallic solids may form from elements having either an incompletely filled band or overlapping bands. Some examples:

Na ($Z=11$) - Metal

Electronic structure: the $1s^2 2s^2 2p^6$ levels are filled with $(2 + 2 + 6) = 10$ of the 11 electrons. The last one goes to level $3s$ which has space for 2 electrons. Consequently a half filled electron band is created from $3s$ orbitals of the N atoms, which is thus both valence and conduction band at the same time.

Even at room temperature ($T=300K$) a few electrons are excited above the Fermi level. An external \mathbf{E} field can add an additional energy $(1/2)m_e v_{drift}^2$ where v_{drift} is the constant average velocity created by the field) because there are empty levels in the neighborhood \Rightarrow good electric and thermal conduction. Level $3p$ overlaps with $3s$ therefore there are even more available levels $(1 + 6) N$ for the last electrons.

Mg (Z=12) - (Semi) Metal

Electronic structure: $1s^2 2s^2 2p^6 3s^2$ levels are filled with $(2 + 2 + 6 + 2) = 12$ electrons: no free levels in highest valence band. But $3p$ overlaps with $3s$ therefore there are 6N possible free levels for the electrons.

Fe (Z=26) - Transition metal

Electronic structure: $1s^2 2s^2 2p^6 3s^2 3p^6 3d^6 4s^2$ Levels $3s$, $4d$ and $4p$ overlap.

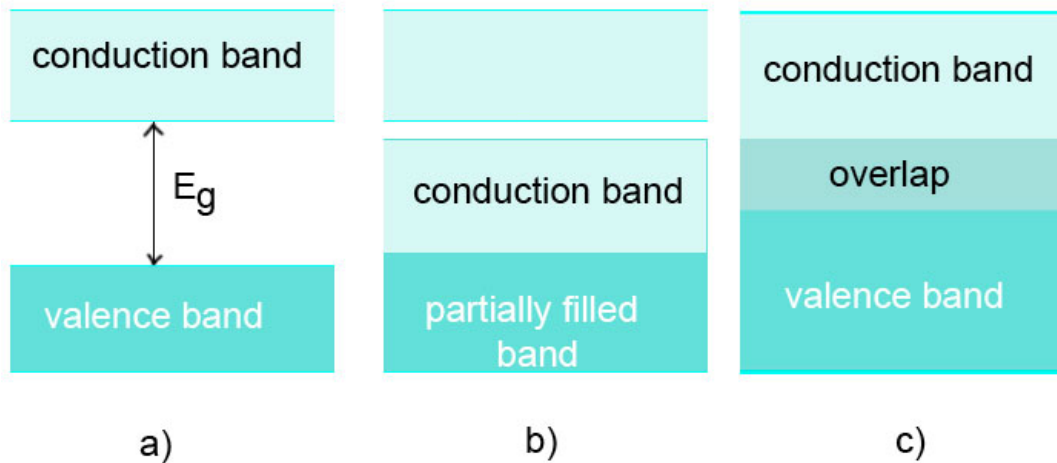


Figure 14.2: Schematic representation of band structures a) insulator (including semi-conductors) b) and c) metals (conductors)

Insulators

In insulators there is an *energy gap* between the completely filled valence band and the empty conduction band. The band width varies. Examples: Lead(II) selenide 0.27 eV, germanium: 0.7 eV, silicon 1.1 eV diamond: 5.5 eV. Because 1 eV corresponds to about $10^4 K$ the thermal energy at room temperatures $k_B T \approx 0.0258 \text{ eV}$ is too small to excite a significant number of electrons into the conduction band.

Important 14.1.1. Semiconductors are insulators with an energy gap around 1 eV or smaller.

When dealing with the electrical conductivity of crystals we find the idea of dispersion relations introduced for lattice vibrations very useful. $E(\mathbf{k})$ dispersion relations connecting the energy and momentum of the electrons may be measured and calculated for the electrons in crystals and these relations determine the band structure. On the schematic pictures in Fig. 14.2 no direction dependencies of the energy levels are displayed, but for the understanding of the electrical properties of crystals we must take them into considerations. We discuss this later in Chapter 16.

14.2 A classical physical model of conductivity in metals. The Drude model

Metals contain moveable electrons that can carry electricity. Although it is impossible to understand their behavior completely using only classical physics, the earliest theory with partial success was the Drude model, proposed in 1900 by Paul Drude, which is based on classical physics. The assumptions of this model are the following:

- In a metal the ion cores (including that of impurity ions) are at rest in lattice points. Lattice defects may also be present though.
- The ion cores are surrounded by unbound conduction electrons. There is no (electromagnetic) interaction between electrons (*independent electron approximation*).
- The only interaction between ion cores, impurities or lattice defects and electrons is collision.
- All collisions are instantaneous events that abruptly change the velocity of the electrons.
- Electrons reach thermal equilibrium only by collisions.
- The probability of a collision during a period of dt is $\sim dt/\tau$, where τ is called *collision time*, *mean free time* or *relaxation time*.

Conduction electrons satisfying these assumptions behave like particles of an ideal gas, therefore they are often referred to as *electron gas*.

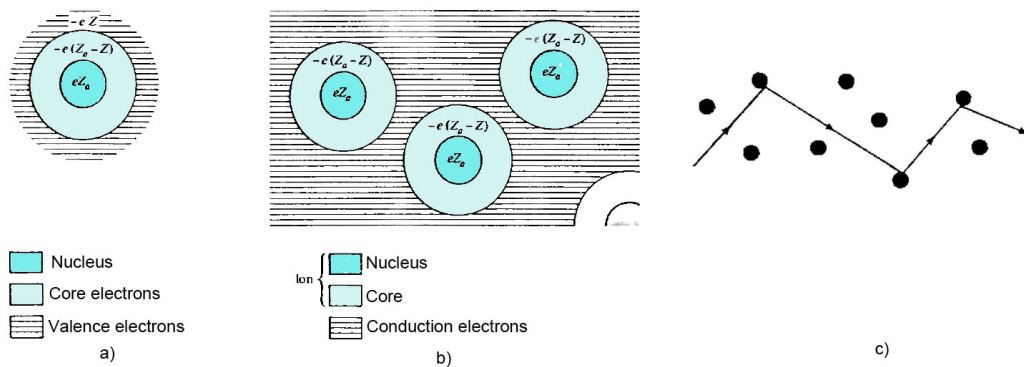


Figure 14.3: a) Schematic representation of an isolated atom (not to scale!) b) atoms in a metal keep the core electrons but the valence electrons form an electron gas c) trajectory of a single electron scattering off the ion cores

Electric conductivity (σ) of the electron gas

Without an external electric field electrons move in random directions with random velocities. The electrons collide with the ion cores after a random Δt time, but these collisions do not change the randomness of the velocities. There is no non-random component present, therefore the average (vectorial) velocity of electrons is 0.

In an external electric field \mathbf{E} all electrons will accelerate in the opposite direction of the field until they suffer a collision. Let us denote the average time between collisions with τ . During this period electron velocities will have an increasing *non-random component*. After time Δt elapsed electrons collide with (or scatter off) something and their velocity is again randomized, i.e. they lose all of the directional velocity component gained during acceleration. The time between collisions is still determined by the random velocity component as it is much larger than the directional one. It can be calculated if we know the velocity distribution. As the Maxwell-Boltzmann velocity distribution was the one known at that time Drude used it to obtain the value of Δt .

The maximum of the directional velocity component parallel with E gained during acceleration is $\mathbf{v}_{\text{accel}} = -e\mathbf{E}\Delta t/m_e$.

The average velocity of a mass point, originally at rest, under a constant acceleration during a time interval of Δt would be half of the maximum velocity $\mathbf{v}_{\text{aver},\Delta t} = \mathbf{v}_{\text{accel},\Delta t}/2$ it reaches. This is the value Drude used in his calculations. Let us introduce the average time τ between collisions with

$$\tau = \langle \Delta t \rangle$$

and express all quantities with it. Interestingly, when the exact classical statistical physical calculations are performed the average velocity expressed with τ will be equal to the maximum velocity $\mathbf{v}_{\text{accel},\tau}$ and not the half of it:

$$\mathbf{v}_{\text{drift}} = -\frac{e\mathbf{E}}{m_e} \tau \quad (14.2.1)$$

The reason for this is that Δt may be different for different collisions, therefore average velocities (which are half of the maximum velocity) for individual collisions will also differ from each other.

According to the differential Ohm's law:

$$\mathbf{j} = \sigma \mathbf{E},$$

where \mathbf{E} is the external electric field and σ is the conductivity. The current density \mathbf{j} in this formula can be expressed with the *drift velocity* $\mathbf{v}_{\text{drift}}$ which is the average ordered (non-random) component of the velocity of the electrons:

$$\mathbf{j} = -ne\mathbf{v}_{\text{drift}}.$$

So

$$\mathbf{j} = \left(\frac{ne^2\tau}{m_e} \right) \mathbf{E}$$

from this

$$\sigma = \frac{ne^2\tau}{m_e} \quad (14.2.2)$$

As a result of collisions the ordered part of the electron velocities will be constant and proportional with the field strength:

$$\mathbf{v}_{drift} = \mu \mathbf{E} \quad (14.2.3)$$

where the constant of proportionality is the *mobility* μ of the electrons

$$\mu = \left(\frac{e\tau}{m_e} \right) \quad (14.2.4)$$

Example 14.1. *Aluminum has three valence electrons per atom, an atomic weight of 0.02698 kg/mol, a density of 2700 kg/m³, and a conductivity of 3.54 10⁷ S/m. Calculate the electron mobility in aluminum. Assume that all three valence electrons of each atoms are free.* **Solution** The number of aluminum atoms per m³ is

$$\begin{aligned} n_a &= 6.0210^{23} \text{ atoms/mol} \cdot 1/0.02698 \text{ mol/kg} \cdot 2700 \text{ kg/m}^3 \\ &= 6.024 \cdot 10^{28} \text{ atoms/m}^3 \end{aligned}$$

Thus the electron density in aluminum is

$$n = 3 \cdot 6.024 \cdot 10^{28} \text{ atoms/m}^3 = 1.807 \cdot 10^{29} \text{ electron/m}^3$$

From (14.2.4)

$$\mu = \frac{\sigma}{ne} = \frac{3.54 \cdot 10^7}{1.807 \cdot 10^{29} \cdot 1.6022 \cdot 10^{-19}} = 1.22 \cdot 10^{-3} \text{ m/s}$$

Calculating τ theoretically is quite complex. For this reason just empirical approaches are given in the following.

One way to determine the value of τ is from (14.2.2), where the density n of conduction electrons is

$$n = L_A \frac{Z\rho_m}{A} \quad (14.2.5)$$

Here $L_A = 6.022 \cdot 10^{23} \text{ 1/mol}$, Z is the number of conduction electrons of one atom, ρ_m is the mass density (g/cm^3), and A is the atomic mass (g/mol). In practice usually the ρ resistance is measured instead of σ ($\rho = 1/\sigma$).

Example 14.2. What is the value of τ in silver at $t = 0^\circ\text{C}$ if the measured resistivity is $1.51 \cdot 10^{-8} \Omega\text{m}$? **Solution** From (14.2.2), (14.2.5) using the definition of ρ

$$\tau = \frac{m_e}{\rho(T) n e^2} = \frac{m_e A}{L_A Z \rho_m e^2 \rho}$$

Silver has a single $5s^1$ electron so $Z = 1$ and the mass density is $\rho_m = 10.49 \text{ g/cm}^3 = 1.049 \cdot 10^4 \text{ kg/m}^3$, $A = 107.8682 \text{ g/mol} = 0.1078682 \text{ kg/mol}$ and $\rho(273\text{K}) = 1.51 \cdot 10^{-8} \Omega\text{m}$. After substitution

$$\underline{\underline{\tau = 4.013 \cdot 10^{-14} \text{ s}}}$$

Another way to determine τ is from the *mean free path* \bar{l} .

$$\tau = \frac{\bar{l}}{\langle v \rangle}$$

For $\langle v \rangle$ Drude used the v_{th} average thermal velocity from the Maxwell-Boltzmann distribution:

$$\langle v \rangle = v_{th} = \sqrt{\frac{3k_B T}{m_e}}$$

However this will lead to an incorrect temperature dependence. From (16.1.1)

$$\tau \sim \frac{1}{\sqrt{T}} \text{ therefore } \sigma \sim \frac{1}{\sqrt{T}} \Rightarrow \rho \sim \sqrt{T}$$

This is a major failure of the Drude model, because the resistivity of metals increases *linearly* with increasing temperature.

At that time it was regarded the greatest success of the Drude model that it explained the empirical Wiedemann-Franz law (1853), which states that

$$\frac{\kappa}{\sigma} \sim T \quad \text{for all metals with about the same constant} \quad (14.2.6)$$

where κ is the thermal conductivity in Fourier's law:

$$j_Q = -\kappa \frac{\partial T}{\partial x} \quad (\text{in 1D}) \quad (14.2.7)$$

$$\mathbf{j}_Q = -\kappa \nabla T \quad (\text{in 3D}) \quad (14.2.8)$$

Drude assumed the bulk of the thermal current is transferred by conduction electrons and the classical ideal gas laws are applicable for the electron gas

$$\frac{\kappa}{\sigma} = \frac{3}{2} \left(\frac{k_B}{e} \right)^2 T \quad (14.2.9)$$

Unfortunately the calculated result is about half of the measured ratio¹.

The electronic specific heat may also be calculated from the Drude model but it also gives the incorrect result

$$C_V = \frac{3}{2}n\bar{l}$$

Yet another failure of the model is that it cannot explain the Hall effect of divalent metals² Mg, Cd and Be, in which the electric charge carriers were found to be positively and not negatively charged.

The three failures of the classical Drude model (the temperature dependence of ρ , the erroneous result of the value in the Wiedemann-Franz law and the electronic specific heat) show that quantum mechanical treatment of the conduction phenomena is required.

In spite of these failures the Drude model was remarkably successful for a first approximation. Some of its failures were corrected by Sommerfeld in his free-electron model of metals.

14.3 Free electron model of metals, the Sommerfeld model

The first step toward a quantum mechanical description of conduction by German physicist Arnold Sommerfeld in 1933 was based on the Drude model supplemented with results from quantum mechanics. These were Sommerfeld's additions to the Drude model:

1. conduction electrons do not interact with the ion cores, so the metal may be represented by a potential box
2. electrons do not interact with each other (independent electron model)
3. even though there are no interactions between electrons still no 2 electrons can be in the same quantum state when we include the spin as a quantum number (Pauli principle)

Because of assumption 2. we can solve the Schrödinger equation for all electrons separately, and according to the other assumptions (V is the volume of the box):

$$\psi = \frac{1}{\sqrt{V}}(\sin k_x x \cdot \sin k_y y \cdot \sin k_z z) \quad (14.3.1)$$

$$\mathcal{E}_{tot} = \frac{p^2}{2m_e} = \frac{\hbar^2 k^2}{2m_e} \quad (14.3.2)$$

¹But only if the calculation is correct. Drude erroneously calculated v_{avg} as half of the correct value and consequently got twice the result of (14.2.9) thus getting the correct result.

²See Section 15.5 in Section 15.5 for details.

Consider a cube shaped potential box with sides L .

Recapping the quantum mechanical problem (See Section 3.5.3) using assumptions (1) and (14.3.2) we find in 1D

$$\psi(x) = \frac{1}{L} \sin kx$$

and

$$\begin{aligned} \psi(0) = \psi(L) &= 0 & \Rightarrow \\ \sin kL &= 0 & \Rightarrow \\ kL &= n\pi & n = 1, 2, \dots \\ k &= \frac{\pi}{L} n & n = 1, 2, \dots \end{aligned}$$

I.e. possible values of k are discrete³.

If we imagine this box is a simple cubic crystal with primitive vectors of length a containing N atoms in any direction i.e. $L = Na$, then the difference between neighboring k -s is very small:

$$\Delta k = \frac{\pi}{Na} \sim 10^{-13} - 10^{-15} \quad 1/m$$

so we may consider k as a (quasi) continuous quantity⁴.

The energy of the electrons as a function of k (the analogue of the dispersion relation for lattice vibrations) is

$$\mathcal{E}_n(k) = \frac{\hbar^2 k^2}{2m_e} = \frac{\hbar^2 \pi^2}{2m_e N^2 a^2} n^2 \quad \text{where } n = 1, 2, \dots \quad (14.3.3)$$

The number of possible k values is infinite in this model. However we will see later (c.f. section 15.1.3. *Bloch functions*) that in periodic potentials the number of k states are limited by N , the number of atoms. Any of these energy states can be occupied by 2 electrons with opposite spins. If each atom contribute s conduction electrons to the crystal then at $T = 0 \text{ K}$ all possible energy states below

$$\mathcal{E}_F (= \mathcal{E}_{max,occupied}) = \frac{\hbar^2 \pi^2}{2m_e a^2} \frac{s}{2} \quad (14.3.4)$$

³We only count positive n 's because $\sin(-\frac{n\pi}{L}) = -\sin(\frac{n\pi}{L})$ and $\sin(n\pi/L)$ are equivalent wave functions as they only differ in a phase factor whose absolute value is 1

⁴The same calculation using the wavelength of the electron instead of its wave vector:

$$\lambda_{electron} \quad \ni \quad n \frac{\lambda_{electron}}{2} = L \quad \text{where } n = 1, 2, \dots$$

from here

$$k = \frac{2\pi}{\lambda_{electron}} = \frac{\pi n}{L} = \frac{\pi}{Na} n \quad \text{where } n = 1, 2, \dots$$

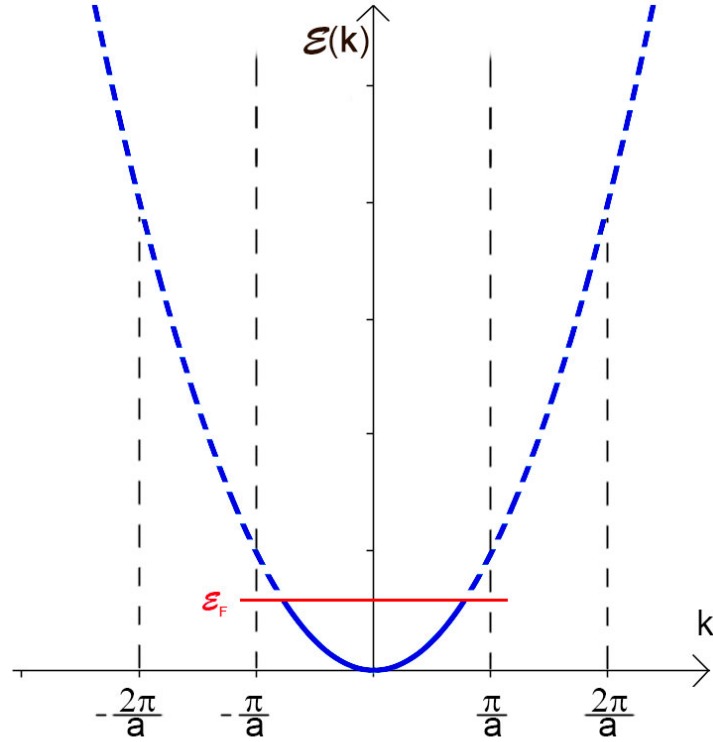


Figure 14.4: The energy of the electrons as a function of k (Energy dispersion relation) in the Sommerfeld model, when all atoms contribute 2 conduction electrons with opposing spins. Dashed line denotes levels unoccupied in the ground state.

will be occupied and all levels above it will be empty. \mathcal{E}_F is called the *Fermi energy* of the system (see also section 9.3). We denote the corresponding k value with k_F and call it the *Fermi wave vector*.

Important 14.3.1. *When one additional electron is added to the number electrons already in the metal the energy of the system will increase with \mathcal{E}_F . Therefore the Fermi energy is the μ_e chemical potential of the electrons.*

Every possible k point occupy the same $\Delta V_k = (2\pi/L)^3$ volume of 3 D k -space, therefore a region of 3 dimensional k -space of a volume of Ω will contain

$$\frac{\Omega}{\Delta V_k} = \frac{\Omega V}{(2\pi)^3} \quad (14.3.5)$$

number of allowed \mathbf{k} points, or equivalently the number of allowed k -values per unit volume of k -space (a.k.a. *density of levels*) is

$$\rho_{e,levels} = \frac{V}{(2\pi)^3} \quad (14.3.6)$$

Since the energy of a one-electron level is directly proportional to the absolute square of its wave vector the direction of \mathbf{k} does not matter. Furthermore N is very large so the Ω volume of k -space occupied may be considered a sphere (*Fermi sphere*). Therefore the number of allowed \mathbf{k} states in the system from (14.3.5) in 3D

$$N_{states} = \rho_{e,levels} \Omega = \frac{4}{3}\pi k_F^3 \left(\frac{V}{8\pi^3} \right) = \frac{k_F^3}{6\pi^2} V \quad (14.3.7)$$

And if every atom gives 2 electrons then the number of electrons ($N_e = N \cdot s$) occupying these states is twice this number:

$$N_e = 2 \frac{k_F^3}{6\pi^2} V \quad \Rightarrow \quad n(\equiv \frac{N_e}{V}) = \frac{k_F^3}{3\pi^2} \quad (14.3.8)$$

The length of the Fermi wave vector is $2\pi/a$ and the allowed k values will be in the interval

$$k \in \left[-\frac{\pi}{a}, \frac{\pi}{a} \right]$$

The Fermi energy from (14.3.4)

$$\mathcal{E}_F = \frac{\hbar^2 k_F^2}{2m_e} \quad (14.3.9)$$

Other definitions:

$$\begin{aligned} p_F &\equiv \hbar k_F && \text{Fermi momentum} \\ v_F &\equiv p_F/m_e && \text{Fermi velocity} \end{aligned}$$

Using (14.3.8) we can estimate the values of these quantities and the results are interesting:

- k_F yields electronic wavelengths corresponding to the inter-atomic distances ($\sim 0.1nm$)
- v_F yields velocities at 0K of the order of 0.1 c! (In classical ideal gases the velocity is 0 at 0K, and even at room temperatures classical particles with the electron mass will only have velocities of about 10^5 m/s $\sim 0.001c$)
- \mathcal{E}_F is about the same as the typical atomic binding energy.

Metal	Li	Na	K	Rb	Cs	Cu	Ag	Au	Mg	Al
$\mathcal{E}_F(eV)$	4.7	3.1	2.1	1.8	1.5	4.1	5.5	5.5	7.3	11.9

Table 14.1: Fermi energies of some metals

Example 14.3. *What is the quasi-free electron density in copper? Calculate the Fermi velocity and momentum too* **Solution** From Table 14.1 and formulas (14.3.8) and (14.3.4)

$$n = \frac{(2m_e \mathcal{E}_F)^{3/2}}{2\pi^2 \hbar^3} = 5.655 \cdot 10^{28} \frac{\text{electron}}{m^3}$$

$$v_F = 1.2 \cdot 10^6 m/s = 0.004 c \quad k_F = 1.0 \cdot 10^{11} 1/m$$

14.3.1 Specific heat of metals

We can express the number of electrons up to the Fermi level using the electron energy instead of the wave vector. The number of possible states for electrons in the energy range $d\mathcal{E}$ around \mathcal{E} is

$$dn_{\mathcal{E},\text{possible}} = g(\mathcal{E})d\mathcal{E}$$

where $g(\mathcal{E})$ is the density of states for a potential box (see (3.5.20)):

$$g(\mathcal{E}) = \frac{8\pi\sqrt{2m_e^3}}{h^3}\sqrt{\mathcal{E}} \quad (14.3.10)$$

The number of states available for electrons up to E is⁵

$$n = \int_0^{\mathcal{E}} g(\mathcal{E})d\mathcal{E} = \frac{16\pi\sqrt{2m_e^3}}{3h^3}\mathcal{E}^{3/2}$$

substituting \mathcal{E} with \mathcal{E}_F and n with the total number of conduction electrons n_{tot} in (14.3.10) we get the value of the density of state function at \mathcal{E}_F :

$$g(\mathcal{E}_F) = \frac{3n_{tot}}{2\mathcal{E}_F} \quad (14.3.11)$$

The number of electrons on these possible states depends on the *probability* that a state is occupied. This probability is determined by the $f_{FD}(\mathcal{E})$ *Fermi-Dirac distribution function*:

$$f_{FD}(\mathcal{E}) = \frac{1}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} \quad (14.3.12)$$

⁵See how summation of quantities depending on the quasi continuous \mathbf{k} vectors can be turned to integration in Appendix 23.5.

So the number of electrons in the energy range $d\mathcal{E}$ around \mathcal{E} is given by

$$dn = f_{FD}(\mathcal{E})g(\mathcal{E})d\mathcal{E}$$

As a consequence of the Pauli principle only electrons with energies near the Fermi energy \mathcal{E}_F can be excited. The electronic part of the internal energy is calculated the usual way, which includes the electronic density of states and the Fermi–Dirac distribution function:

$$U_e = \int \mathcal{E} \cdot g(\mathcal{E}) \cdot f_{FD}(\mathcal{E}) d\mathcal{E} = \int \mathcal{E} \cdot g(\mathcal{E}) \cdot \frac{1}{e^{(\mathcal{E}-E_F)/k_B T} + 1} d\mathcal{E}$$

The total integral may be written as a sum of 3 integrals: an integral from 0 to $(\mathcal{E}_F - k_B T)$, an integral between $(\mathcal{E}_F - k_B T)$ and $(\mathcal{E}_F + k_B T)$ and an integral from $(\mathcal{E}_F + k_B T)$ to ∞ . The first one is approximately constant ($k_B T \ll E_F$), the second one may be approximated⁶ by using the value of $g(\mathcal{E})$ at \mathcal{E}_F and the last one is negligible (≈ 0):

$$U_e \approx U_0 + \frac{\pi^2}{6} (k_B T)^2 g(\mathcal{E}_F) + 0$$

From this:

$$C_V^{(el)} = \left(\frac{\partial U}{\partial T} \right)_V = \frac{\pi^2}{3} k_B^2 g(\mathcal{E}_F) T = \frac{\pi^2 R}{2} \frac{k_B T}{\mathcal{E}_F} \quad (14.3.13)$$

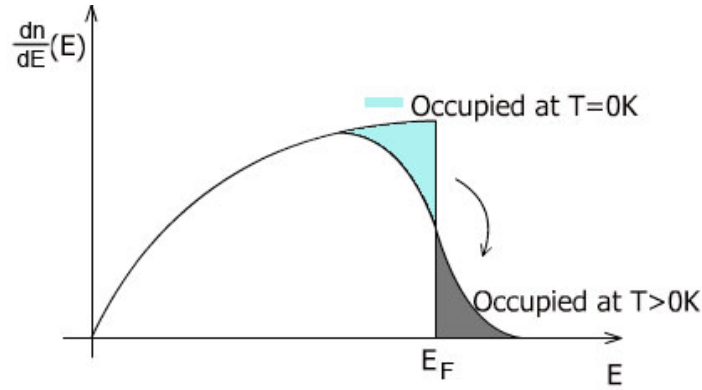


Figure 14.5: Energy distribution of electrons in the Sommerfeld model

⁶This integral can be approximated because f_{FD} only changes in a range of magnitude $2 k_B T$ around \mathcal{E}_F and in a smooth way.

14.3.2 Conductivity

The number of one electron levels in the velocity interval d^3v around velocity \mathbf{v} is

$$dn_{levels}(\mathbf{v}) = f(\mathbf{v})d^3v$$

where $f(\mathbf{v})$ is the *velocity distribution function* calculated by using $\mathbf{v} = \hbar\mathbf{k}/m_e$.

On the other hand $dn_{levels}(\mathbf{v})$ is equal to the number of levels in an interval d^3k around $\mathbf{k}(\mathbf{v})$

$$dn_{levels}(\mathbf{k}) = 2 \left(\frac{V}{8\pi^3} d^3k \right)$$

The number of electrons in these states then is given by⁷

$$dn = f_{FD}(\mathcal{E})dn_{levels}$$

Substituting $d^3k = (m_e/\hbar)^3 d^3v$ dn

$$\begin{aligned} dn(\mathbf{v}) &= dn(\mathbf{k}) \quad \text{where } \mathbf{v} = \hbar\mathbf{k} \\ dn(\mathbf{k}) &= f_{FD}(\mathcal{E}) \cdot 2 \left(\frac{V}{8\pi^3} \right) d^3k = \\ &= \frac{1}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} \left(\frac{V}{4\pi^3} \right) d^3v = \\ &= \left(\frac{Vm_e^3}{4\hbar^3\pi^3} \right) \frac{1}{e^{(\frac{1}{2}m\mathbf{v}^2-\mathcal{E}_F)/k_B T} + 1} d^3v \end{aligned}$$

i.e.

$$f(\mathbf{k}) = \left(\frac{Vm_e^3}{4\hbar^3\pi^3} \right) \frac{1}{e^{(\frac{1}{2}m\mathbf{v}^2-\mathcal{E}_F)/k_B T} + 1} \quad (14.3.14)$$

Without an external electric field the electron distribution in k -space has spherical symmetry and the average velocity is 0. When an external $\mathbf{E} = -\text{grad } \varphi(\mathbf{r})$ electric field is turned on it modifies this distribution by modifying the potential with $-e\varphi(\mathbf{r})$. This causes the displacement of the *Fermi sphere* in the (opposite) direction of the \mathbf{E} field. The velocity would increase indefinitely unless some mechanism prohibits it. Sommerfeld thought that the scattering mechanism of the Drude model is responsible for this.

Because of the exclusion principle only electrons with energies near to \mathcal{E}_F can be scattered in collisions (because there are no free levels for electrons with energies lower than about $\mathcal{E}_F - k_B T$), therefore their thermal (average) velocities will be equal to the v_F Fermi velocity. The relaxation time - mean free path relation becomes:

$$\tau = \frac{\bar{\ell}}{v_F} \quad (14.3.15)$$

⁷For brevity we will omit the explicit use of \mathbf{k} dependence, unless it is important to display.

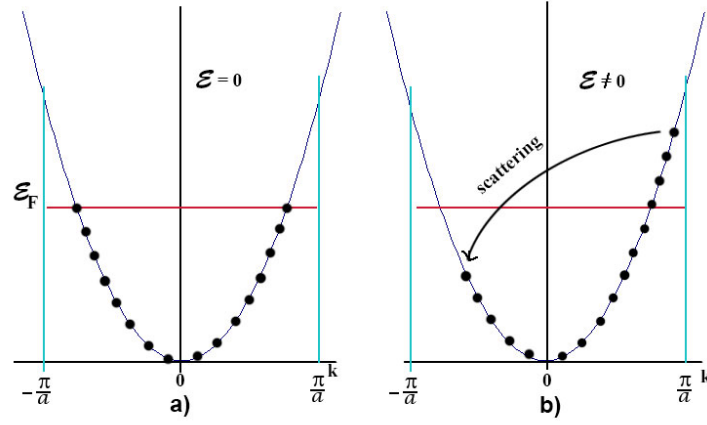


Figure 14.6: Conceptual drawing of conduction in metals according to the Sommerfeld model in 1 D. a) w.o. external field b) with an external \mathbf{E} field.

Substituting this into the conductivity formula (14.2.2) of the Drude model:

$$\sigma = \frac{ne^2\tau}{m_e} = \frac{ne^2}{m_e} \frac{\bar{\ell}}{v_F} \quad (14.3.16)$$

In contrast with the Drude model the v_F thermal velocity is independent of the temperature and $\bar{\ell}$ is inversely proportional with it, therefore

$$\sigma \sim \bar{\ell} \sim T^{-1} \quad \Rightarrow \quad \rho \sim T \quad (14.3.17)$$

i.e. the resistivity is proportional to the temperature which conforms to the experimentally observed behavior. Unfortunately v_F is about 10-20 times larger than v_{th} in the Drude model which leads to mean free paths in the order of $10^{-7} m$, much larger than the expected ones which are in the order of magnitude of interatomic distances ($\sim 10^{-9} m$).

The Sommerfeld free-electron model successfully explained other phenomena too for which the Drude-model either failed or did not accounted for. To name a few: the specific heat of metals, the work function, the thermionic emission and the contact potential. But it has its own failures. Some of them are:

- It did not explain why some materials are insulators? (e.g. carbon is an insulator in the form of diamond, while a conductor in the form of graphite.)
- It did not explain the temperature, relaxation time and magnetic field dependence of the Hall effect⁸ as it predicts a constant $R_H = -1/nec$ value.

⁸Section 15.5

- It did not explain the *magnetoresistance*, according to which the resistivity of a wire perpendicular to a homogeneous magnetic field depends on the field strength.

14.4 Work function, thermionic emission and contact potential

14.4.1 Work function

The work function is the minimum energy that must be given to an electron to leave the crystal. The method of energy transfer may be supplied by photons (*photoelectric effect*) or thermal vibrations (*thermionic emission*).

The real crystal potential near the surface can be approximated with the one in Fig. 14.7. Inside the crystal the potential is periodic (a sum of the Coulomb potentials of the periodically arranged ion cores) and the superposition of the potential of the neighboring atoms creates a potential lower than that in the vacuum. At the surface the atoms do not have neighbors at one side so the Coulomb potential of the surface atoms in the vacuum will not diminish.

Similarly while inside the crystal the electron density is periodic, at the surface this periodicity is broken. A few cells near the surface will have an electron deficiency (which gives rise to a positive *surface charge*), while some of the electrons will be outside, near the surface (as a negative surface charge) which will produce a *double layer* at the surface. Only electrons at or near to the Fermi energy can leave the crystal. For an electron to become free an external source must supply an energy to move it from the Fermi energy (measured from the vacuum level so it is now negative) to the vacuum level plus the work needed to overcome the W_s potential of the double layer. That is the formula for the work function is :

$$W = W_s - \mathcal{E}_F$$

The work functions on different faces of a crystal may be different, in which case there will be a potential difference between the different faces. The whole crystal still remain neutral, because the sum of the microscopic surface charges over all the faces will cancel.

14.4.2 Thermionic emission

At very high temperatures electrons may get enough energy to leave the metal. The current density of these electrons can be calculated by generalizing the well known differential formula for the current density:

$$\mathbf{j} = -en < \mathbf{v} >$$

For cases where $\mathbf{v} = \mathbf{v}(\mathbf{k})$

$$\langle \mathbf{v}(\mathbf{k}) \rangle = \frac{1}{n} \frac{1}{V} \sum_{\mathbf{k}} f(\mathbf{k}) \mathbf{v}(\mathbf{k})$$

$$\mathbf{j} = -e \frac{1}{V} \sum_{\mathbf{k}} f(\mathbf{k}) \mathbf{v}(\mathbf{k})$$

This formula may be re-written as an integral using (23.5.1) and taking into account the 2 spin orientations of electrons:

$$\mathbf{j} = - \left(\frac{e}{4\pi^3} \right) \int_{\mathbf{k}} f(\mathbf{k}) \mathbf{v}(\mathbf{k}) d^3\mathbf{k}$$

Now suppose the metal surface is perpendicular to the x -axis and there is vacuum outside the crystal (if $x \geq 0$)

$$\mathbf{j}_x = -e \left(\frac{1}{4\pi^3} \right) \int_{\mathbf{k}} f(\mathbf{k}) v_x(\mathbf{k}) d^3\mathbf{k}$$

If the work function is W then electrons in the vacuum have a total energy of

$$\mathcal{E}_{tot} = \mathcal{E}_F + W + \frac{1}{2} m_e v_x^2(\mathbf{k}) = \mathcal{E}_F + W + \frac{\hbar^2 k^2}{2m_e}$$

substituting this into the formula of $f(\mathbf{k})$

$$f(\mathbf{k}) = \frac{1}{e^{(\hbar^2 k^2 / 2m_e + W) / k_B T} + 1}$$

W typically is a few eV, $W/k_B T \sim 10^4$, therefore

$$f(\mathbf{k}) \simeq e^{-(\hbar^2 k^2 / 2m_e + W) / k_B T}$$

$$\mathbf{j}_x = -e \left(\frac{1}{4\pi^3} \right) \int_{\mathbf{k} x > 0} v_x e^{-(\hbar^2 k^2 / 2m_e + W) / k_B T} d^3\mathbf{k}$$

$$\mathbf{j}_x = -e \left(\frac{1}{4\pi^3} \right) e^{-W/k_B T} \int_{\mathbf{k} x > 0} \frac{\hbar k_x}{m_e} e^{-\hbar^2 k^2 / 2m_e k_B T} d^3\mathbf{k}$$

The current per unit area emitted by the surface is given by:

$$\mathbf{j}_x = -\frac{e m_e}{2\pi^2 \hbar^3} (k_B T)^2 e^{-W/k_B T} = -\frac{4\pi m_e e}{h^3} (k_B T)^2 e^{-W/k_B T} \quad (14.4.1)$$

This is the *Richardson-Dushman* equation that gives the temperature dependence of thermionic emission.

14.4.3 Contact potential

Suppose two metals with different Fermi energies are contacted. Where the metals touch electrons may move freely from one metal to the other and they will, because the Fermi energy (chemical potential) of the electrons at the two sides differ. When equilibrium is reached \mathcal{E}_F will be the same for both metals. As a consequence there will be a net flow of charge from one metal to the other until both metals becomes charged with a net negative and positive charge respectively.

The potential difference that arise from this process is called the *contact potential* and can be calculated as the difference of the Work functions of the two metals:

$$-e(\varphi - \varphi') = W - W'$$

Measurement of the Contact Potentials

The contact potential cannot be measured by simple closed electronic circuit of the two metals and a galvanometer, because in the circuit the sum of the contact potentials (between metal A and B, between metal A and the galvanometer, between metal B and the galvanometer) must be 0 otherwise a perpetuum mobile could be created.

However there exists a simple method invented by Kelvin to measure contact potentials. Suppose we prepare a plane surface on both metals then contact them together in a closed circuit with an electrometer and a variable potential bias. As the Fermi levels in all metals will become the same contact potentials will form at every contacts. Now move the two metals apart so that the plane faces form a parallel capacitor. The potential between the two faces will remain the contact potential but the capacitance will vary depending on the distance of the faces. Therefore the contact potential can be calculated by either keeping the bias constant and measuring the current that flows while the faces are being moved, or by adjusting the bias so that no current flows, in which case the contact potential will be equal to the bias which just cancels it.

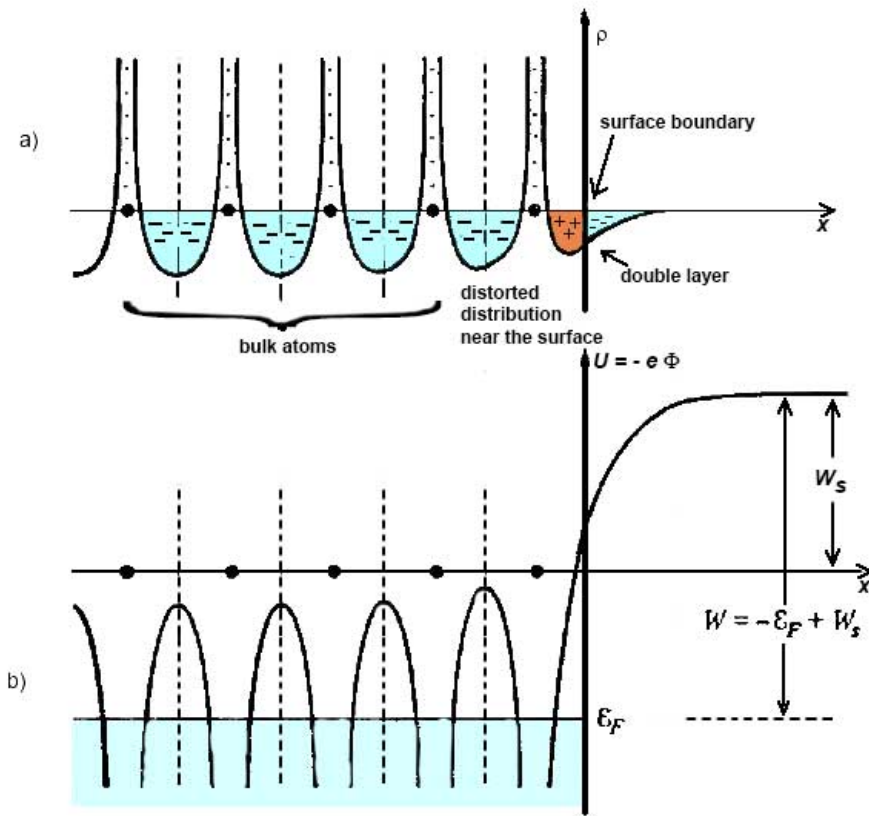


Figure 14.7: a) density of electrons in the crystal ($\sim |\psi|^2$)
The electron distribution is distorted near the surface relative to the bulk. b) crystal potential, Fermi energy and work function (W). The crystal potential is distorted near the surface. Closed circles are the equilibrium ion sites. W_s is the potential energy of the double layer.

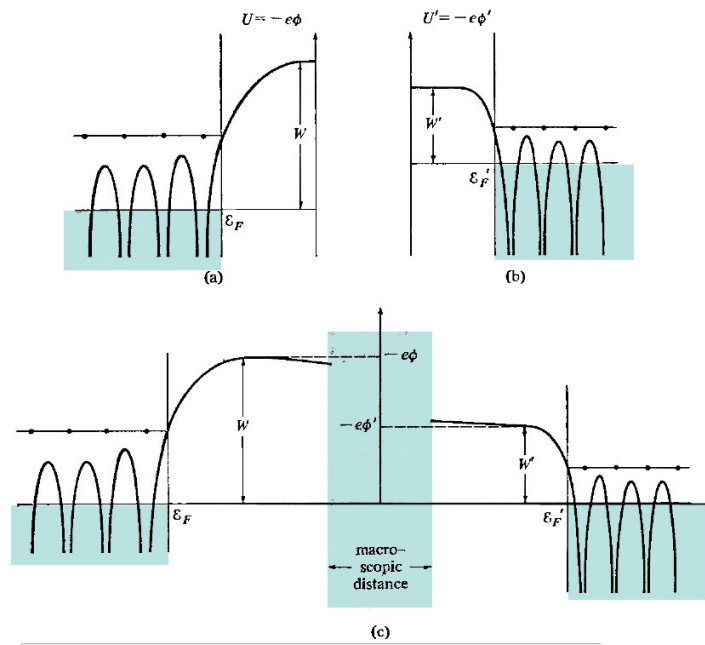


Figure 14.8: a) and b) two metals with different \mathcal{E}_F c) in equilibrium \mathcal{E}_F is the same but the surface potentials differ, i.e. a contact potential is created.

Chapter 15

Electrons in conductors

15.1 Quantum mechanics of electrons in periodic lattices. Adiabatic principle. Brillouin-zone. Bloch functions

We concluded in Section 14.3 that to correctly describe the conductivity we must use correct quantum mechanical calculations. This means the solution of the Schrödinger equation of the whole crystal. The Hamiltonian of the system can be written as a sum of the Hamiltonians for the electrons, the ions and the interaction between electron and ions:

$$H = H_{ion\ cores} + H_{electrons} + H_{ion-electron}$$

For a system of N ions and K electrons these are

$$\begin{aligned} H_{ion\ cores} &= \sum_{j=1}^N \frac{\mathbf{p}_j^2}{2M_{ion}} + V_{ion}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) \\ H_{electrons} &= \sum_{j=1}^K \frac{\mathbf{p}_j^2}{2m_e} + V_{electron}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K) \\ H_{ion-electron} &= V_{i-e}(\mathbf{R}_1, \dots, \mathbf{R}_N, \mathbf{r}_1, \dots, \mathbf{r}_K) \end{aligned}$$

Here $\mathbf{R}_1, \dots, \mathbf{R}_N$ are not lattice vectors, just the positions of the ions and $\mathbf{r}_1, \dots, \mathbf{r}_K$ are the positions of the electrons.

This will give us an unsolvable system of roughly 10^{24} coupled coordinates. So we must simplify it a bit.

Because the crystal is neutral its electrostatic potential energy is 0. We can select the potential between the ions to be so that when ions are at their equilibrium positions

$\mathbf{R}_j^{(0)}$ ($j = 1, 2, \dots, N$) the total potential energy will be 0.

$$V_{ion}(\mathbf{R}_1^{(0)}, \dots, \mathbf{R}_N^{(0)}) = 0$$

15.1.1 The Adiabatic Principle

The mass of the ion cores is about 2000-20000 times larger than the electron mass, therefore the velocities of the electrons will be much higher than the velocity of any ion. At every given moment the ions will only feel an average field due to the electrons, while for the electrons the lattice will be almost at rest at all times. The lattice vibrations being much slower than the motion of the electrons will manifest themselves in that ions are not in their equilibrium position. The electrons will follow the movement of the ions *adiabatically*.

Therefore we may be able to study the motion of ions and electrons separately. This is called the *adiabatic principle*.

Each and every electron feels the slowly varying potential of the ion cores and the (almost) instantaneous field of other electrons. Taking the positions of the ions fixed in their equilibrium position¹ the equation for the system of electrons will only contain $H_{electrons}$ and $H_{ion-electron}$.

$H_{ion-electron}$ only depends on the interaction of an electron with all of the ions, so each electron feels this interaction separately, i.e. V_{i-e} is a sum of one-electron potentials:

$$V_{i-e}(\mathbf{R}_1, \dots, \mathbf{R}_N, \mathbf{r}_1, \dots, \mathbf{r}_K) = \sum_{j=1}^K V(\mathbf{r}_j)$$

As the consequence of the translational symmetry of the crystal the potential will be the same at equivalent positions in all primitive cells, so

$$V(\mathbf{r}_j + \mathbf{R}) = V(\mathbf{r}_j)$$

where \mathbf{R} is a lattice vector.

We should like to write $V_{electron}$ as a sum of one-particle potentials, because then we would need only to solve independent Schrödinger equations for each electrons. If this is possible then our previous use of the picture of independent, non interacting electrons, i.e. the existence of an electron gas is justified.

15.1.2 Hartree-Fock method

This problem -at least in principle - can be solved with the *Hartree-Fock* („self-consistent”) method:

¹A more exact calculation would be unnecessarily complicated and will not give us any advantage so we omit it. We will only consider the movement of the ion cores when discussing lattice vibrations.

1. solve the one-electron Schrödinger equations when the electron moves in the periodic potential field of the ion cores, then
2. calculate the charge density using the square of the absolute value of these one-electron wave functions
3. calculate the electrostatic potential from this charge density at the position of every electron to obtain an approximation of the potential from the other electrons
4. add this potential to the periodic potential of the ions
5. solve the one-electron Schrödinger equation for every electron using this potential
6. compare this wave function with the one obtained in the previous iteration
7. if these differ significantly then continue from step (2) otherwise you are done

Important 15.1.1. *Because the Hartree-Fock wave function contains all of the interactions between the electrons, the particles this wave function represent are not the same electrons we started with. Figuratively speaking these electrons differ from the „bare” electrons we started with because they are “dressed up” with the interaction of the lattice and of the other electrons, therefore these are non-interacting particles of charge $-e$ moving in a periodic potential created by the lattice ions and by the other (bare) electrons.*

15.1.3 Bloch electrons

The Hamiltonian of this system can be written in the form:

$$H = \sum_{j=1}^K H_o(\mathbf{p}_j, \mathbf{r}_j) \quad \text{where} \quad H_o = \frac{\mathbf{p}^2}{2m_e} + V(\mathbf{r})$$

The V potential is periodic:

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

and the periodicity is given by the lattice vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$. In Appendix 23.6 we determined the form of the wave function of an electron in a (weak) periodic potential. The result:

$$\psi(x) = u(x)e^{ikx} \quad \text{where} \quad u(x + na) = u(x) \quad (15.1.1)$$

i.e $u(x)$ is a lattice periodic function. It also depends on k so we will sometimes denote it by $u_k(x)$

A wave function of this functional form is called *Bloch function* and the corresponding particle is the *Bloch electron*. The Bloch function is a *crystal orbital*, as it is delocalized

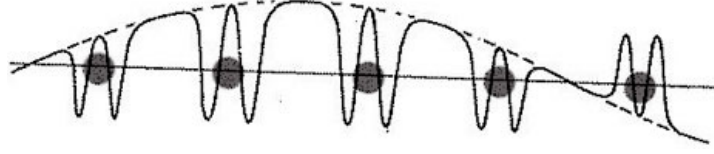


Figure 15.1: The Bloch function (solid line) is a free electron wave function (dashed line) modulated by a lattice periodic function

throughout the solid, and not localized around any particular atom. Thus the electron is shared by the whole crystal².

15.2 Crystal momentum of Bloch electrons. Dispersion relations

The momentum of the Bloch electron is calculated the usual way

$$\begin{aligned}
 p &= \int \psi^* \hat{p} \psi dx = \frac{\hbar}{i} \int \psi^* \frac{d}{dx} \psi dx = \\
 &= \frac{\hbar}{i} \int u^*(x) e^{-ikx} \frac{d}{dx} u(x) e^{ikx} dx = \\
 &= \frac{\hbar}{i} \int u^*(x) e^{-ikx} (u'(x) + u(x)(ik)) e^{ikx} dx = \\
 &= \frac{\hbar}{i} \int ik u^*(x) u(x) dx + \frac{\hbar}{i} \int u^*(x) u'(x) dx = \\
 &= \hbar k \underbrace{\int u^*(x) u(x) dx}_{=1} + \frac{\hbar}{i} \int u^*(x) u'(x) dx
 \end{aligned}$$

where the value of the first integral for normalized ψ -s is 1. The total momentum can be written as the sum of the momentum of a free electron of wave vector \mathbf{k} and a p_u momentum that describes the interaction with the crystal.

$$p = \hbar k + p_u \quad (15.2.1)$$

Here $\hbar k$ is called the *crystal momentum* or *quasi momentum* of the Bloch electron. As you can see the crystal momentum is not the same as the momentum of a free electron,

²In Section 15.4 we will see that such delocalized wave functions can be constructed even from wave functions of localized valence electrons.

because $p \neq \hbar k$. The corresponding kinetic energy according to Appendix 23.7 is

$$\mathcal{E}_{kin}(k) = \frac{\hbar^2 k^2}{2m_e} + \mathcal{E}_{cryst}(k) \quad (15.2.2)$$

Therefore the kinetic energy of a Bloch electron is not $\frac{p^2}{2m_e}$ with p from (15.2.1) but the sum of the kinetic energy corresponding to the quasi-momentum k and an interaction energy with the crystal. This means that although Bloch-electrons look like free electrons with a modified amplitude their energy is not completely kinetic.

The total energy vs k (*the Bloch electron dispersion relation*) can be determined by solving the Schrödinger equation for the unknown $u_k(x)$ function:

$$\begin{aligned} H\psi &= \mathcal{E}\psi \\ \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right) u_k(x) e^{ikx} &= \mathcal{E}_k u_k(x) e^{ikx} \\ \left(-\frac{\hbar^2}{2m} \left(\frac{d^2 u_k}{dx^2} + 2ik \frac{du_k}{dx} - k^2 u_k \right) + V(x) u_k(x) \right) e^{ikx} &= \mathcal{E}_k u_k(x) e^{ikx} \\ \left(\frac{-\hbar^2}{2m} \left(\frac{d}{dx} + ik \right)^2 + V(x) \right) u_k(x) &= \mathcal{E}_k u_k(x) \end{aligned} \quad (15.2.3)$$

$$H' u_k(x) = \mathcal{E}_k(x) u_k(x)$$

where there is a periodic boundary condition for $u_k(x)$ with the periodicity of the $V(x)$ crystal potential:

$$u_k(x + R) = u_k(x).$$

Those k crystal momenta that differ from each other only by some integer multiple of $2\pi/a$ correspond to the same k quasi-momentum. As a consequence k can be confined to the first Brillouin zone (or to any convenient primitive cell of the reciprocal lattice). Let k' is a value outside the first Brillouin zone, then it can be written as

$$k' = k + K$$

where K is a reciprocal lattice vector and k is inside the first Brillouin zone. Then in (15.1.1) (now explicitly denoting the k dependence of ψ and u)

$$\psi_{k'}(x) = u_{k'}(x) e^{ik'x} = u_{k'}(x) e^{ikx} e^{iKx} = u_{k'}(x) e^{ikx}$$

Notice that the k' index of u_k does not become k , which means that the exact form of u depends not on k only but on K as well.

Equation (15.2.3) then can be regarded as an eigenvalue problem restricted to a single cell of the crystal. As in the case of the Born-Karman periodic boundary condition this restriction of the wave function into a finite volume of space gives rise to an infinite number of *discrete* eigenvalues at every possible k . These eigenvalues will differ for different K values. It follows that the $\mathcal{E} = \mathcal{E}(k)$ dispersion relation will have more than one branches. As all K which corresponds to values of k' in the n -th Brillouin zone belongs to the same branch we may use the index of the branch instead of K :

$$\mathcal{E}(k') = \mathcal{E}(k + K) = \mathcal{E}_n(k) \quad (15.2.4)$$

$$u_{k'}(x) = u_{n,k}(x) \quad (15.2.5)$$

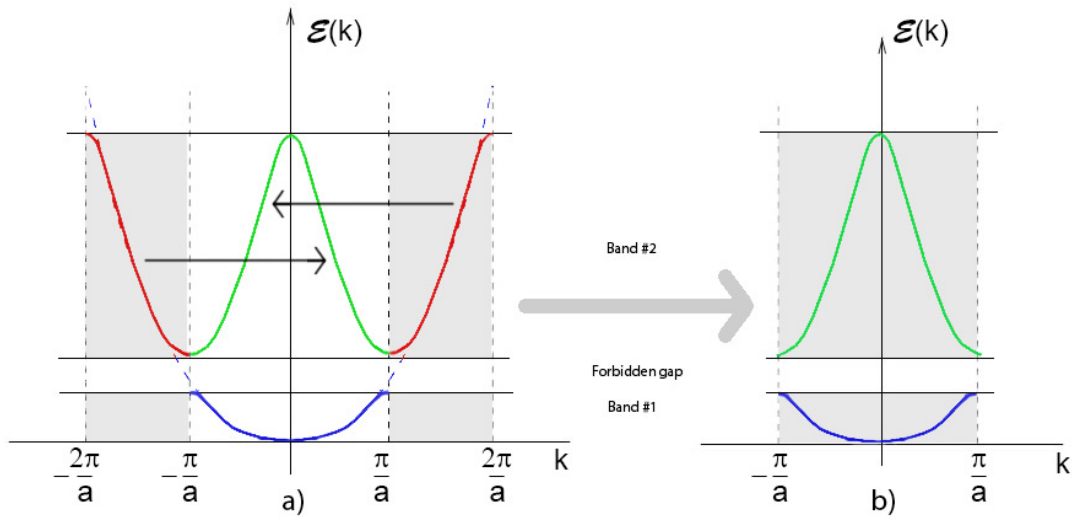


Figure 15.2: Dispersion relation of Bloch electrons. Extended (a) and reduced (b) zone pictures. On the extended zone picture the blue curve is the branch of the dispersion relation in the first, the red ones are the branches in the second Brillouin zone. Bands are shown as gray areas. The reduced zone picture (blue and green curves) is constructed from the extended one by moving all branches of the dispersion relation that lie in the n -th Brillouin zone (red for zone #2) back into the first Brillouin zone with a reciprocal lattice vector of length $K = \frac{2\pi}{a} \cdot n$ as denoted by the gray arrows. The green curve is the result.

In an ideal crystal the branches of the dispersion relation have a horizontal tangent at the edges of the Brillouin zone with gaps between them. The energy ranges these branches represent are called *energy bands*. Therefore the index of the branch n in (15.2.4) is called the *band index*.

No electron may have an energy between the top of a lower lying band and the bottom of the next (upper lying) band: there appears a *forbidden gap* between the bands.

Although for every k in the first Brillouin zone we can find an infinite number of energy levels (*reduced zone picture*) we may also display an *extended* view of the bands, when all branches are drawn in their own (2nd, 3rd, etc) Brillouin zone as on the left side of Fig. 15.2.

A comparison of the free electron and Bloch electron dispersion relation near the zone boundary is shown in Fig. 15.3.

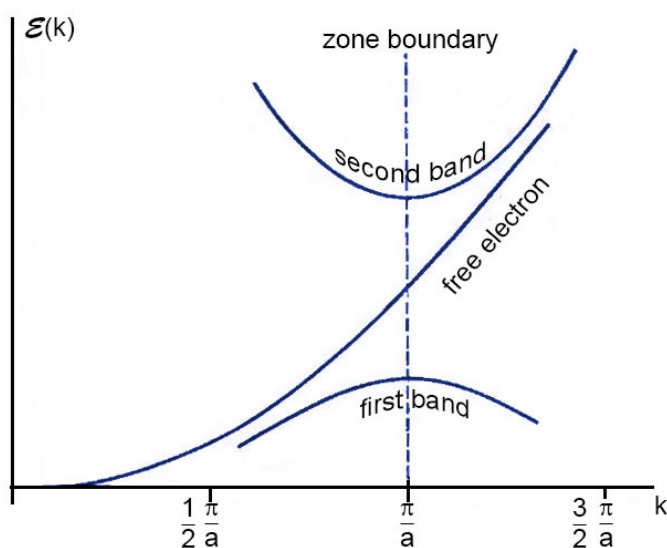


Figure 15.3: Dispersion relation of Bloch electrons near the zone boundary in the extended zone picture. In this figure we selected the region between $\frac{\pi}{2a}$ and $\frac{3\pi}{2a}$ as the first Brillouin zone and projected the curves into this region.

The smaller the magnitude of the crystal potential the smaller is this forbidden gap. In the $V \rightarrow 0$ limit (which from a quantum mechanical point of view corresponds to the Sommerfeld model: (almost) free electron in a box) the gap vanishes and in the reduced zone picture the free electron dispersion relation is folded back into the first Brillouin zone (see Fig. 15.4).

How can we explain the formation of the forbidden gap in a “plausible way”?

Bloch electrons are waves in the crystal. The wavelength of Bloch electrons is in the range of the lattice constant³, therefore the interaction of the wave function with

³Like for X-rays.

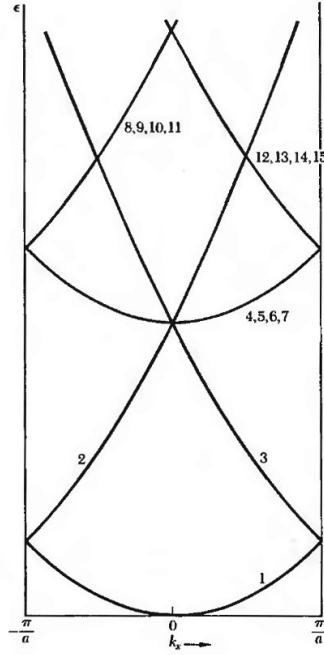


Figure 15.4: Dispersion relation of Bloch electrons in the limit of $V \rightarrow 0$ (*quasi free electron model*). The numbers refer to the line sections of the extended dispersion relation, which are projected back to the first Brillouin zone. (Line sections are numbered consecutively in every Brillouin zone. E.g. '2' denotes the line section from the 2nd Brillouin zone above π/a , while '3' is the line section from the same zone below $-\pi/a$.)

the lattice can be described as a reflection by the lattice planes. The condition of a constructive interference is given by the Bragg condition:

$$2a \sin \theta = n\lambda \quad \text{where } n = 1, 2, \dots$$

For an angle of incidence perpendicular to the zone boundary ($\theta = 90^\circ$) the path difference becomes

$$2a = n\lambda$$

which corresponds to a phase difference of

$$\Delta\varphi = \frac{2\pi}{\lambda} 2a$$

Constructive interference for perpendicular incidence occurs, when

$$\Delta\varphi = 2n\pi$$

$$k = \frac{2\pi}{\lambda} = n \frac{\pi}{a} \quad (15.2.6)$$

i.e. the lattice reflects back electrons with this momentum (i.e. $k' = -k$). This means that the momentum of an electron accelerated by e.g. an electrical field is reversed when it reaches the border of the first Brillouin zone (see next section for details), therefore electrons can not leave the first Brillouin zone when their energy changes continuously, neither can they move from one branch to another one this way.

Important 15.2.1. *When the crystal momentum of an electron is increased by a constant or slowly varying external field reaches the zone boundary it will be reflected back to the opposite zone boundary, therefore*

- *in the extended zone picture no such field can make the electron leave the Brillouin zone it occupies and move it to any other Brillouin zone, or*
- *in the reduced zone picture no such field can make the electron leave the actual branch of the dispersion relation.*

Every branch of the dispersion relation corresponds to an energy *band*, so this statement is equivalent to the following:

Important 15.2.2. *No constant or slowly varying external field can move an electron from one band to another band.*

Constant or slowly varying external fields change the electron energy as a function of its crystal momentum as described by the dispersion relation. Any process that excites an electron from one band to another one must change the electron energy abruptly. Such processes are collisions with photons, because the momentum of a photon is negligible compared to the momentum of a Bloch electron⁴.

The change in momentum at the boundary of the B-zone is $k' - k = \frac{2\pi}{a}$ which is a reciprocal lattice vector.

Bloch electrons are particles traveling with a constant velocity. They have a wave function of non diminishing amplitude. Therefore in perfect crystals at T=0K (when there are no lattice vibrations with which they can interact⁵) they can travel freely without any resistance. This is in striking contrast with the Drude- or Sommerfeld models, where the ion cores are the scattering centers independent of the temperature. Therefore the appearance of the resistivity requires that either lattice imperfections or lattice vibrations are to be present.

⁴When the material in question has an *indirect gap* – see near Figure 15.6 and Figure 15.3 – both the energy and the momentum may change in the excitation process. This requires a simultaneous collision with a photon and a *phonon* (the quantum of lattice vibrations).

⁵No interaction is possible with the zero-point vibrations.

Dispersion relations in 3D

In 3D the Bloch electron energy dispersion relation depends on the direction in which the electron travels. This can be visualized the same way as for the lattice vibrations: by selecting special high symmetry points in k -space and drawing the branches of the dispersion relation along directions connecting two high symmetry points in k -space. The dispersion relation is then called the *band structure* of the solid. The schematic

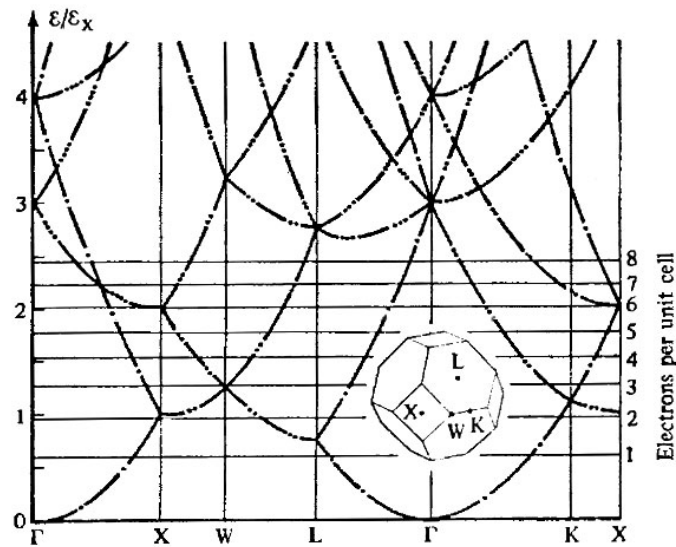


Figure 15.5: Dispersion relation in 3D for an FCC Bravais lattice. The horizontal lines give Fermi-energies for the indicated number of electrons per primitive cell. The number of dots on a curve specifies the number of degenerate free electron levels represented by the curve.

band pictures in Fig. 14.2 show just the allowed and forbidden regions of the energy axis marked by the branches of the dispersion relations and do not give the correct band structure. For instance the maximum of a lower lying band and the minimum of the upper lying band may be found at different \mathbf{k} values. A more detailed representation than in Fig. 14.2 is in Fig. 15.6. The band overlap if present need not occur in the same directions or at the same k values in the Brillouin zone. Materials with overlapping bands are metals or *semi-metals*. The alkali metals and the noble metals have only one valence electron per primitive cell, so they have to be metals. The alkaline earth metals however have two electrons per primitive unit cell, so they could have been insulators, but because their bands overlap in energy an electron can move into an empty band with only a small amount of energy readily provided by thermal excitations.

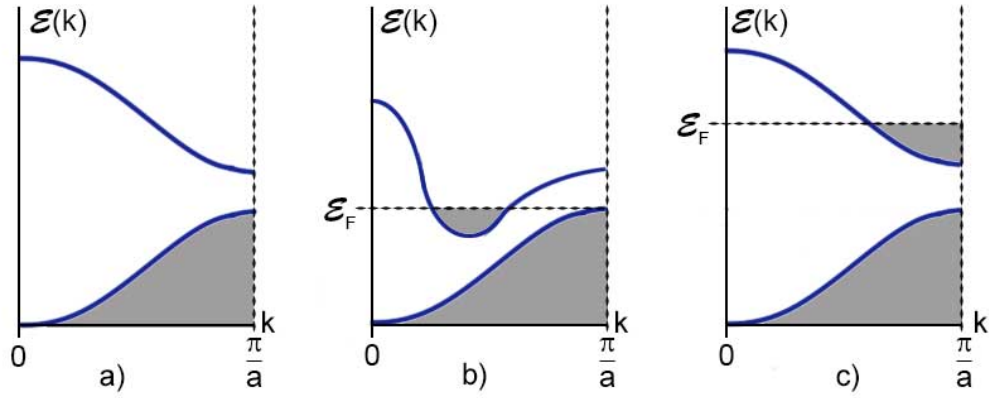


Figure 15.6: Band structure of insulators and metals. a) insulator, b) metal with overlapping bands and c) metal with non-overlapping partially filled conduction bands

15.3 Kinematics of electrons. Effective mass

A localized electron is characterized by a localized wave function, called a *wave packet*. There exist two different velocities for such a wave: the *phase velocity* $v_{ph} = \omega/k$ and the *group velocity* $v_g = d\omega/dk$ ⁶. In one dimension

$$\begin{aligned} v_g(k) &:= \frac{d\omega}{dk} \Rightarrow & v_g(k) &= \frac{1}{\hbar} \frac{d\mathcal{E}}{dk} \\ v_g(0) = 0 & \quad \text{and} & v_g(\pm\pi/a) &= 0 \end{aligned}$$

In regions far both from 0 and $(\pm\pi/a)$ v_g is almost the same as for a free electron:

$$v_g(k) \approx \frac{\hbar k}{m} \quad (15.3.1)$$

If a constant F force is acting on the electron

$$F = \hbar \frac{dk}{dt} \quad (15.3.2)$$

and

$$F = \frac{dm v_g(k)}{dt} = \frac{1}{\hbar} \frac{d}{dt} \left(m \frac{d\mathcal{E}}{dk} \right)$$

⁶ It is worth mentioning that for electrons the phase velocity is k dependent even in vacuum. That is the reason why an electron wave packet changes shape in vacuum too as time progresses.

i.e.

$$\hbar \frac{dk}{dt} = \frac{dmv_g(k)}{dt} = \frac{1}{\hbar} \frac{d}{dt} \left(m \frac{d\mathcal{E}}{dk} \right) \quad (15.3.3)$$

but as we saw before when the k crystal momentum reaches the edge of the Brillouin zone it will be reflected back ($k' = -k$), i.e. it will reappear at the other end of the Brillouin zone.

This behavior may be further examined in the following way: According to (15.3.3) far from 0 and the zone boundaries ($\pm\pi/a$) :

$$\frac{1}{\hbar} \frac{d}{dt} \left(m \frac{d\mathcal{E}}{dk} \right) = \hbar \frac{dk}{dt} \quad (15.3.4)$$

We may make this formula valid for all k values by replacing $m(= m_e)$ with an m_{eff} *effective mass*, defined a suitable way. Because

$$\begin{aligned} \frac{d}{dt} \left(m \frac{d\mathcal{E}}{dk} \right) &= m \frac{d}{dk} \frac{dk}{dt} \left(\frac{d\mathcal{E}}{dk} \right) = m \frac{d^2\mathcal{E}}{dk^2} \frac{dk}{dt} \\ m \left(\frac{1}{\hbar^2} \frac{d^2\mathcal{E}}{dk^2} \right) \hbar \frac{dk}{dt} &= \hbar \frac{dk}{dt} \end{aligned} \quad (15.3.5)$$

After canceling the common factors we find that if the effective mass is defined the following way:

$$\frac{1}{m_{eff}} = \frac{1}{\hbar^2} \frac{d^2\mathcal{E}}{dk^2} \quad (15.3.6)$$

then we can use

$$v_g(k) = \frac{\hbar k}{m_{eff}} \quad (15.3.7)$$

instead of formula (15.3.1) everywhere in the Brillouin zone. In 3D the effective mass depends on the direction the electron travels, i.e. on the components of \mathbf{k} . The 3D formula equivalent to (15.3.6) is:

$$\left(\frac{1}{m_{eff}} \right)_{i,j} = \frac{1}{\hbar^2} \frac{\partial^2 \mathcal{E}}{\partial k_i \partial k_j} \quad i, j = 1, 2, 3 \quad (15.3.8)$$

The effective mass incorporates the effects of the crystal structure on the dynamics of the electron. This underlines our previous statement that Bloch-electrons are not “ordinary” electrons, but particles “dressed up” with the interaction of the lattice.

The magnitude and even the sign of the second derivative of \mathcal{E} changes and correspondingly the magnitude and sign of the effective mass also changes as k is varied:

condition	m_{eff}	remark
$\mathcal{E}(k)$ has minimum at k	$m_{eff} > 0$	e.g. on the lowest branch when $k = 0$
$\mathcal{E}(k)$ has maximum at k	$m_{eff} < 0$	e.g. on the lowest branch when $k = \pm \frac{\pi}{a}$
$\mathcal{E}(k)$ has inflection point at k	$m_{eff} = \infty$	

The effective mass is positive at the bottom of a band and negative at the top of a band. When the curvature of the lower lying band is not the same as the curvature of the upper lying band the absolute values of the effective masses at the same k in these two bands differ. The wider is the band near its extremum the smaller is the magnitude (absolute value) of the effective mass there.

Example 15.1. *The dispersion relation of electrons in the valence and conduction bands near the band edges is approximated by the following functions:*

$$\mathcal{E}_v(k) = -3.024 \cdot 10^{-20} (k - 2.45 \cdot 10^8)^2 + 13 \quad [eV]$$

$$\mathcal{E}_c(k) = 4.65 \cdot 10^{-20} k^2 + 11.9 \quad [eV]$$

Express the effective masses of electrons in units of the free electron mass $m_e = 9.1 \cdot 10^{-31} kg$. **Solution In 1D**

$$\frac{1}{m_{eff}} = \frac{1}{\hbar^2} \frac{d^2 \mathcal{E}(k)}{dk^2}$$

The energies (in eV) converted to Joule are:

$$\mathcal{E}_v(k) = -4.845 \cdot 10^{-39} (k - 2.45 \cdot 10^8)^2 + 2.08 \cdot 10^{-18} \quad [J]$$

$$\mathcal{E}_c(k) = +7.450 \cdot 10^{-39} k^2 + 1.91 \cdot 10^{-18} \quad [J]$$

The second derivative of both functions gives twice the coefficient of the 2nd order terms, and so the electron effective masses in the conduction and valence bands are:

$$m_{eff}^{(c)} = -\frac{\hbar^2}{1.490 \cdot 10^{-38}} = -7.46 \cdot 10^{-31} [kg] = -0.819 m_e$$

$$m_{eff}^{(v)} = \frac{\hbar^2}{9.690 \cdot 10^{-39}} = 1.14 \cdot 10^{-30} [kg] = 1.260 m_e$$

Substituting the effective mass into formulas (14.2.2) and (14.2.4):

$$\sigma = \frac{ne^2\tau}{m_{eff}} \quad (15.3.9)$$

$$\mu = \left(\frac{e\tau}{m_{eff}} \right) \quad (15.3.10)$$

Because the effective mass depends on k and even at the same k value it may be different in different bands the mobility of charge carriers depends on k and the band index.

Bloch oscillations

While examining the behavior of the electrons in periodic lattices Bloch and Zener predicted that *the motion of electrons in a perfect crystal under the influence of a constant electric field would be oscillatory instead of uniform*

Background 15.3.1. *Let us describe the behavior of the electron in the lowest band in Fig. 15.2), under the influence of a constant E electric field!*

When $k = 0$ the momentum and the velocity are 0 and the acceleration is E/m_{eff} .

When k is near to 0 the wave function is a Bloch function with $u_k(x) \approx 1$ (almost a free electron wave function) and m_{eff} is positive. Near the origin the $\mathcal{E}(k)$ function may be approximated by a parabola. Until this approximation becomes invalid the effective mass is constant.

As the $p(k) = \hbar(k + G)$ momentum⁷ (see (15.2.1)) increases k also increases with it but m_{eff} remains constant. The $v_g(k) = \frac{1}{\hbar} \frac{d\mathcal{E}}{dk}$ velocity of the electron increases linearly and the acceleration is constant.

After the parabolic approximation fails more and more of the increase in p is transferred from the electron to the lattice. This corresponds to the increase in m_{eff} , (as the curvature of $\mathcal{E}(k)$ changes) which becomes ∞ at the inflection point. The acceleration decreases. The acceleration becomes 0 at the inflection point. Above the inflection point m_{eff} is negative. This means that as the momentum changes from k to $k + \Delta k$ the momentum transfer to the lattice from the electron is larger than the momentum transfer from the applied force to the electron. Although k still increases the acceleration is negative (deceleration), the velocity decreases, as the approach to the Bragg reflection (see previous section) results in an overall decrease of the forward momentum of the electron.

At the zone boundary $k = \frac{\pi}{a}$ (top of the band) the electron wave function which until now was a running wave becomes a standing wave instead. This standing wave may be written as the sum of two running waves that run in opposite directions. The velocity will be 0 and the acceleration remains negative. This is the turning point for the electron's movement. Up till now the electron traveled in the direction of the force.

A further increase of the total momentum p does not move k out of the first Brillouin zone, because the Bragg reflection changes k from $\frac{\pi}{a}$ to $-\frac{\pi}{a}$ that is the electron enters the Brillouin zone from the other side. The acceleration is still negative. The velocity becomes negative, the electron starts to move in a direction opposite to the force.

This means the electron turns back, and no longer travels in the original direction .

⁷ G is a vector of the reciprocal space.

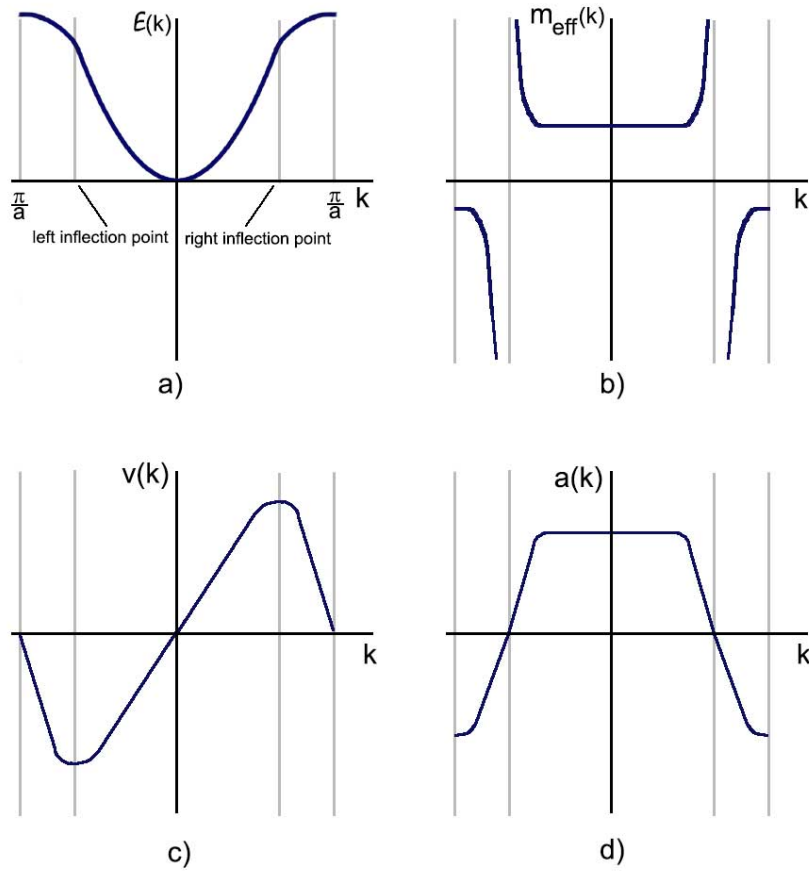


Figure 15.7: Bloch oscillations. a) The dispersion relation, b) m_{eff} in the first Brillouin zone, c) velocity and and d) acceleration throughout the first Brillouin zone

As k grows the magnitude of the velocity increases. When nearing the left inflection point the effective mass decreases toward minus infinity and the absolute value of the acceleration decreases. In the left inflection point the effective mass is infinite again (and negative) the acceleration becomes 0 .

A further increase of k changes the sign of m_{Eff} back to positive again and this is accompanied by the increase of the velocity (its magnitude decreases). Again the parabolic approximation becomes more and more valid and the acceleration will be constant.

When k reaches 0 the state of the electron will be the same as it was at the beginning and the process continues. This is the second turning point for the electron.

At first sight this oscillatory movement precludes an electric current but this is incorrect. We only discussed the motion of a single Bloch electron originally at rest. But the velocity

(wave vector) distribution without the external field is random while if an electric field is applied this distribution will have a non-random (drift) component. That is the *average momentum* will not be zero.

Interband transitions

As stated earlier interband transitions may occur as a result of photon absorption. These can be *direct* or *indirect* transitions for *direct gap* and *indirect gap* materials respectively (see Fig. 15.8)

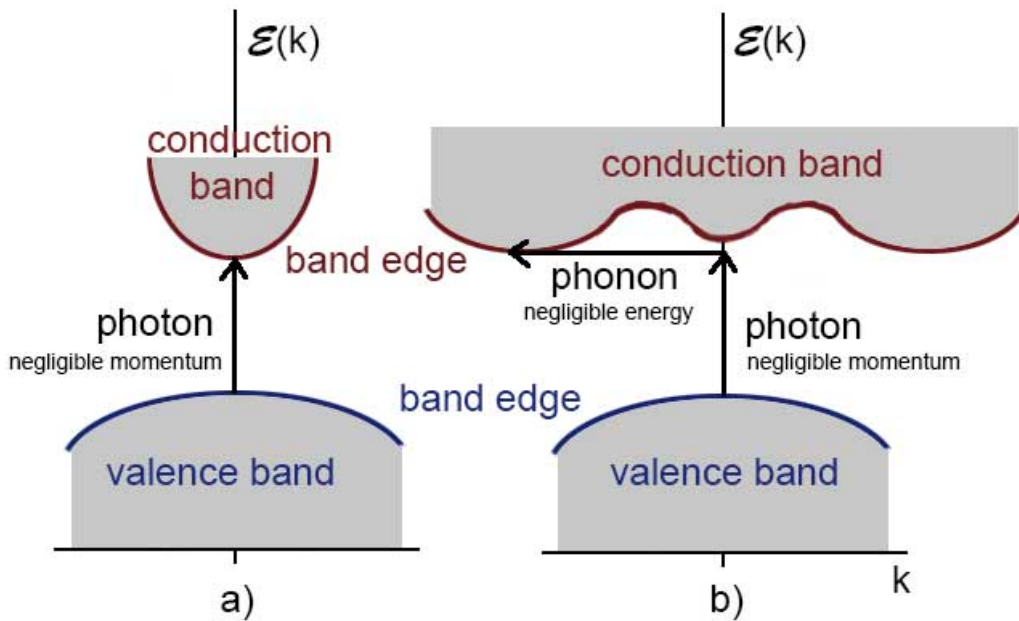


Figure 15.8: a) direct transition. The bottom of the conduction band is at the same k as the top of the valence band b) indirect transition. This involves both a photon and a phonon, because the bottom of the conduction band and the top of the valence band are at different k .

Transition may occur between any levels in the two bands for which both the momentum and the energy conservation is fulfilled. On the figure only the threshold transitions are shown. The threshold frequency ω_g for absorption by the direct transition determines the energy gap $\mathcal{E}_g = \hbar\omega_g$. The momentum of the electron remains almost the same in the transition because the wave vector k of the absorbed photon is very small. The absorption threshold for indirect transition is greater than the width of the gap: $\hbar\omega_g = \mathcal{E}_g + \hbar\omega_{ph}$, where ω_{ph} is the frequency of an emitted phonon of wave vector $\mathbf{G} \simeq -\mathbf{k}$. At higher temperatures where more phonons are already present in the system it is also possible

to absorb a photon and a phonon at the same time, in this case the required threshold energy is smaller: $\hbar\omega_g = \mathcal{E}_g - \hbar\omega_{ph}$.

Direct and indirect transitions may be distinguished by optical absorption experiments. Fig. 15.9 shows the threshold frequency in insulators at 0K.

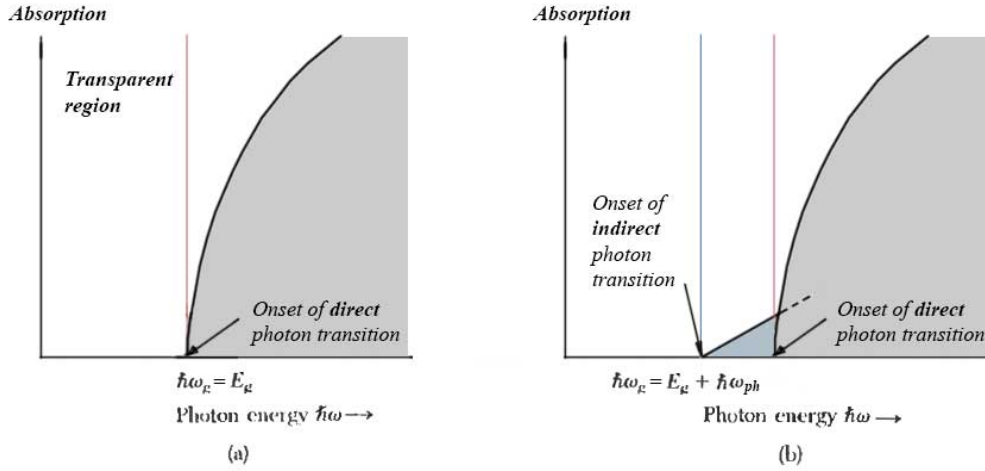


Figure 15.9: a) direct gap optical absorption b) indirect gap optical absorption

15.4 Width of the energy bands. Tight binding model.

The calculations of the band structure of a material are very complex. But some general features, like the width of the bands can be determined more easily. The width of a band can be calculated if we know the exact $\psi(\mathbf{r})$ Bloch function for the band. In one dimension e.g.

$$\mathcal{E} = \frac{\int \psi^* \hat{H} \psi dx}{\int \psi^* \psi dx} \quad (15.4.1)$$

If Bloch function ψ depends on some parameters, then \mathcal{E} will also depend on the same parameters and the band width can be determined using this. So how can we obtain a usable Bloch function?

Surprisingly Bloch functions can be created from slightly overlapping *localized* (atomic) wave functions of valence electrons:

$$\psi(x) = \sum_{n=0}^{N-1} e^{ikna} \varphi(x - na) = \underbrace{\sum_{n=0}^{N-1} e^{-ik(x-na)} \varphi(x - na)}_{u(x)} \cdot e^{ikx} = u(x) e^{ikx} \quad (15.4.2)$$

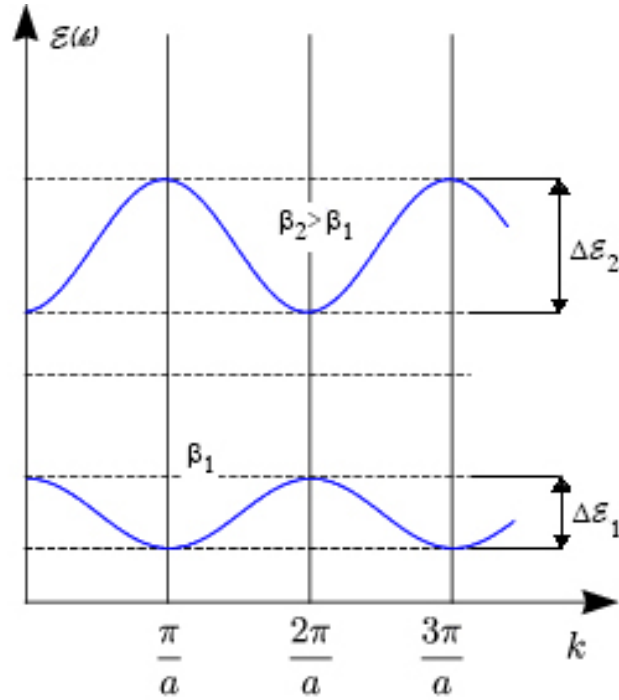


Figure 15.10: Results of a tight-binding calculation. The overlap of the wavefunctions is larger for the upper band, than for the lower one.

Because the constituent φ wave functions describe electrons are localized to the corresponding ion cores this is called the *tight binding model*. This is a good model for insulators and transition metals.

Details of the calculation can be found in Appendix 23.8. The result:

$$\mathcal{E} = \mathcal{E}_{atomic} - \alpha - 2\beta \cos ka \quad (15.4.3)$$

where α and β are integrals depending on the selected atomic wave functions and their overlaps. The possible energy values lie in a band determined by the term $2\beta \cos ka$. The cosine may change between -1 and +1, so the width of this band is 4β .

Parameter β is determined by the overlap of pairs of atomic wave functions:

$$\beta \propto \sum_{n,l} \int \varphi^*(x - la) \Delta V_p(x) \varphi(x - na) dx$$

Usually the overlap of atomic orbitals from atoms that are not nearest neighbors can be neglected and we need to calculate only terms for $l = n \pm 1$. This leads to narrow d bands and wide s bands.

Important 15.4.1. Integrals

$$\int \varphi^*(x - (n \pm 1)a) \Delta V_p(x) \varphi(x - na) dx \quad (15.4.4)$$

are called bond energy or two center integrals and they are the most important elements of the tight-binding model.

Higher lying states corresponds to larger overlap, i.e. a larger β parameter.

15.5 Conduction of metals. Electrons and Holes

In a completely full band, which does not overlap with any other band, electrons cannot change their momentum as a response to an external electric field, because there are no free electronic states available. No electric current can flow in these materials i.e.

$$\mathbf{j}_{full\ band} = 0.$$

unless one or more electrons are excited to the next band either by thermal excitations or by any other mechanisms.

However in metals either there are unoccupied levels in a band or a completely filled band and an other empty or partially filled band overlap. In both cases current may flow.

Important 15.5.1. Bands that are empty or partially filled at $T = 0\ K$ are called conduction bands, while bands completely filled at $T = 0\ K$ are called valence bands.

The current from a single electron moving with $\mathbf{v}(\mathbf{k})$ velocity in the conduction band is

$$\mathbf{j} = -e\mathbf{v}(\mathbf{k})$$

The total current from all electrons then⁸:

$$\mathbf{j}_e = -e \sum_{\substack{\text{occupied levels} \\ \text{in band}}} \mathbf{v}(\mathbf{k}) = -e \frac{1}{8\pi^3} \int_{\substack{\text{occupied levels} \\ \text{in band}}} \mathbf{v}(\mathbf{k}) d^3k$$

We can add to and subtract from \mathbf{j}_e the current of imaginary electrons occupying the empty levels without changing the current:

$$\mathbf{j}_e = \mathbf{j}_{\substack{\text{occupied levels} \\ \text{in band}}} + \underbrace{(\mathbf{j}_{\substack{\text{empty levels} \\ \text{in band}}} - \mathbf{j}_{\substack{\text{empty levels} \\ \text{in band}}})}_{=0}$$

⁸See (23.5.1)

Reordering the terms gives:

$$\mathbf{j}_e = \underbrace{\mathbf{j}_{\text{occupied levels in band}} + \mathbf{j}_{\text{empty levels in band}}}_{\mathbf{j}_{\text{full band}}=0} - \mathbf{j}_{\text{empty levels in band}}$$

i.e.

$$\mathbf{j}_{\text{occupied levels in band}} = -\mathbf{j}_{\text{empty levels in band}} \quad \text{or}$$

$$-\frac{e}{8\pi^3} \int_{\text{occupied levels in band}} \mathbf{v}(\mathbf{k}) d^3k = +\frac{e}{8\pi^3} \int_{\text{empty levels in band}} \mathbf{v}(\mathbf{k}) d^3k$$

Important 15.5.2. *The current usually attributed to the negatively charged electrons present in the band can also be expressed as a current of positively charged particles, called holes that occupy levels not filled with electrons.*

Because $\mathbf{j}_e = \mathbf{j}_h$ the electrical current of a partially full band may be described either as a current of electrons or as a current of holes, but we can use only one of these descriptions at any time in the same band.

The electron and hole current densities are:

$$\mathbf{j}_e \equiv -\frac{e}{8\pi^3} \int_{\text{occupied levels in band}} \mathbf{v}(\mathbf{k}) d^3k \mathbf{j}_h \equiv +\frac{e}{8\pi^3} \int_{\text{empty levels in band}} \mathbf{v}(\mathbf{k}) d^3k \quad (15.5.1)$$

Instead of saying that we have electrons in the conduction band we may also say that we have holes in the conduction band.

When a single electron with wave vector k_e is excited by a photon of negligible wave vector ($k_p \approx 0$) i.e. negligible momentum from the valence band to the conduction band then the total momentum of the valence band becomes $-\hbar k_e$. Wave vector $-k_e$ may be ascribed to a hole in the valence band. In this way one hole is an alternative description of a band with one missing electron.

When we excite electrons out of the valence band into the conduction band then the total electric current will be:

$$\mathbf{j}_e = \mathbf{j}_{\text{conduction band}} + \mathbf{j}_{\text{valence band}} \quad (15.5.2)$$

The current in the conduction band is described as a current of electrons. However it makes more sense to describe the current in the valence band not as current of all the remaining electrons in the band but as current of a single movable hole in a band which otherwise contains no holes. The conduction in this band will be *hole conduction*. Holes will behave exactly as positively charged particles, which have the same effective mass

than the electrons in the same band at the same k would have. In the valence band holes will be the *majority carriers*.

The effective mass of electrons in the conduction band and of holes in the valence band may differ, therefore according to equations (15.3.9) and (15.3.10) the conductivity and mobility of holes and electrons in the different bands together with the corresponding current densities may also be different. Therefore it is also possible that the hole current dominates, i.e. the observed current may be the current of holes instead of electrons. The sign of the dominating charge carriers can be determined using the Hall effect.

The Hall effect

If an electric current flows through a conductor in a magnetic field, the magnetic field exerts a transverse force on the moving charge carriers which tends to push them to one side of the conductor. This is most evident in a thin flat conductor as illustrated. A buildup of charge at the sides of the conductors will balance this magnetic influence, producing a measurable voltage between the two sides of the conductor. The presence of this measurable transverse voltage is called the Hall effect⁹ after E. H. Hall who discovered it in 1879.

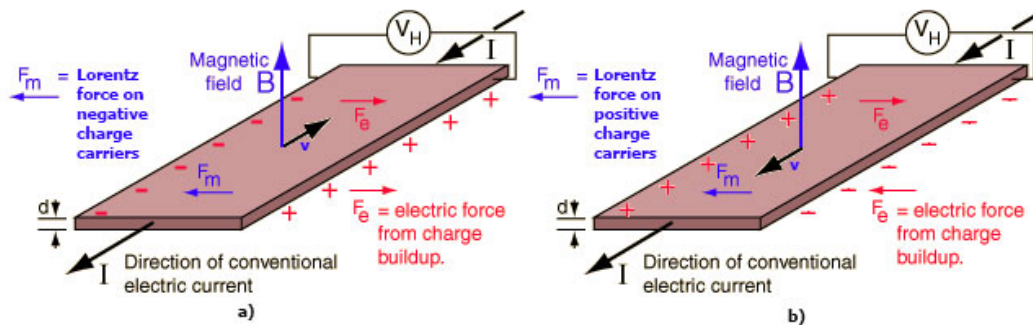


Figure 15.11: The Hall effect. a) for electrons b) for holes Both types of charge builds up on the same face of the sample.

It is easy to prove¹⁰ that $U_H = \frac{IB}{nq d}$. The sign of U_H determines the sign of the charge of the charge carriers ($q = \pm e$).

⁹The measurement of large magnetic fields on the order of a Tesla is often done by making use of the Hall effect. A thin film Hall probe is placed in the magnetic field and the transverse voltage (on the order of microvolts) is measured.

¹⁰In equilibrium the total Lorentz force acting on the q charge must be 0 as the force from the electric field of the charges at the opposite faces just cancels the force acting on the charge that moves with v velocity in the B magnetic induction. In our case v is orthogonal to B and both force acts perpendicular to the current: $q(E + vB) = 0$. Let us denote the width and thickness of the sample with w and d !

15.5.1 Effective mass of electrons and holes

In most metals electric current is carried by electrons, but there exist divalent metals (Mg, Cd) in which the current is carried by holes and not by electrons as the sign of the Hall voltage show for samples made from these metals . What is the reason for this?

Explanation:

the electron configuration of these metals are:

Mg : $1s^2 2s^2 2p^6 3s^2$,

Cd : $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2$

Both metals have a completely filled outer shell which results in a completely filled band. The fact that these are metals hints that there must exist an overlapping empty conduction band. When electrons are excited from one band to the other their effective mass may change. Mobility of the charge carriers is inversely proportional to their effective mass by the formula (C.f. (15.3.10)):

$$\mu = \frac{e\tau}{m_{eff}} \quad (15.5.3)$$

In Mg and Cd electrons in the conduction band have larger effective mass than holes in the valence band. This will result in a greater mobility of holes (with smaller effective mass compared to electrons) i.e. a larger hole current and smaller electron current. In this case both electrons and holes contribute to the total current, because they are in different bands, but the sign of the Hall voltage is determined by the dominating hole current.

Because $I = n q v_{drift} A$, where $A = w d$ and $U = E w$ it follows that

$$\frac{U_H}{w} = - \frac{I B}{q n w d}$$

Chapter 16

Semiconductors

16.1 Homogeneous semiconductors. Charge carrier concentrations. Donors and acceptors

Diamond, silicon, germanium and GaAs each have two atoms of valence four, so there are 8 electrons per primitive cell; their bands do not overlap, and the pure crystals are insulators at $T = 0$ K. Their forbidden gaps are:

Diamond - $\mathcal{E}_g = 5.5\text{eV}$, Si - $\mathcal{E}_g = 1.1\text{eV}$, Ge - $\mathcal{E}_g = 0.7\text{eV}$, GaAs - $\mathcal{E}_g = 1.42\text{eV}$. Let us recap from Section 14.1: Semiconductors are insulators with an energy gap around 1 eV.

16.1.1 Intrinsic semiconductors

Intrinsic semiconductors are pure perfect crystals. Only electrons excited into the conduction band and the remaining holes in the valence band may carry electricity. As the concentration of these are small they are very bad conductors. The probability of an electron to be excited to the conduction band can be expressed by the Boltzmann-factor:

$$\mathcal{P}(T) = e^{-\frac{\mathcal{E}_g}{k_B T}}$$

Example 16.1. *Compare the electron excitation probabilities in silicon at room temperature and at $T = 450$ K.* **Solution** Substituting into the probability formula: at room temperature

$$\mathcal{P}(300\text{K}) = 3.32 \cdot 10^{-19},$$

while at $T = 450\text{K}$:

$$\mathcal{P}(450\text{K}) = 4.79 \cdot 10^{-13}$$

So the ratio of these probabilities is

$$\frac{\mathcal{P}(450K)}{\mathcal{P}(300)} = 1.44 \cdot 10^6$$

i.e. the number of electrons available for conduction is about 10^6 larger at 450K than at 300 K. In the same range the increase in the resistivity because of lattice vibrations is about linear. As a consequence resistivity of Si decreases with increasing temperature, i.e. has a *negative thermal coefficient* contrary to metals. The same is true for all semiconductors.

When electrons are excited from the valence band to the conduction band holes appear in their place in the valence band. Because hole-electron pairs are continually created by thermal agitation of a semiconductor lattice, it might seem that the number of holes and free electrons would continually increase with time. This does not happen because free electrons are continually recombining with holes. At any temperature, an equilibrium is reached when the creation rate of hole-electron pairs is equal to the recombination rate.

Important 16.1.1. *The mean lifetime $\tau_n(s)$ of a free electron is the average time that the electron exists in the free state before recombination. The mean lifetime $\tau_p(s)$ for the hole is defined similarly. In an intrinsic semiconductor, $\tau_n(s)$ is equal to $\tau_p(s)$ because the number of free electrons must be equal to the number of holes. However, the addition of foreign atoms (impurities) to the semiconductor lattice can cause the mean lifetimes to be unequal.*

In contrast with metals where the conduction can be attributed to either movable holes or movable electrons *in the same band* but not both at the same time, in semiconductors the electrons and the holes move in different bands, therefore both electron and hole currents exist independently.

Holes are positively charged particles whose e charge is of the same magnitude and opposite sign as the electron charge. The electric current in a semiconductor is the sum of currents of electrons and holes:

$$\mathbf{j} = \mathbf{j}_e + \mathbf{j}_h = -n_e e \langle \mathbf{v}_e \rangle + n_h e \langle \mathbf{v}_h \rangle$$

Substituting $\langle v_e \rangle = \mu_e \mathbf{E}$ and $\langle v_h \rangle = \mu_h \mathbf{E}$ and $n_i \equiv n_e = n_h$

$$\mathbf{j} = e(n_e \mu_e + n_h \mu_h) \mathbf{E} \quad (16.1.1)$$

$$\begin{aligned} \mathbf{j} &= n_i e (\mu_e + \mu_h) \mathbf{E} \quad \text{and} \\ \mathbf{j} &= \sigma \mathbf{E} \end{aligned}$$

$$\sigma = n_i e (\mu_e + \mu_h) \quad (16.1.2)$$

Example 16.2. A rod of intrinsic Si is 1 cm long and has a diameter of 1mm. At room temperature, the intrinsic concentration in the silicon is $n_i = 1.5 \cdot 10^{16} \text{m}^{-3}$. The electron and hole mobilities are $\mu_e = 0.13 \text{m}^2 \text{V}^{-1} \text{s}^{-1}$ and $\mu_h = 0.05 \text{m}^2 \text{V}^{-1} \text{s}^{-1}$. Calculate the conductivity σ of the silicon and the resistance R of the rod. **Solution**

$$\sigma = n_i e (\mu_e + \mu_h) = 4.33 \cdot 10^{-4} \text{ } \Omega \text{m}$$

$$R = \frac{l}{\sigma d^2 \pi / 4} = 29.4 \text{M}\Omega$$

Both the curvature of the branches of the electron dispersion relation and the sign of the effective masses of conduction electrons and valence holes differ. As an example let us compare the band structure of Si and Ge.

Both Ge and Si have a diamond structure (which can be viewed as either an fcc lattice with 2 atoms basis or 2 single atom fcc lattices displaced by 1/4th of the main diagonal of the unit cell). Therefore the reciprocal lattice is a bcc lattice. At the band edges (i.e. at the minimum and maximum positions) the constant energy surfaces in 3D are ellipsoids of revolution and can be written using the effective masses (here denoted by an asterix):

$$\mathcal{E}_{cond}(\mathbf{k}) = E_c + \frac{\hbar^2}{2} \left(\frac{k_1^2}{m_{e,1}^*} + \frac{k_2^2}{m_{e,2}^*} + \frac{k_3^2}{m_{e,3}^*} \right) \quad (16.1.3)$$

$$\mathcal{E}_{val}(\mathbf{k}) = E_v + \frac{\hbar^2}{2} \left(\frac{k_1^2}{m_{h,1}^*} + \frac{k_2^2}{m_{h,2}^*} + \frac{k_3^2}{m_{h,3}^*} \right) \quad (16.1.4)$$

From Fig. 16.1 the band gap E_g is 1.12 eV. Because the conduction band minimum and valence band maximum are at a different k vector Si has an *indirect gap*. (C.f. Section 15.3.) There are 6 equivalent valleys in the conduction band (corresponding to the same effective masses) in the $\langle 100 \rangle$ direction: $(k_m, 0, 0)$, $(-k_m, 0, 0)$, $(0, k_m, 0)$, $(0, -k_m, 0)$, $(0, 0, k_m)$, $(0, 0, -k_m)$, where $k_m = 5 \text{ } 1/\text{nm}$. The effective mass of these *anisotropic* minima is characterized by a longitudinal mass along the corresponding equivalent $\langle 100 \rangle$ directions and two transverse masses in the plane perpendicular to the longitudinal direction. The two transverse effective masses of holes and electrons in Si (and Ge) are equal. There are also 3 maxima at $k = 0$ called light and heavy hole bands and a so called *split-off* band.

As you can see Ge is also an indirect gap semiconductor (see Fig. 16.2). Like Si Ge also has 6 equivalent valleys but in the $\langle 111 \rangle$ direction, it has light, heavy and split-off hole bands and the corresponding transverse, longitudinal light and heavy hole and split-off effective masses.

A semiconductor with a direct gap is gallium arsenide. Fig. 16.3 shows its band structure.

Important 16.1.2. *In the following we use the simple model that electrons in the conduction band are free particles in a potential box whose bottom is at \mathcal{E}_c , i.e. $\mathcal{E} = \mathcal{E}_c + \frac{\hbar^2 k^2}{2m_e}$ where m_e denotes the effective mass for density of states calculations of the electron in the conduction band. Similarly holes are free particles in a potential box whose bottom¹ is at \mathcal{E}_v with an effective mass for density of states calculations m_h .*

The density of states then are given by (C.f. equation (14.3.10)):

$$g_c(\mathcal{E}) = \frac{8\pi\sqrt{2m_e^3}}{h^3} \sqrt{\mathcal{E} - \mathcal{E}_c} \quad \mathcal{E} \geq \mathcal{E}_c$$

$$g_v(\mathcal{E}) = \frac{8\pi\sqrt{2m_h^3}}{h^3} \sqrt{\mathcal{E}_v - \mathcal{E}} \quad \mathcal{E} \leq \mathcal{E}_v$$

Therefore the *effective mass for density of state calculations* is the one which provides the density of states using the expression for one isotropic maximum or minimum. For instance for a single band minimum described by a longitudinal and two transverse effective masses (e.g. Si, Ge) the effective mass for density of states calculations is the geometric mean of the three masses:

$$m_c = N_{min} \sqrt[3]{m_l m_t m_t}$$

where N_{min} is the number of equivalent minima.

The effective masses of electrons and holes in Si and Ge are in Table 16.1. The *effective mass for conductivity* of electrons m_{cc} is what we must use for mobility or diffusion constant calculations. For cubic isotropic semiconductors with anisotropic dispersion relations minima (again e.g. Si and Ge), one has to sum over the effective masses in the different minima along the equivalent directions. The resulting effective mass for bands which have ellipsoidal constant energy surfaces is given by:

$$m_{cc} = \frac{3}{\frac{1}{m_l} + \frac{1}{m_l} + \frac{1}{m_t}}$$

As we discussed, in homogeneous perfect semiconductor crystals, the so called *intrinsic semiconductors*, the number of conduction electrons (n_c) and (movable) valence band holes (p_v) are equal. Because the probability that a level is filled in with electrons is given by the Fermi-Dirac distribution function $f(\mathcal{E})$ and holes are missing electrons, the

¹Hole energy increases as electron energy decreases.

electrons			
	rel. eff.mass	Si	Ge
longitudinal	m_l/m_e	0.98	1.59
transverse	m_t/m_e	0.19	0.0815
dens.of.states	m_c/m_e	0.36	0.22
conduct.	m_{cc}/m_e	0.26	0.12
holes			
	rel. eff.mass	Si	Ge
heavy	m_h/m_e	0.49	0.33
light	m_{lp}/m_e	0.16	0.043
split-off band	m_{so}/m_e	0.24	0.084
dens.of.states	m_v/m_e	0.81	0.34

Table 16.1: Effective masses in Si and Ge

probability of a level to have a hole equals to $(1 - f(\mathcal{E}))$. The expectation values of n_c and p_v can be calculated by²

$$n_c(T) = \int_{\mathcal{E}_c}^{\infty} g_c(\mathcal{E}) \frac{d\mathcal{E}}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} \quad (16.1.5)$$

$$p_v(T) = \int_{-\infty}^{\mathcal{E}_v} g_v(\mathcal{E}) \left(1 - \frac{d\mathcal{E}}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} \right) \quad (16.1.6)$$

$$= \int_{\mathcal{E}_v}^{\infty} g_v(\mathcal{E}) \frac{d\mathcal{E}}{1 + e^{(\mathcal{E}_F - \mathcal{E})/k_B T}} \quad (16.1.7)$$

where $g_c(\mathcal{E})$ and $g_v(\mathcal{E})$ are the density of states in the conduction and valence bands respectively.

$$n_c(T) = \frac{8\pi\sqrt{2m_e^3}}{h^3} \int_{\mathcal{E}_c}^{\infty} \frac{\sqrt{\mathcal{E} - \mathcal{E}_c}}{e^{(\mathcal{E}-\mathcal{E}_F)/k_B T} + 1} d\mathcal{E} \quad (16.1.8)$$

$$p_v(T) = \frac{8\pi\sqrt{2m_h^3}}{h^3} \int_{-\infty}^{\mathcal{E}_v} \frac{\sqrt{\mathcal{E}_v - \mathcal{E}}}{e^{(\mathcal{E}_F - \mathcal{E})/k_B T} + 1} d\mathcal{E} \quad (16.1.9)$$

These integrals cannot be expressed analytically at non-zero temperatures³. We either

²Strictly speaking the upper limit for n_c should be $\mathcal{E}_{top\ of\ the\ conduction\ band}$ and the lower limit for p_v should be $\mathcal{E}_{bottom\ of\ the\ valence\ band}$, but the f_{F-D} function goes to 0 there, so we can use $\pm\infty$ instead.

calculate them approximately or numerically. For *non-degenerate* semiconductors, i.e. for semiconductors, where \mathcal{E}_F is at least $3k_B T$ away from either band edge⁴ the Maxwell-Boltzmann distribution function can be used instead of the Fermi-Dirac one because $(\mathcal{E} - \mathcal{E}_F)$ is large (c.f. Chapter 9). In this case⁵:

$$n_c(T) = \frac{8\pi\sqrt{2m_e^3}}{h^3} \int_{\mathcal{E}_c}^{\infty} \sqrt{\mathcal{E} - \mathcal{E}_c} e^{-(\mathcal{E} - \mathcal{E}_F)/k_B T} d\mathcal{E}$$

$$p_v(T) = \frac{8\pi\sqrt{2m_h^3}}{h^3} \int_{-\infty}^{\mathcal{E}_v} \sqrt{\mathcal{E}_v - \mathcal{E}} e^{-(\mathcal{E}_F - \mathcal{E})/k_B T} d\mathcal{E}$$

the integrals can be calculated analytically⁶

$$n_c(T) = N_c e^{-\frac{\mathcal{E}_c - \mathcal{E}_F}{k_B T}} \quad (16.1.10a)$$

$$p_v(T) = P_v e^{-\frac{\mathcal{E}_F - \mathcal{E}_v}{k_B T}} \quad (16.1.10b)$$

³At 0K the calculation is simple because the Fermi-Dirac function is 1 below \mathcal{E}_F and 0 above it. Therefore

$$\begin{aligned} n_0 \equiv n_c(0) &= \int_{\mathcal{E}_c}^{\mathcal{E}_F} g_c(\mathcal{E}) d\mathcal{E} \\ &= \frac{8\pi\sqrt{2m_e^3}}{h^3} \int_{\mathcal{E}_c}^{\mathcal{E}_F} \sqrt{\mathcal{E} - \mathcal{E}_c} d\mathcal{E} \\ &= \frac{16\pi\sqrt{2m_e^3}}{3h^3} (\mathcal{E}_F - \mathcal{E}_c)^{3/2} \quad \text{for } \mathcal{E}_F \geq \mathcal{E}_c \end{aligned}$$

similarly

$$p_0 \equiv p_v(T) = \frac{16\pi\sqrt{2m_h^3}}{3h^3} (\mathcal{E}_v - \mathcal{E}_F)^{3/2} \quad \text{for } \mathcal{E}_v \geq \mathcal{E}_F$$

⁴This is true to Si, Ge and Ga As.

⁵when $|\mathcal{E} - \mathcal{E}_F| \gg k_B T$

$$\frac{1}{e^{(\mathcal{E} - \mathcal{E}_F)/k_B T} + 1} \approx e^{-(\mathcal{E} - \mathcal{E}_F)/k_B T}$$

$$1 - \frac{1}{e^{(\mathcal{E} - \mathcal{E}_F)/k_B T} + 1} = \frac{1}{e^{(\mathcal{E}_F - \mathcal{E})/k_B T} + 1} \approx e^{-(\mathcal{E}_F - \mathcal{E})/k_B T}$$

⁶Calculate the integrals by variable substitution. E.g. for n_c let $y \equiv \sqrt{\mathcal{E} - \mathcal{E}_c}$. Then $d\mathcal{E} = 2y dy$ and at the lower limit $y(\mathcal{E} = \mathcal{E}_c) = 0$ The exponent becomes $\mathcal{E} - \mathcal{E}_F = y^2 + \mathcal{E}_c - \mathcal{E}_F$ and the integral:

$$\int_0^{\infty} y^2 e^{-(y^2 + \mathcal{E}_c - \mathcal{E}_F)/k_B T} dy = 2e^{-(\mathcal{E}_c - \mathcal{E}_F)/k_B T} \int_0^{\infty} y^2 e^{y^2/k_B T} dy = \frac{\sqrt{\pi k_B T}^3}{2} e^{-(\mathcal{E}_c - \mathcal{E}_F)/k_B T}$$

where N_c and P_v are the *effective density of states* in the conduction and valence bands respectively:

$$N_c(T) = 2 \cdot \left(\frac{2\pi m_e k_B T}{h^2} \right)^{\frac{3}{2}} \quad (16.1.11)$$

$$P_v(T) = 2 \cdot \left(\frac{2\pi m_h k_B T}{h^2} \right)^{\frac{3}{2}} \quad (16.1.12)$$

In this formula again the *effective mass for density of states* must be used.

Example 16.3. Calculate the effective carrier densities of states in the conduction and valence bands of germanium and silicon at 300 K. **Solution** Substitute the effective masses for the density of states from Table 16.1 into (16.1.11) and (16.1.12).

cm^{-3}	Ge	Si
$N_c(300K)$	1.02 10^{19}	2.81 10^{19}
$P_v(300K)$	5.64 10^{18}	1.83 10^{19}

The general formula from the calculations above can be expressed in the simple form

$$N_c(T) = 2.5 \cdot 10^{19} \left(\frac{m_c}{m_e} \right)^{\frac{3}{2}} \left(\frac{T}{300K} \right)^{\frac{3}{2}} \quad (16.1.13)$$

$$P_v(T) = 2.5 \cdot 10^{19} \left(\frac{m_h}{m_e} \right)^{\frac{3}{2}} \left(\frac{T}{300K} \right)^{\frac{3}{2}} \quad (16.1.14)$$

A very useful fact is that the product $n_c(T)p_v(T)$ for non degenerate semiconductors⁷ is independent of the \mathcal{E}_F Fermi energy:

$$n_c(T)p_v(T) = N_c(T)P_v(T)e^{-\frac{\mathcal{E}_c - \mathcal{E}_v}{k_B T}} = N_c(T)P_v(T)e^{-\frac{\mathcal{E}_g}{k_B T}} \quad (16.1.15)$$

so if we know the concentration of either of these *at a given temperature*(!) we can calculate the other one. This is the *mass action law* for non degenerate semiconductors. This formula remains valid (see Section 16.1.2) even for non intrinsic semiconductors⁸.

For intrinsic semiconductors

$$n_c(T) = p_v(T) \quad (16.1.16)$$

$$n_i \equiv n_c(T) = p_v(T) = \sqrt{N_c(T)P_v(T)} e^{-\frac{\mathcal{E}_g}{2k_B T}} \quad (16.1.17)$$

⁷Intrinsic semiconductors are usually non-degenerate.

⁸i.e. when a semiconductor is *doped* (see below)

From (16.1.16), (16.1.11) and (16.1.12):

$$N_c e^{-\frac{\mathcal{E}_F - \mathcal{E}_c}{k_B T}} = P_v e^{-\frac{\mathcal{E}_v - \mathcal{E}_F}{k_B T}}$$

taking the logarithm of both sides and reordering the resulting equation we can obtain \mathcal{E}_F

$$\mathcal{E}_F = \frac{1}{2}(\mathcal{E}_c + \mathcal{E}_v) + \frac{1}{2}k_B T \ln \left(\frac{P_v}{N_c} \right) \quad (16.1.18)$$

and using (16.1.11) and (16.1.12) and that $\frac{1}{2}(\mathcal{E}_c + \mathcal{E}_v) = \frac{1}{2}\mathcal{E}_g + \mathcal{E}_v$

$$\mathcal{E}_F = \frac{1}{2}\mathcal{E}_g + \mathcal{E}_v + \frac{3}{4}k_B T \ln \left(\frac{m_h}{m_c} \right) \quad (16.1.19)$$

When the effective masses are equal ($m_h = m_c$) the logarithm is zero. Then:

$$\mathcal{E}_F = \frac{1}{2}\mathcal{E}_g + \mathcal{E}_v \quad \text{where } m_h = m_e \quad (16.1.20)$$

that is for non-degenerate intrinsic semiconductors with equal electron and hole effective masses the Fermi level lies in the middle of the forbidden gap. In the general case the chemical potential shifts toward the band with the lower effective mass, but this deviation can be practically neglected even at room temperature.

16.1.2 Extrinsic (doped) semiconductors

To create semiconductor devices intrinsic semiconductors, i.e. pure single crystals with negligible concentration of impurities (foreign atom content) are not sufficient. For practical use semiconductors are *doped*, i.e. a well known concentration of selected foreign elements called *dopants* are introduced in the crystal in *very small concentrations* to change its electrical properties⁹. Dopant concentrations are $\approx 10^{13} - 10^{16} \text{ cm}^{-3}$ which is $10^{-9} - 10^{-6}$ times smaller than the concentration of the semiconductor atoms ($\approx 10^{22} \text{ cm}^{-3}$). Dopants can come in two flavors: they either have more or fewer valence electrons than the atoms (molecules) that form the crystal.

Background 16.1.1. *For Group IV semiconductors¹⁰ such as silicon, germanium, and silicon carbide, the most common dopants are acceptors from Group III or donors from*

⁹The color of some gemstones is also caused by dopants. For example, ruby and sapphire are both aluminum oxide, the former getting its red color from chromium atoms, and the latter doped with any of several elements, giving a variety of colors.

¹⁰When discussing periodic table groups, semiconductor physicists always use an older notation, not the current IUPAC group notation. For example, the carbon group is called "Group IV", not "Group 14"

Group V elements. By doping pure Group IV semiconductors with Group V elements such as phosphorus, extra valence electrons are added that become un-bonded from individual atoms and allow the compound to be an electrically conductive n-type semiconductor. Doping with Group III elements, which are missing the fourth valence electron, creates "broken bonds" (holes) in the silicon lattice that are free to move. The result is an electrically conductive p-type semiconductor.

The energy required to remove an electron from a donor atom can be approximated using a hydrogen-like model. After all, the donor atom consists of a positively charged ion and an electron just like the proton and electron of the hydrogen atom. The difference however is that the average distance between the electron and the donor ion is much larger since the electron occupies one of the outer orbitals. By modeling the surroundings of an atom with a continuous medium of permittivity ϵ the Bohr radius of the n -th orbital of the single valence electron will be larger than that for H in vacuum by a factor of $\epsilon_r m_e / m_e^*$, where we used an asterisk to denote the effective electron mass:

$$r_n = \frac{2(2\pi)^3 n^2}{m_e^* e^2} \epsilon_o \epsilon_r$$

The ionization energy¹¹ will be correspondingly smaller:

$$\mathcal{E}_{ion} = 13.6 \frac{m_e^*}{m_e \epsilon_r^2} [eV]$$

For silicon $\epsilon_r = 11.7$, $m_e^* = 0.12 m_e$ so the ionization energy is $\mathcal{E}_{Si,ion} \approx 13.6 \cdot 0.12 / 11.7^2 = 0.012 eV$. The exact value depends on the actual donor atom. By similar consideration we can see for acceptors too that their "ionization energy" (or the energy for electron capture) in Si is also small.

A justification of our starting assumption is given by calculating the first Bohr radius in Si using the well known value for the first Bohr radius in vacuum $a_o = 0.05 nm$. The result is $r_o^{Si} \approx 8 nm$ which is very large relative to the lattice constant, therefore we may regard the neighboring Si atoms as a continuous medium.

Thus the donor atoms can be easily ionized giving the system movable conduction electrons and creating immobile holes localized at the donor atoms. Similarly the acceptor atoms trap electrons making them immobile and add movable holes to the system. In the band structure these atoms give *shallow* donor or acceptor levels in the forbidden gap near the band edges.

Conduction in doped semiconductors is different from the intrinsic case, because dopants contribute movable charge carriers (electrons or holes) for conduction thus increasing the conductivity, while keeping opposite charges (holes or electrons) localized.

¹¹Remember the H ionization energy is 13.6 eV.

acceptor in	B	Al	Ga	In	Te
Si	0.046	0.057	0.065	0.16	0.26
Ge	0.01	0.01	0.061	0.011	0.01
donor in	P	As	Sb	Bi	
Si	0.046	0.057	0.065	0.16	.
Ge	0.01	0.01	0.061	0.011	.

Table 16.2: Shallow donor and acceptor levels in Si and Ge in eV.

The neutrality of the semiconductor requires that the total of positive charges per unit volume from p_v movable holes, and N_d^+ immobile donor ions and negative charges from n_c movable electrons and N_a^- immobile acceptor ions is zero, therefore:

$$N_d^+ + p_v = N_a^- + n_c \quad (16.1.21)$$

($N_d^+ \leq N_d$ and $N_a^- \leq N_a$ are the ionized part of the donor and acceptor atoms respectively.)

Semiconductors in which the majority carriers are electrons, i.e. in which there are more donors than acceptors are called *n-type semiconductors*, while semiconductors in which the majority carriers are holes, i.e. in which there are more acceptors than donors are called *p-type semiconductors*.

For n-type semiconductors ($N_d^+ > N_a^-$ where $N_d^+ \approx N_d$ and $N_a^- \approx N_a$):

$$n_c \approx (N_d - N_a) \quad p_v \approx \frac{n_i^2}{N_d - N_a} \quad (16.1.22)$$

while for p-type semiconductors ($N_a > N_d$):

$$p_v \approx (N_a - N_d) \quad n_c \approx \frac{n_i^2}{N_a - N_d} \quad (16.1.23)$$

Example 16.4. A rod of n-type extrinsic Si is 1 cm long and has a diameter of 1mm. At room temperature, the donor concentration is $5 \cdot 10^{14} \text{ atom/cm}^3$ and this corresponds to 1 impurity for 10^8 Si atoms. A steady $2\mu\text{A}$ current is flowing through the bar. Determine the electron and hole concentrations, the conductivity and the voltage across the rod. The intrinsic electron concentration in silicon is¹² $n_i = 1.01 \cdot 10^{10} \text{ m}^{-3}$. The electron mobility is $\mu_e = 0.13 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$. **Solution n-type:** $N_a \approx 0$. **From (16.1.22)** $n = N_d = 5 \cdot 10^{20} \text{ 1/m}^3$,

$$p_v = \frac{n_i^2}{N_d} = 4.5 \cdot 10^{13} \text{ 1/m}^3$$

¹²The previously accepted value before 1991 was $1.45 \cdot 10^{10} \text{ m}^{-3}$

From (16.1.1) and using that $p_v \ll n_c$

$$\sigma = en_c \mu_e = 0.104 S/m$$

The voltage across the rod:

$$U = \frac{I l}{\sigma A} = 0.12 V$$

Fermi level of doped semiconductors

Consider an n-type semiconductor. Near $T = 0K$ all electrons in the conduction band must come from the donor levels. The excitation energy required for this is small and this situation corresponds to an intrinsic semiconductor with a band gap of $\mathcal{E}_c - \mathcal{E}_{donor}$. Similarly for acceptor levels $\mathcal{E}_{acceptor} - \mathcal{E}_v$ can be used. So in this case¹³ The Fermi energies $\mathcal{E}_F^{(d)}$ and $\mathcal{E}_F^{(a)}$ attributed to the dopants are (c.f. equation (16.1.19))

$$\mathcal{E}_F^{(d)} = \frac{1}{2} (\mathcal{E}_c + \mathcal{E}_{donor}) + \frac{3}{4} k_B T \ln \left(\frac{m_h}{m_c} \right) \quad (16.1.24a)$$

$$(16.1.24b)$$

$$\mathcal{E}_F^{(a)} = \frac{1}{2} (\mathcal{E}_v + \mathcal{E}_{acceptor}) + \frac{3}{4} k_B T \ln \left(\frac{m_h}{m_c} \right) \quad (16.1.24c)$$

The equivalent of equations (16.1.10) in the case when the number of electrons from the valence band and holes from the conduction band are negligible relative to the ones due to the dopants¹⁴:

$$n_c^{(d)}(T) = N_c e^{-\frac{\mathcal{E}_c - \mathcal{E}_F^{(d)}}{k_B T}} \quad \text{for an n-type semiconductor} \quad (16.1.25a)$$

$$p_v^{(a)}(T) = P_v e^{-\frac{\mathcal{E}_F^{(a)} - \mathcal{E}_v}{k_B T}} \quad \text{for a p-type semiconductor} \quad (16.1.25b)$$

Substituting both carrier concentrations n_c and p_v in (16.1.25) with their equivalent intrinsic n_i gives:

$$n_c^{(d)}(T) = n_i e^{-\frac{\mathcal{E}_F^{(d)} - \mathcal{E}_F}{k_B T}} \quad \text{for an n-type semiconductor} \quad (16.1.26a)$$

$$p_v^{(a)}(T) = n_i e^{-\frac{\mathcal{E}_F - \mathcal{E}_F^{(a)}}{k_B T}} \quad \text{for a p-type semiconductor} \quad (16.1.26b)$$

¹³Even at so low temperatures where the excitation of the donor atoms is negligible too there is still some conduction, because the wave function of the localized donor electrons overlap even for very small donor concentrations. This is called *impurity band conduction*

¹⁴C.f. (16.1.10).

Example 16.5. Determine the ratio of conduction electrons from P dopants in Si to the intrinsic electron concentration at the following temperatures: room temperature, 100°C and 500°C! Is it possible for the intrinsic electron concentration to become larger than the one due to the dopants? **Solution a)**

P is a donor atom, therefore the ratio of the conduction electrons from P and from the valence band can be calculated according to (16.1.26), using (16.1.24a) and neglecting the factor $\ln \left(\frac{m_c}{m_h} \right)$:

$$\begin{aligned} \frac{n_c^{(d)}(T)}{n_i(T)} &= e^{-\frac{\mathcal{E}_F^{(d)} - \mathcal{E}_F}{k_B T}} \\ &= e^{-\left(\frac{\mathcal{E}_d - \mathcal{E}_v}{2 k_B T} \right)} \\ &= e^{-\left(\frac{\mathcal{E}_d - \mathcal{E}_c + \mathcal{E}_g}{2 k_B T} \right)} \end{aligned}$$

The value of $\mathcal{E}_d - \mathcal{E}_c$ from Table 16.2 is -0.046 eV , $\mathcal{E}_g = 1.12 \text{ eV}$, and $k_B T$ at room temperature (300 K) equals to 0.0258 eV so

$$\frac{n_c^{(d)}(T)}{n_i(T)} = e^{-\left(\frac{1.12 - 0.046}{2 \cdot 0.0258} \right)} = \underline{\underline{1.09 \cdot 10^9}}$$

Similarly at 100°C (373K) and 500°C (773K)

$$\begin{aligned} \frac{n_c^{(d)}(373K)}{n_i(373K)} &= \underline{\underline{1.08 \cdot 10^7}} \\ \frac{n_c^{(d)}(773K)}{n_i(773K)} &= \underline{\underline{3.17 \cdot 10^3}} \end{aligned}$$

The conduction electron concentration in P doped Si at room temperature is $\approx 10^9$ times larger than the electron concentration in intrinsic Si, and about a thousand times as large even at the very high temperature of 737 K !

b)

From this formulas it seems that the intrinsic electron concentration may never reach $n_c^{(d)}$ and the two concentrations may never even be equal. But this is not so. We made some assumptions, which become invalid at higher temperatures. First we neglected the logarithmic terms when we calculated the Fermi energies and second we neglected the fact that the donor concentration is very small and the number of conduction electrons from donors has an upper limit, i.e. the electron density due to dopants has *saturation*. Therefore it is possible for the intrinsic electron concentration to exceed the electron concentration from dopants.

Important 16.1.3. *Because donor and acceptor levels are very close to the band edges it is much more easy to excite charge carriers from them to the near band. Therefore even at higher temperatures the overwhelming majority of the charge carriers (electrons in the conduction and holes in the valence band) come from donor or acceptor atoms.*

When the temperature increases the ionized portion of the ionized donor or acceptor levels will become larger. At the same time the number of conduction electrons that come from the valence band or the hole concentration that remains behind also increases. But because the dopants are almost all ionized even at very low temperatures the increase in charge carrier concentrations due to dopants is much smaller than that of the intrinsic charge carriers. This is the *saturation region*. Increasing the temperature further the intrinsic charge carrier concentration will exceed the one due to the dopants. This is the *intrinsic region*. As a consequence the Fermi energy will shift towards the intrinsic value as is depicted in Fig 16.7 for an n-type semiconductor.

As we mentioned in the previous section the law of mass action holds true even for extrinsic semiconductors. That is

$$n_c(T) \cdot p_v(T) = n_i^2(T) \quad (16.1.27)$$

This means that when we create an n-type semiconductor we not only increase the conduction electron concentration but decrease the valence hole concentration at the same time as well. A short qualitative explanation why it is so is given in Appendix 23.9.

16.2 Semiconductor structures. The p-n junction. Applications

Homogeneous intrinsic semiconductors are rare to find in practice. In every semiconductor device special inhomogeneous structures are used. These are created by locally varying the level of doping (number of donor and acceptor states) when the semiconductor device is fabricated. A short description of the fabrication process is in Appendix 23.10.

16.2.1 Inhomogeneous semiconductors. The (unbiased) p-n junction.

In inhomogeneous semiconductors the concentration of donors and acceptors is different at different parts of the material. The simplest such structure is the p-n junction when

a p-type and an n-type layer meets (see Fig. 16.8)

$$N_d(x) = \begin{cases} 0 & x \leq -l_n \\ 0 < N_d(x) < N_d & -l_n < x < l_p \\ N_d & x \geq 0 \end{cases} \quad (16.2.1)$$

$$P_v(x) = \begin{cases} N_a & x > 0 \\ 0 < N_a(x) < N_a & -l_n < x < l_p \\ 0 & x \leq 0 \end{cases} \quad (16.2.2)$$

Here l_n and l_p are the widths at the n- and p- sides of the boundary between the differently doped sides where the concentration of one of the dopant drops to 0 while the concentration of the other dopant increases up to its bulk level, called the *transition region*. Its width is $\approx 1 - 1000$ nm.

Imagine that the two types of semiconductors has just been connected. Because at the p-type side there are fewer electrons than holes and at the n-type side fewer holes than electrons a diffusion current of electrons and holes starts. As majority charge carriers move to the opposite side where they are minority carriers a $\varphi(x)$ potential arises between the n and p-type sides. After a short while this potential difference grow high enough to stop the diffusion current. There will be a region wider than the transition region where no movable charge carriers could be found, only unmovable negative and positive charges. The width of this region is $\approx 10-1000$ nm. This is called either the *space charge region* or the *depletion region*¹⁵. The calculation of the $n_c(x)$ and $p_v(x)$ concentrations is in Appendix 23.11.

If we simplify our task by setting l_p and l_n to 0 (i.e. abrupt change of doping at the boundary) the simple Poisson equation is

$$\frac{d\varphi(x)}{dx} = \begin{cases} 0 & x > d_n \\ -\frac{e N_d(x)}{\epsilon} & 0 < x < d_n \\ +\frac{e N_a(x)}{\epsilon} & -d_p < x < 0 \\ 0 & x < -d_p \end{cases} \quad (16.2.3)$$

The result with simple integration:

$$\varphi(x) = \begin{cases} \varphi(\infty) & x > d_n \\ \varphi(\infty) - \frac{e N_d}{2\epsilon}(x - d_n)^2 & 0 < x < d_n \\ \varphi(-\infty) + \frac{e N_a}{2\epsilon}(x + d_p)^2 & -d_p < x < 0 \\ 0 & x < -d_p \end{cases} \quad (16.2.4)$$

¹⁵Originally the $\mathcal{E}_{F,p}$ and $\mathcal{E}_{F,n}$ Fermi levels are different at the two sides, at the p-type region it will be closer to the valence band at the n-type region closer to the conduction band. After equilibrium is reached the Fermi level at both sides will be the same. C.f. contact potential

The boundary conditions that must be satisfied are the continuity of $\varphi(x)$ and its first derivative are explicitly obeyed by these equation. If we write them for $x = 0$ we can determine the values of d_n and d_p ¹⁶.

$$\begin{aligned} d_n &= \sqrt{\frac{N_a}{N_d(N_d + N_a)} \frac{\epsilon \Delta \varphi}{2e}} \\ d_p &= \sqrt{\frac{N_d}{N_a(N_d + N_a)} \frac{\epsilon \Delta \varphi}{2e}} \end{aligned} \quad (16.2.5)$$

Example 16.6. Calculate the total voltage difference (the built in potential) between the *n*-type and *p*-type part for a uniformly doped Silicon *p-n* junction with $N_d = N_a = 10^{17} \text{ cm}^{-3}$ at room temperature. The intrinsic carrier density is $1.45 \cdot 10^{14} \text{ m}^{-3}$ Will the built-in voltage increase or decrease with an increase in temperature? **Solution From formula (23.11.4)**

$$V_{p-n} \equiv \Delta \varphi = \frac{1}{e} k_B T \ln \left(\frac{N_d N_a}{n_i^2} \right) \quad (16.2.6)$$

$$\begin{aligned} V_{p-n} &= 1.38 \cdot 10^{-23} \left[\frac{J}{K} \right] 300 [K] \frac{1}{1.6022 \cdot 10^{-19} [As]} \ln \left(\frac{10^{21} 10^{21}}{(1.45 \cdot 10^{14})^2} \right) \\ &= 0.82V \end{aligned}$$

Substituting back the expression (16.1.17) of $n_i^2(T)$ and (16.1.14) we find that V_{p-n} is of the form:

$$e V_{p-n} = \text{const}_1 k_B T - \text{const}_2 k_B T \ln T + \mathcal{E}_g$$

When T increases the change in the term containing $-k_B T \ln T$ is larger than the change in the term containing $k_B T \rightarrow$ the voltage decreases.

Example 16.7. Determine the widths d_n and d_p and the electrical field strength for the Si *p-n* junction of the previous example. The relative permittivity of Si is $\epsilon_r = 16.0$.

¹⁶Continuity of $\varphi'(x)|_{x=0}$ gives

$$N_d d_n = N_a d_p$$

i.e. the excess positive charge on the *n*-side of the junction is the same as the excess negative charge on the *p*-side. From the continuity of $\varphi(x)$ at $x = 0$:

$$\frac{e}{2\epsilon} (N_d d_n^2 + N_a d_p^2) = \Delta \varphi$$

From these two equations d_n and d_p can be determined.

Solution We may write (16.2.5) in a more convenient form:

$$\begin{aligned} d_n &= 5257 \sqrt{\frac{N_a}{N_d(N_d + N_a)} \frac{\epsilon_r \Delta\varphi}{2}} [nm] \\ d_p &= 5257 \sqrt{\frac{N_d}{N_a(N_d + N_a)} \frac{\epsilon_r \Delta\varphi}{2}} [nm] \end{aligned} \quad (16.2.7)$$

Substituting the data from the previous example gives $d_n = d_p = 425.8$ nm. The magnitude of E is $\Delta\varphi/(d_n + d_p) = 0.82V/4.258 \cdot 10^{-7} \text{ m} = 1.93 \cdot 10^7 \text{ V/m}$.

The Fermi energy is the (electro)chemical potential for a semiconductor. In inhomogeneous semiconductors we can define a position dependent chemical potential by

$$\mu_e(x) = \mathcal{E}_F + e\varphi(x) \quad (16.2.8)$$

With this (23.11.1) can be written in the form:

$$\begin{aligned} n_c(x) &= N_c(T) e^{-(\mathcal{E}_c - \mu_e(x))/k_B T} \\ p_v(x) &= P_v(T) e^{-(\mu_e(x) - E_v)/k_B T} \end{aligned} \quad (16.2.9)$$

These are precisely the form of relations (16.1.10) for homogeneous semiconductors, except that \mathcal{E}_F is replaced by the position dependent electrochemical potential. Therefore the p-n junction may be described by either having constant band and impurity energies and a position dependent $\mu_e(x)$ electrochemical potential¹⁷ or by having position dependent bands and impurity energies and a constant electrochemical potential (Fig. 16.10).

16.2.2 The biased p-n junction.

Things become really interesting when an external voltage (bias) is applied across the p-n junction. We shall take V positive (*forward bias*) if its application raises the potential of the p -side with respect to the n -side, in the opposite case it is negative (*reverse bias*). When $V = 0$ as above there is a depletion layer of 10-100 nm in extent about the transition point where the doping changes from n -type to p -type. Because of the lack of carriers this layer has a much higher electrical resistance than the homogeneous regions. Most of the voltage drop will occur in this region:

$$V = V_{homog} + V_{depl} \approx V_{depl}$$

¹⁷Even though $\mu_e(x)$ is not equivalent with \mathcal{E}_F it is sometimes called the Fermi energy and denoted by $\mathcal{E}_F(x)$.

This modifies the total potential difference:

$$\Delta\varphi = \Delta\varphi_0 - V \quad (16.2.10)$$

where $\Delta\varphi_0$ is the potential difference in (16.2.6). This change in $\Delta\varphi$ changes the values of d_n and d_p according to (16.2.5):

$$\begin{aligned} d_n(V) &= d_n|_{V=0} \cdot \left(1 - \frac{V}{\Delta\varphi_0}\right) \\ d_p(V) &= d_p|_{V=0} \cdot \left(1 - \frac{V}{\Delta\varphi_0}\right) \end{aligned} \quad (16.2.11)$$

When $V = 0$ no current flows through the junction. When $V \neq 0$ electron and hole currents of the same values will flow. The total current is the sum of the electron and hole currents. It follows it is sufficient to deal with only one of them. Let us consider the current of the holes! It has two components:

- *Generation current* (j_{hole}^{gen})
Holes are continuously generated on the n-side by thermal excitation of electrons to the conduction band. Although these are minority carriers there, they still have an important role in the total current. Any hole generated near the junction may wander into it then it is swept over to the p-side by the strong electric field in the depletion layer. This current is insensitive to the magnitude of V .
- *Recombination current* ($j_{hole}^{rec}(V)$)
A hole current flows from the p-side to the n-side. But in that direction the potential change presents a barrier to the holes. Only holes which have high enough thermal energy may go in that direction. This current depends on V exponentially with the proportionality constant C :

$$j_{hole}^{rec}(V) = C e^{-e(\Delta\varphi_0 - V)/k_B T} \quad (16.2.12)$$

The total current $j_{hole}^{tot}(V)$ is the difference of the recombination and generation currents:

$$j_{hole}^{tot}(V) = j_{hole}^{rec}(V) - j_{hole}^{gen} \quad (16.2.13)$$

When $V = 0$ the total current is 0 and the generation and recombination current must be equal

$$j_{hole}^{tot}(0) = 0 = j_{hole}^{rec}(0) - j_{hole}^{gen} \quad \text{and} \quad j_{hole}^{rec}(0) = C e^{-\Delta\varphi/k_B T}$$

From this C can be determined:

$$C = j_{hole}^{gen} e^{\Delta\varphi/k_B T}$$

Consequently the factor $e^{-e\Delta\varphi_0/k_BT}$ is cancelled, therefore

$$j_{hole}^{tot}(V) = j_{hole}^{gen} (e^{V/k_BT} - 1) \quad (16.2.14)$$

$$\begin{aligned} j^{tot}(V) &= j_h^{tot}(V) + j_e^{tot}(V) \\ j^{tot}(V) &= (j_h^{gen} + j_e^{gen}) (e^{V/k_BT} - 1) \end{aligned} \quad (16.2.15)$$

The current exponentially depends on V as seen in Fig. 16.12.

Diode

This is the characteristics of a *rectifier* or diode. When forward bias is applied the current flows freely while with reverse bias the current is very small. Putting a semiconductor diode into an electric circuit where alternating current flows effectively allows current to flow only when the direction of the instantaneous voltage corresponds to the forward bias. It is worth to note that too high voltages damage the p-n junction.

16.2.3 Transistors

The Bipolar Junction Transistor (BJT)

The invention of the Bipolar Junction Transistor in 1948 turned the fate of electronics, which up to that time only used relatively large, fragile *vacuum tubes*.

The BJT is a 3 terminal (**B**ase, **E**mitter, **C**ollector) electronic device. The “Bipolar” in the name implies that in a BJT both types of charge carriers are used. They come in two flavors: there are PNP and the NPN transistors, where the letters refer to the types of the three regions of the device. In a PNP transistors the holes, in an NPN transistor the electrons are the majority carriers. As the mobility of electrons is usually larger than that of the holes NPN transistors are faster devices than PNP transistors.

In typical operation, the base-emitter junction is forward biased and the base-collector junction is reverse biased. The collector-emitter current (*collector current* in short) is controlled by the much smaller base-emitter current (*base current* in short), but can also be viewed as controlled by the base-emitter voltage (voltage control). These views are related to the current-voltage relation of the base-emitter junction, which is just the usual exponential current-voltage curve of a p-n junction (diode). The collector current flows through only the collector and the base current only through the base electrode, while according to Kirchhoff’s law the sum of these must flow through the emitter. It may seem that BJTs can be considered as two diodes with a shared anode.

This is true only when just two electrodes are used¹⁸ However when voltage applied between both B and E and B and C the behavior is different.

Background 16.2.1. *In an NPN transistor, for example, when a positive voltage (forward bias) is applied to the base–emitter junction, the equilibrium between thermally generated carriers and the repelling electric field of the depletion region becomes unbalanced, allowing thermally excited electrons to enter the base region (and a smaller current of holes flows from base to emitter). The base region is narrow therefore just a fraction of these electrons can recombine in it or reach the base electrode. Although there is a reverse bias between the base and the collector most of the electrons wander (or "diffuse") through the base into the collector, which means the base–collector junction does not operate like a diode.*

Without the forward bias no current can flow between the base–emitter diode, therefore no current will flow from the emitter to the collector.

The electrons in the base are only called minority carriers because the base is doped p-type which would make holes the majority carrier in the base.

To minimize the percentage of carriers that recombine before reaching the collector–base junction, the transistor’s base region must be thin enough that carriers can diffuse across it in much less time than the semiconductor’s minority carrier lifetime. In particular, the thickness of the base must be much less than the diffusion length of the electrons. The collector–base junction is reverse–biased, and so little electron injection occurs from the collector to the base, but electrons that diffuse through the base towards the collector are swept into the collector by the electric field in the depletion region of the collector–base junction. The thin shared base and asymmetric collector–emitter doping is what makes a bipolar transistor different from two separate and oppositely biased diodes connected in series.

The Field Effect Transistor (FET)

FETs has 3 terminal connectors¹⁹:

Source (S) through which the current I_S enter the device. This is connected to a heavily doped region.

Drain (D) , through which the current I_D leave the device. This is also connected to a heavily doped region.

¹⁸When the first portable “transistor radios” become available the number of transistors was the main selling point, therefore many radio contained a surplus number of transistors used as diodes.

¹⁹Most FETs also have a fourth terminal called the body, base, bulk, or substrate. This fourth terminal serves to bias the transistor into operation; it is rare to make non-trivial use of the body terminal in circuit designs, but its presence is important when setting up the physical layout of an integrated circuit. In discreet FETs the bulk terminal usually is connected to the source.

Gate (G) , the terminal whose voltage relative to the source controls the I_D current. The voltage is related to the strength of the electric field between the terminals thus the name “Field Effect” transistor.

In FETs only one kind of charge carriers are used. It is a *unipolar* device. The current that flows from *source* to the *drain* is the current of majority charge carriers, electrons or holes. It is flowing through an active channel induced in the doped substrate by the voltage between the *gate* and the source.

There are many kinds of FETs. The substrate of a FET is doped to produce either an n-type or a p-type semiconductor. The drain and source may be doped of opposite type to the channel (*depletion mode FET*), or doped of similar type to the channel (*enhancement mode FET*), but both source and drain must be doped with the same type of dopant and more heavily than the substrate. Field-effect transistors are also distinguished by the method of insulation between channel and gate.

The most widely used FET type is the metal-oxide-semiconductor field-effect transistor and we will only describe it here. The structure of a MOSFET is in Fig 16.14.

MOSFETs are based on the modulation of charge concentration by a MOS capacitance between the source electrode and the gate electrode²⁰, located above the body and insulated from all other device regions by a gate dielectric layer which in the case of a MOSFET is an oxide, such as silicon dioxide.

Background 16.2.2. *A traditional metal-oxide-semiconductor (MOS) structure is obtained by growing a thin (fraction of a micron) layer of silicon dioxide (SiO_2) on top of a silicon substrate and depositing a layer of metal or polycrystalline silicon (the latter is commonly used). As the silicon dioxide is a dielectric material, its structure is equivalent to a planar capacitor, with one of the electrodes replaced by a semiconductor.*

When a voltage is applied across a MOS structure, it modifies the distribution of charges in the semiconductor. In the structure in Fig 16.14 a positive voltage, V_{GS} , from gate to body creates a depletion layer by forcing the positively charged holes away from the gate-insulator/semiconductor interface, leaving exposed a carrier-free region of immobile, negatively charged acceptor ions. As V_{GS} increases, hole concentration decreases, and the region near gate behaves progressively more like intrinsic semiconductor material as the excess hole concentration is zero.

If V_{GS} is higher than a V_{th} threshold, electrons from the heavily doped source and drain regions enter this region. The high concentration of negative charge carriers form a thin *inversion layer* located next to the interface between the semiconductor and the insulator. This inversion layer serves as the channel. The thickness of this channel is controlled by the applied V_{GS} (or more accurately by $V_{GS} - V_{th}$).

In an n-channel depletion-mode device, a negative gate-to-source voltage causes a depletion region to expand in width and squeezes the channel from the sides, narrowing

²⁰Source and Body electrodes are usually connected.

it. If the depletion region expands to completely close the channel, the resistance of the channel from source to drain becomes large, and the FET is effectively turned off like a switch. Likewise a positive gate-to-source voltage increases the channel size and allows electrons to flow easily.

Advantages of FET

The main advantage of the FET is its high input resistance, on the order of 100M ohms or more. Thus, it is a voltage-controlled device, and shows a high degree of isolation between input and output. It is a unipolar device, depending only upon majority current flow. It is less noisy and is thus found in FM tuners for quiet reception. It is relatively immune to radiation. It exhibits no offset voltage at zero drain current and hence makes an excellent signal chopper. It typically has better thermal stability than a Bipolar Junction Transistor (BJT).

Disadvantages of FET

It has relatively low gain-bandwidth product compared to a BJT. The MOSFET has a drawback of being very susceptible to overload voltages, thus requiring special handling during installation.

16.3 Metal–semiconductor junctions

A metal–semiconductor (M–S) junction is a type of junction in which a metal comes in close contact with a semiconductor material. It is the oldest practical semiconductor device. M–S junctions can either be rectifying or non-rectifying. The rectifying metal–semiconductor junction forms a Schottky barrier, this is used in a device known as the *Schottky diode*, while the non-rectifying junction is called an *ohmic contact*.

Rectifying (Schottky) junction

When a metal is contacted with an n-type semiconductor where the $\mathcal{E}_{F,s}$ Fermi energy is larger than $\mathcal{E}_{F,m}$ in the metal, electrons move into the metal thus creating a depletion layer in the semiconductor and thus giving rise to a potential barrier of $e\phi_m - e\phi_s$. Applying a V voltage may increase or decrease this barrier: $\phi^{tot} = \phi_m - \phi_s - V$. The probability of an electron to go through this barrier $\propto e^{-\Delta\mathcal{E}/k_B T}$. Therefore the current is small when reverse bias is applied, in which case the current is dominated by electron flow from the metal to the semiconductor and increasing exponentially with forward bias when the current is dominated by electron flow from the semiconductor to the metal.

Therefore like a p-n junction Schottky barriers too are rectifying junctions. Table 16.3 compares the properties of a p-n and a Schottky junction.

p-n junction	Schottky junction
Reverse current due to minority carriers diffusing to the depletion layer leads to strong temperature dependence	Reverse current due to majority carriers that overcome the barrier leads to less temperature dependence
Forward current due to minority carrier injection from n- and p-sides	Forward current due to majority injection from the semiconductor
Forward bias needed to make the device conducting (called <i>cut-in</i> or <i>knee voltage</i>) is large	The cut-in voltage is quite small
Switching speed controlled by recombination (elimination) of minority injected carriers	Switching speed controlled by thermalization of "hot" injected electrons across the barrier \sim few picoseconds
Recombination in depletion region	Essentially no recombination in depletion region

Table 16.3: Comparison of the rectifying properties of a p-n junction and a metal—semiconductor junction.

Ohmic contact

An ohmic contact is a contact between a semiconductor and a metal is a contact whose resistance is voltage independent. Such a contact can be the result of a negative or zero Schottky barrier height or of heavy doping. Frequently the creation of ohmic contacts includes a high temperature step which causes the deposited metals to form an alloy with the semiconductor or the high temperature anneal reduces the barrier height at the interface.

- Heavy doping (N_d or $N_a \sim 10^{24}m^{-3}$) in the semiconductor causes a very thin depletion width and electrons can tunnel across this barrier leading to ohmic behavior.
- Diffusion of metals into the semiconductor creates continuous concentration change leading to ohmic behavior.

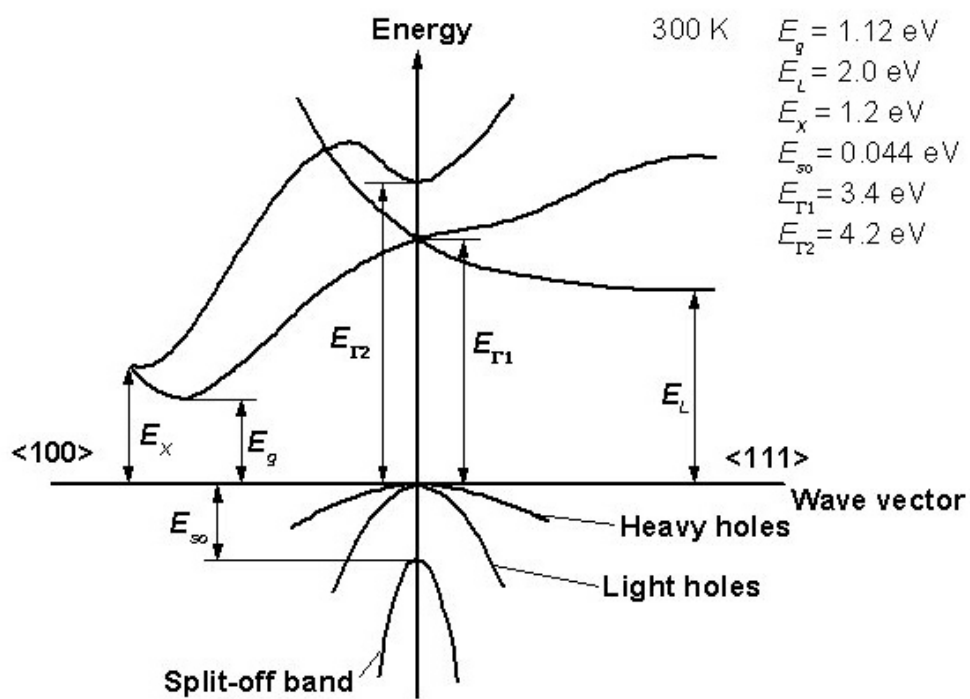


Figure 16.1: Band structure of Si in the $\langle 100 \rangle$ and $\langle 111 \rangle$ directions. Observe that the band minimum at $k = 0$ is not the lowest one.

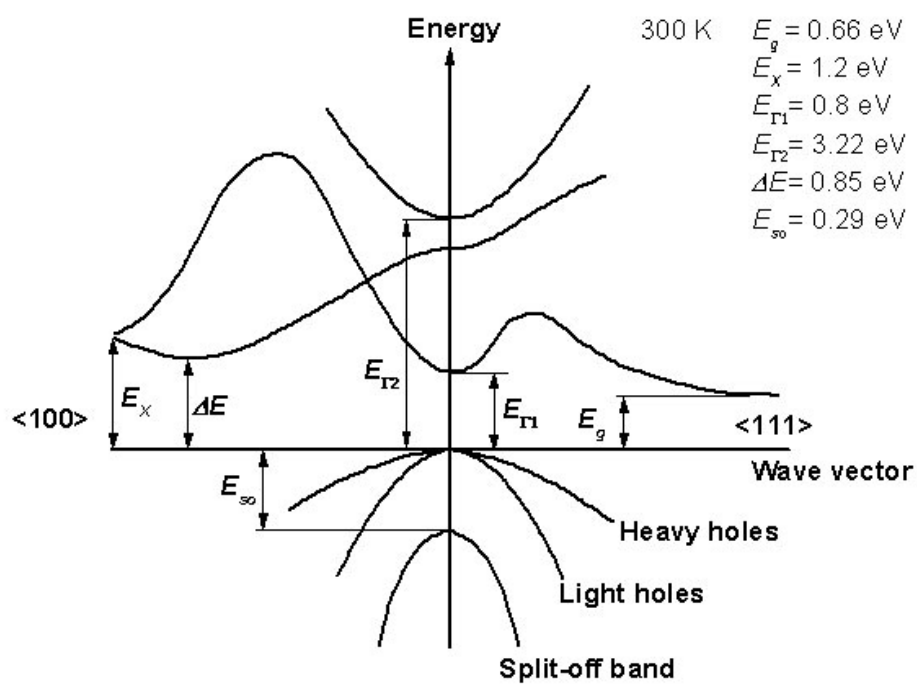


Figure 16.2: Band structure of Ge in the $\langle 100 \rangle$ and $\langle 111 \rangle$ directions.

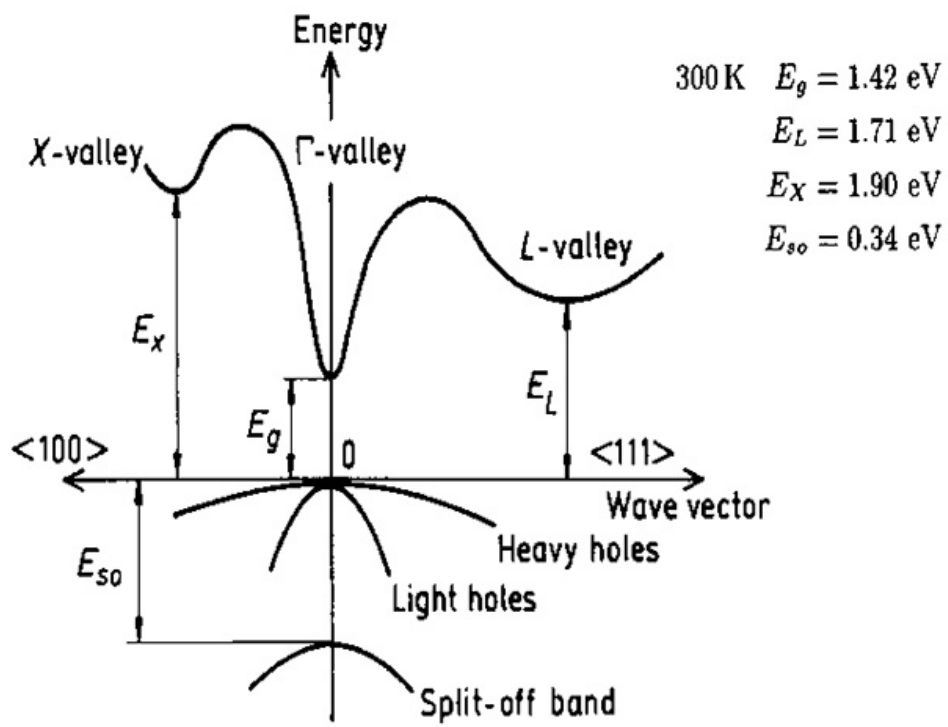


Figure 16.3: Band structure of gallium arsenide in the $\langle 100 \rangle$ and $\langle 111 \rangle$ directions.

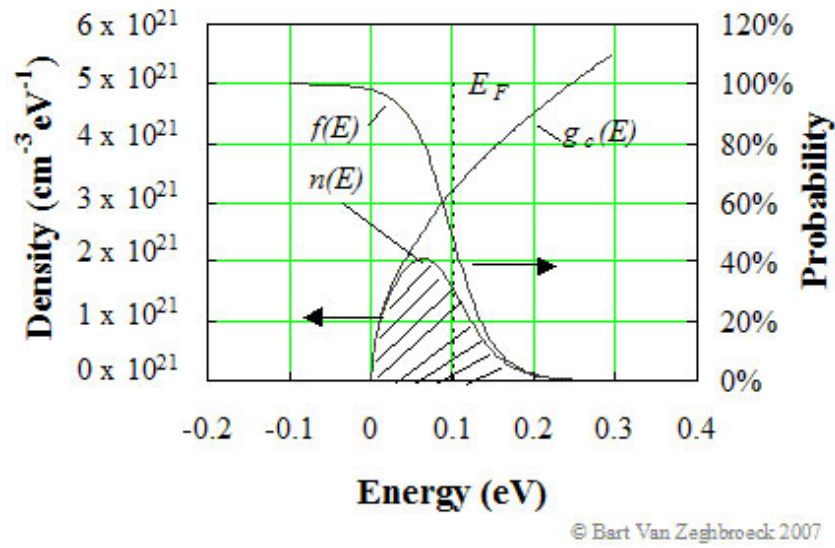


Figure 16.4: The carrier density integral. Shown are the density of states, $g_c(E)$, the density per unit energy, $n(E)$, and the probability of occupancy, $f_{FD}(E)$. The carrier density, n_o , equals the crosshatched area.

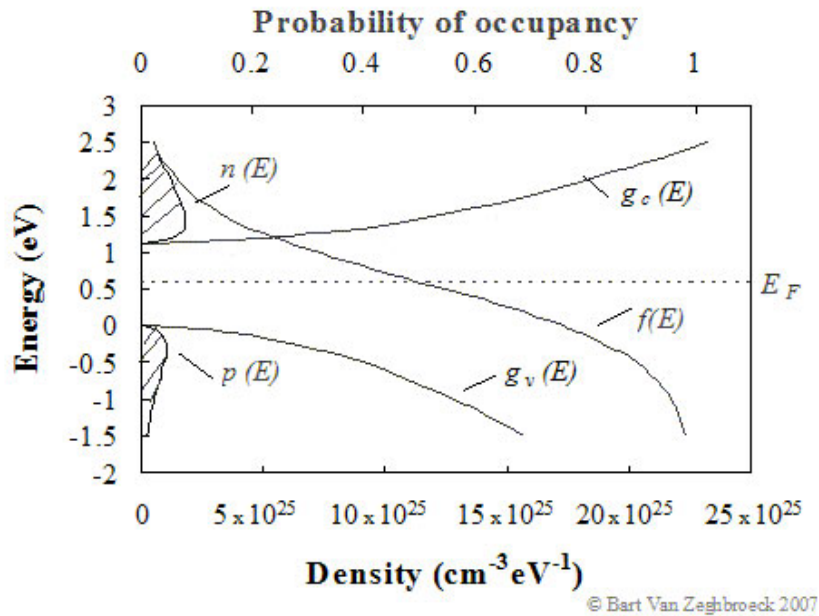


Figure 16.5: The density of states and carrier densities in the conduction and valence band. The crosshatched area indicates the electron and hole densities.

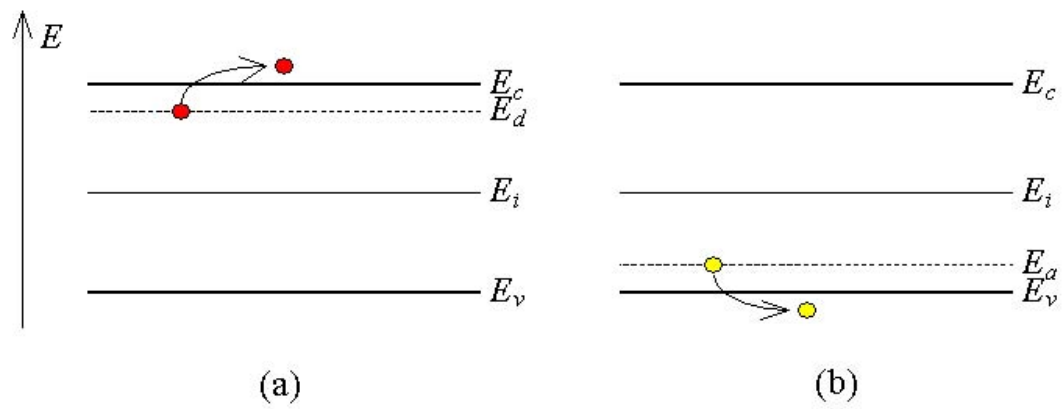


Figure 16.6: Levels and ionization of a) a shallow donor and b) a shallow acceptor

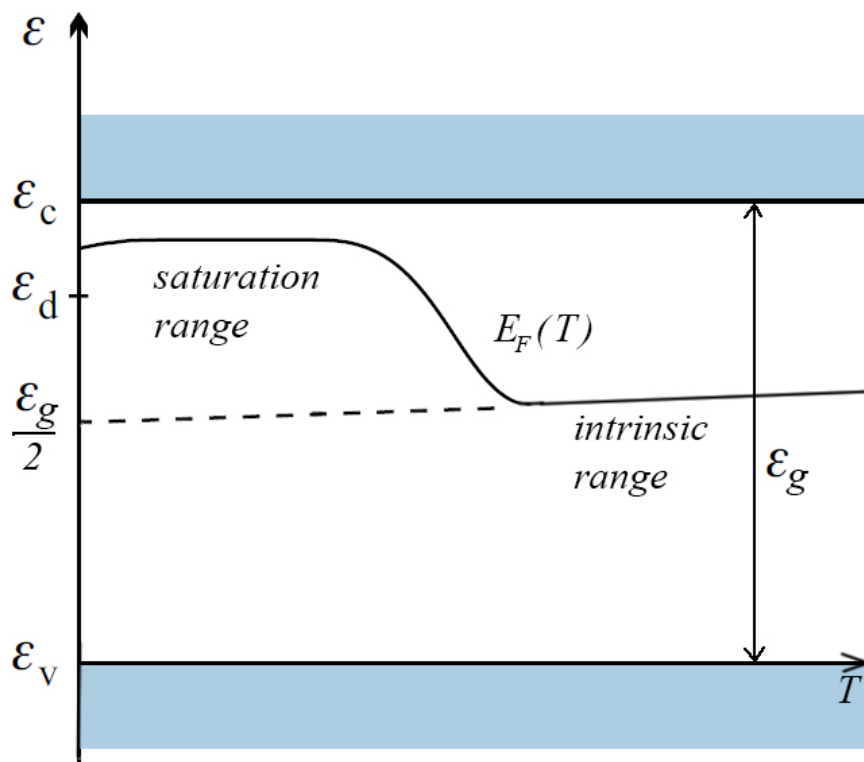


Figure 16.7: Schematic view of the Fermi level in an n-type semiconductor.

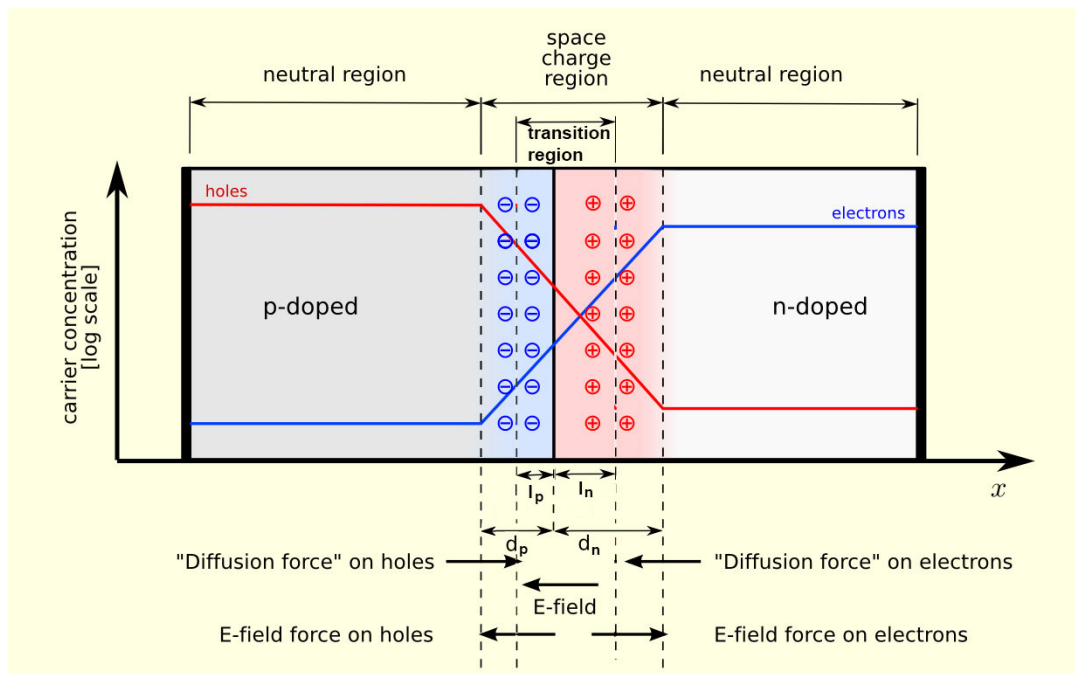


Figure 16.8: The p-n junction without an external voltage.

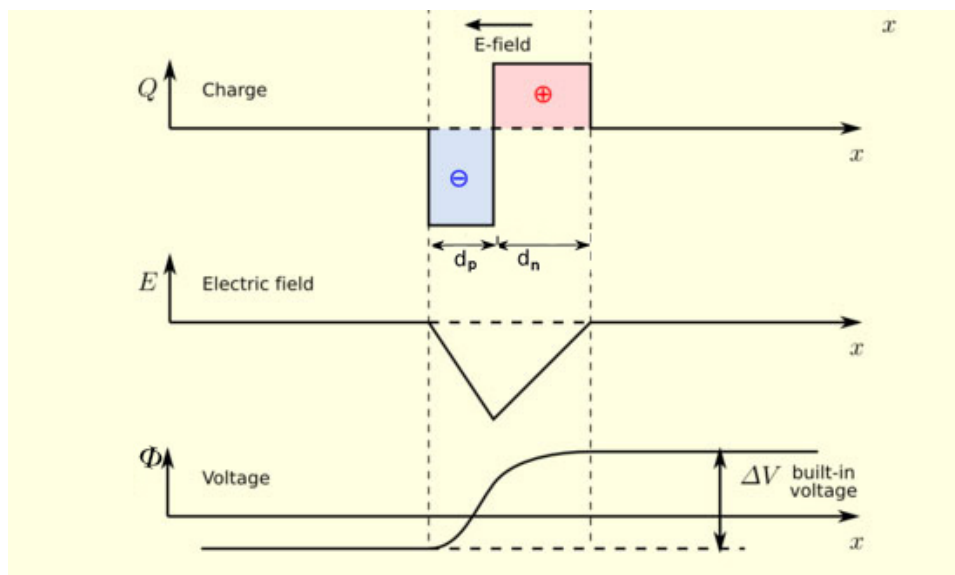


Figure 16.9: Charge density, electric field, and voltage in a p-n junction in thermal equilibrium with zero-bias voltage applied.

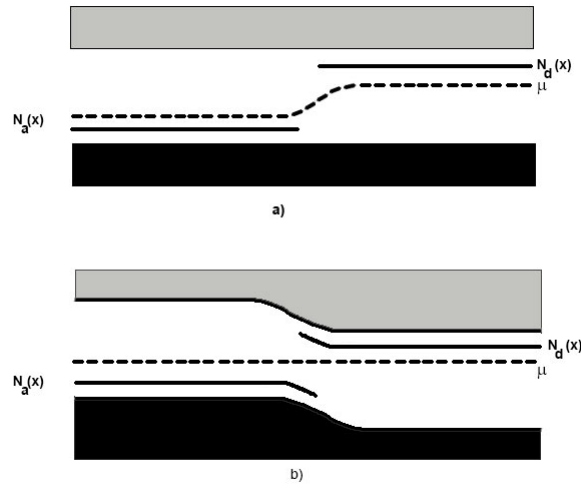


Figure 16.10: Two equivalent ways to describe the p-n junction.

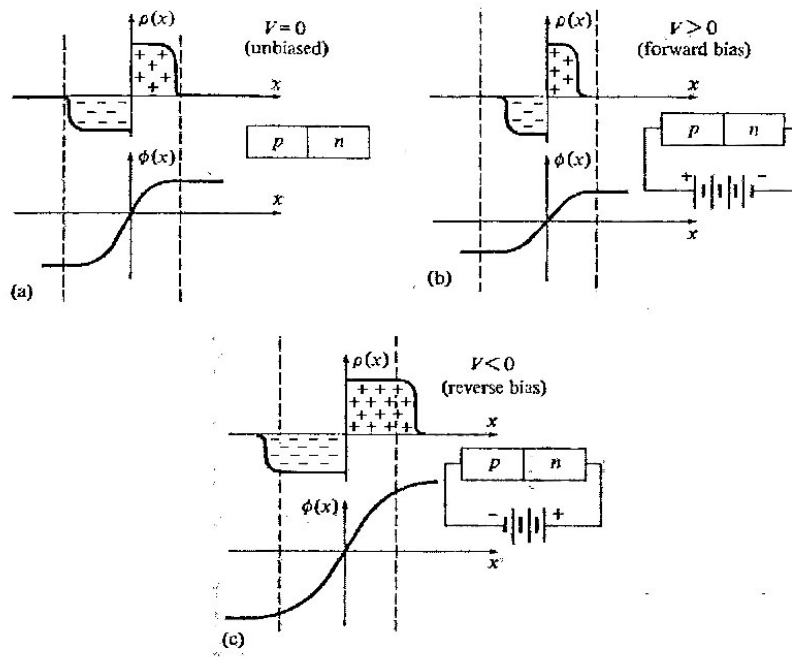


Figure 16.11: Charge density ρ and potential φ for a) unbiased, b) forward biased and c) reverse biased p-n junction

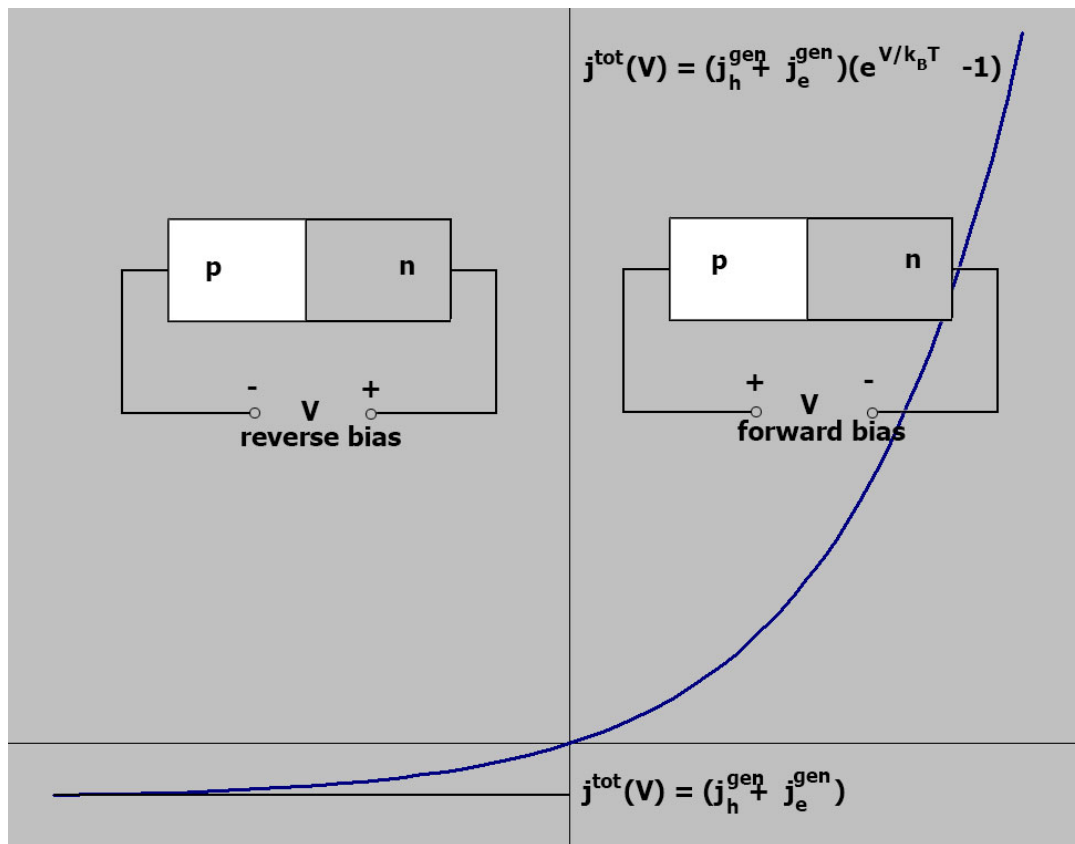


Figure 16.12: Current vs applied voltage for a p-n junction.

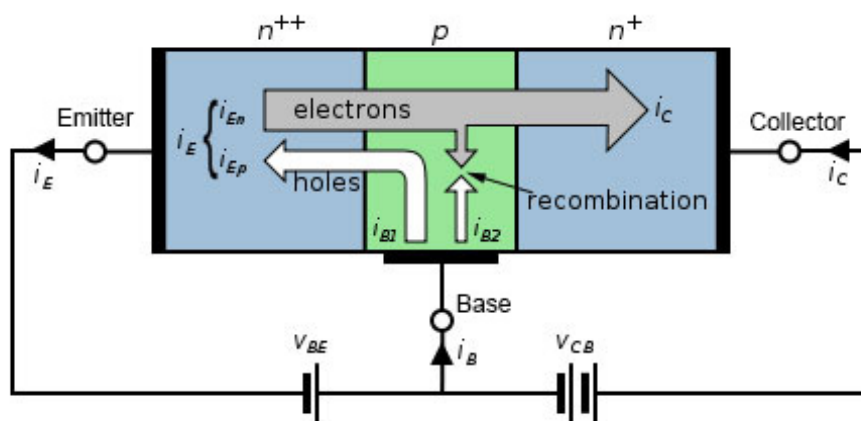


Figure 16.13: Schematics of an NPN Bipolar Junction Transistor

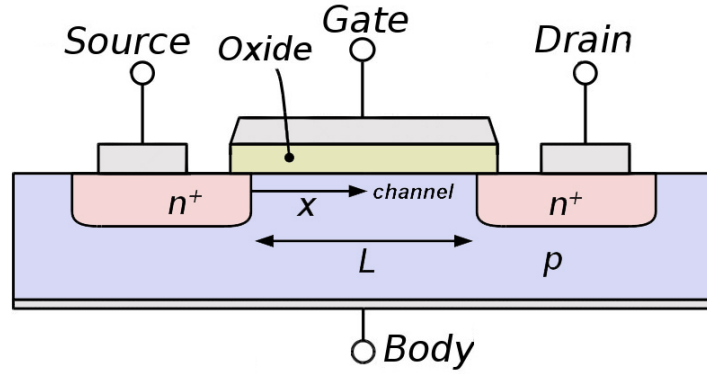


Figure 16.14: The structure of an n type Metal-Oxid-Semiconductor FET (MOSFET). Note that although in this figure the source and drain electrodes are symmetrical in electrical circuits FETs must be connected correctly.

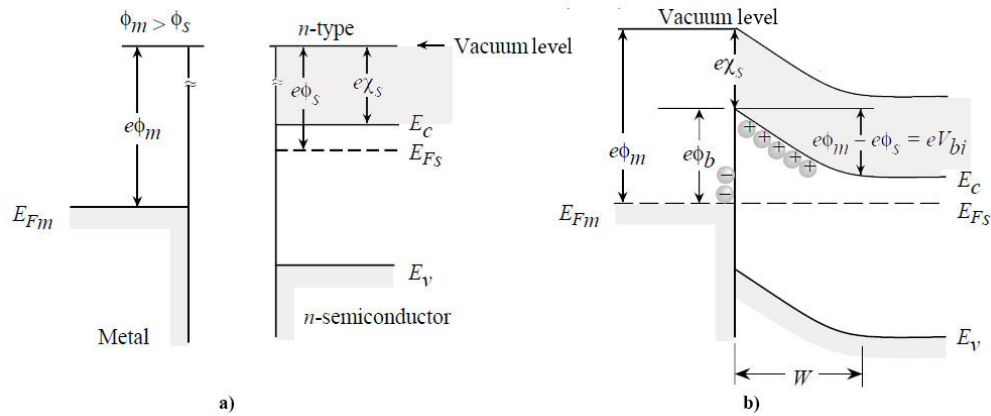


Figure 16.15: Band profiles for a) unconnected metal and an n-type semiconductor, b) Schottky junction.

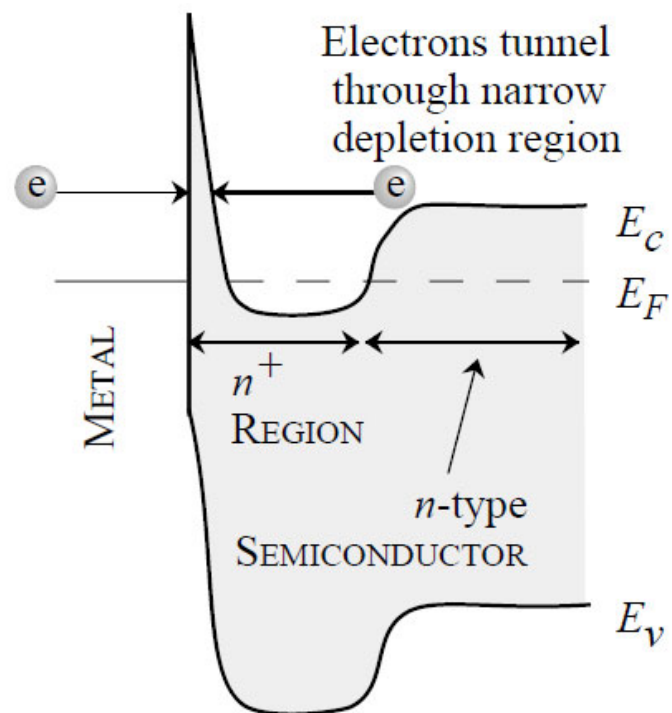


Figure 16.16: Ohmic contact between a metal and a semiconductor

Chapter 17

Superconductivity

17.1 Conductivity revisited. Superconductivity. Type I and II superconductors. High temperature superconductors

Superconductivity was discovered in 1911 by H. Kamerlingh Onnes who successfully liquified helium in 1908, when he investigated the low temperature resistivity of mercury in 1911. The reason he selected mercury was that it could be made very pure by distillation, and the resistivity at low temperatures are dominated by impurity effects as we stated above. He expected the resistivity to smoothly tend to zero which it would reach at 0 K, but to his surprise he found that mercury went through a *phase transition* and its resistance suddenly dropped to zero at 4.2K instead. This phenomenon is called *superconductivity*. The temperature at which it occurs is called the *critical temperature* (T_c).

In the Bloch model the Bloch wave function already incorporates the effect of the periodic potential and lattice vibrations are not considered, so Bloch electrons move freely in the crystal with a constant \mathbf{k} momentum *at any temperature*. When an external field acts on the system of electrons it will distort the Fermi sphere and \mathbf{k} of all electrons will increase, but still the electrons will move without any scattering¹. But in real metals the conductivity is not infinite. Even ideal metal crystals has a finite conductivity at $T > 0 K$ temperature.

An assumption of the Bloch model was that ion cores are at their equilibrium position in a perfect (infinite/periodic) crystal. Scattering of Bloch electrons may occur at *scattering centre* (described below) which are characterized by their *scattering cross section* σ_s ². The probability of scattering in a unit time interval is the inverse of the

¹C.f. Equation 15.2.6

²Don't confuse σ_s with the conductivity σ !

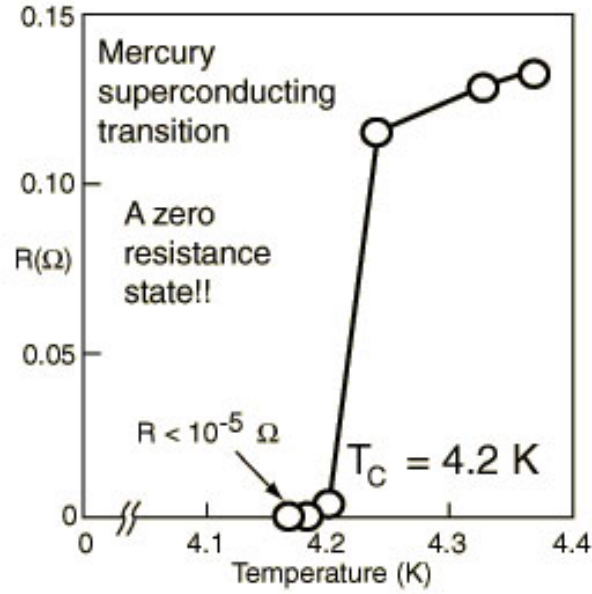


Figure 17.1: H. K. Onnes, Commun. Phys. Lab.12,120, (1911)

average scattering time (or relaxation time):

$$\frac{1}{\tau} = \sigma_s(v) n_s v_F \quad (17.1.1)$$

where the scattering cross section $\sigma_s(v)$ may depend on the velocity v , n_s is the density of the scattering centers, and v_F is the Fermi velocity, as only electrons near the Fermi level can take part in conduction. If more than one mutually exclusive scattering processes are possible the probability of scattering

$$\frac{1}{\tau} = \sum_i \sigma_{s,i} n_{s,i} v_F = \sum_i \frac{1}{\tau_i} \quad (17.1.2)$$

This is Mathiessen's rule.

The following scattering mechanisms are possible:

- Scattering on crystal defects characterized by the *scattering cross section* $\sigma_{s,def}$. This is independent of the temperature.
- Scattering on small amplitude lattice vibrations (i.e. far from the melting point) with $\sigma_{s,vib}$. This is proportional to the temperature.

The resulting resistivity as we show in Appendix 23.12 is

$$\rho = A + BT \quad (17.1.3)$$

Measured $\rho(T)$ curves are shown in Fig 17.2.

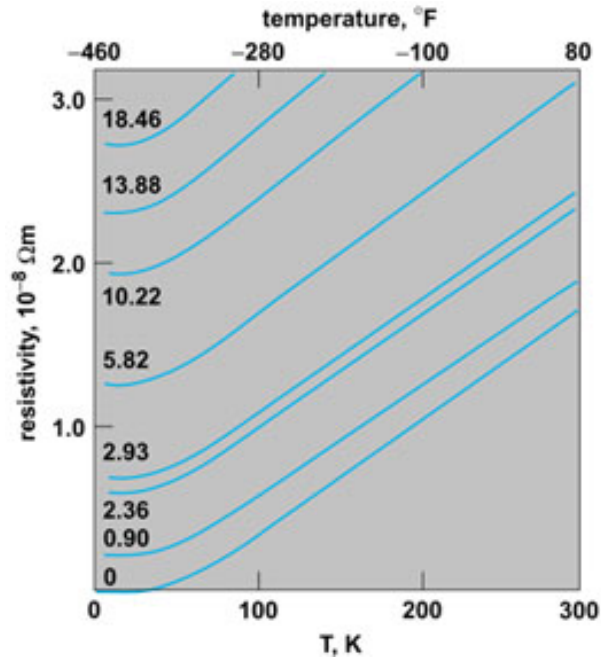


Figure 17.2: Resistivity of the disordered copper-zinc alloy system. Numbers on the curves give the concentration of zinc in atomic percent. (After W. E. Henry and P. A. Schroeder, The low-temperature resistivities and thermopowers of α -phase copper-zinc alloys, Can. J. Phys., 41:1076-1093, 1963)

Superconductors

Superconductors are materials displaying zero resistivity at finite temperatures. Superconductivity differs from the “simple” 0 resistivity state achieved at 0K in the Bloch model for all very pure and perfect crystals, because

- The material need not to be perfect crystal, crystal defects may be present in it.
- 0 resistivity appears at $T > 0K$ temperatures, where there are lattice vibrations. Usually $T_c < 30K$, but there are high temperature ($T_c \sim 100K$) superconductors too.

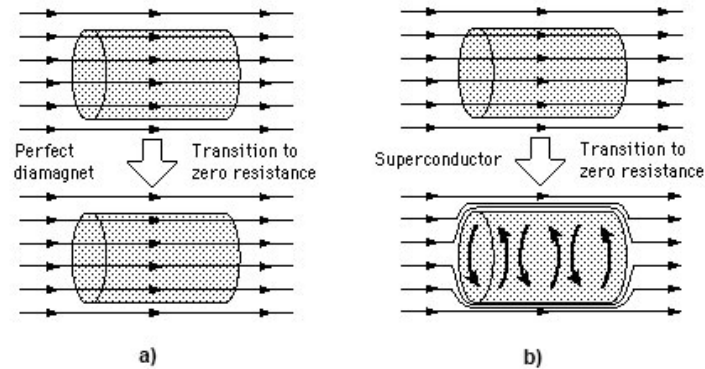


Figure 17.3: The Meissner effect in Type I superconductors. a) when a perfect diamagnet goes through a transition from finite to zero conductivity without becoming a superconductor, magnetic fields inside it remain untouched, b) the same transition for a superconductor induces currents inside the material which expels the magnetic field from the material.

- When a material makes the transition from the normal to superconducting state, it actively excludes magnetic fields from its interior. This is called the *Meissner effect*.

The Meissner effect

Every zero resistivity material shows perfect diamagnetism. Zero resistance implies that if you tried to magnetize the material, current loops would be generated to exactly cancel the imposed field (Lenz's law). If a conductor already had a steady magnetic field through it before this transition and was then cooled to a zero resistance state, during which it becomes a perfect diamagnet, the magnetic field would be expected to stay the same. A superconductor however actively excludes any magnetic field present when it makes the phase change to the superconducting state. This is depicted in Fig. 17.3.

If a small magnet is brought near a superconductor, it will always be repelled because induced *supercurrents*, i.e. superconducting currents that can flow without resistance, will produce mirror images of each pole. If a small permanent magnet is placed above a superconductor, it can be levitated by this repulsive force independent of its orientation at any height if the magnetic field of the magnet is strong enough at the superconductor.

Type I and II superconductors

When high enough external magnetic fields are applied the superconductive state vanishes depending on the type of the superconductor:

Type I there exists a single critical field B_c , above which all superconductivity is lost.

Type II there exist two critical fields $B_c^{(1)}, B_c^{(2)}$, between which they allow partial penetration of the magnetic field constrained in filaments within the material. These filaments are in the normal state, surrounded by *supercurrents* in what is called a vortex state. Such materials can be subjected to much higher external magnetic fields and still remain superconducting.

BCS theory of superconductivity

In 1972 the Physical Nobel Price was awarded to John Bardeen, Leon Cooper, and Robert Schrieffer for their successful model of Type I superconductors, what is now commonly called the BCS theory.

In the BCS theory a slight attractive force arises between electrons with opposite \mathbf{k} wave vector close to the Fermi level through interaction with the crystal lattice, which binds them in pairs. These pairs are called *Cooper pairs*. This attractive force is due to lattice vibrations that is the reason why the coupling to the lattice is called a phonon interaction.

In a material at $T > 0K$ whose all bands are either completely filled, or completely empty and there is no overlap between a filled and empty band or in any other material at $T = 0K$ no current may flow, because there are no empty energy levels for electrons to move under the influence of an external electric field. But Cooper has shown that if there exists an attractive interaction between electrons then the system of electrons becomes unstable against formation of bounded pairs of electrons of opposite \mathbf{k} values, now called *Cooper pairs*. This occurs, because the energy of these pairs is lower than the sum of the energy of the individual electrons.

Important 17.1.1. *The critical temperature of superconductors of different isotopes of the same atoms is found to be inversely proportional to the mass of the isotope used in the material. This hints that the cause of the attraction between electrons is related to lattice vibrations.*

In a solid the appearance of such an attractive interaction can be explained qualitatively using a simplified non-quantum physical argument:

When (classical) electrons are moving in a crystal they not only repel other electrons but also distort the lattice by attracting the ion cores. This small distortion creates a small

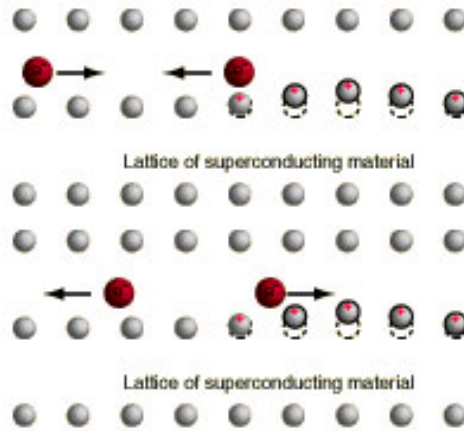


Figure 17.4: A visual model of the Cooper pair formation. A passing electron attracts the lattice, causing a slight ripple toward its path. Another electron passing in the opposite direction is attracted the resulting net positive charge. This creates a coupling between these two electrons.

net positive charge around the electron which attracts other electrons thus electron pairs may be formed. The magnitude of this interaction is about 10^{-3} eV which is equivalent to $T \sim 10$ K. Therefore at higher temperatures this attraction is easily suppressed by thermal vibrations.

But this model is incomplete and does not explain why only electrons with \mathbf{k} vectors of same magnitude and of opposite direction are affected.

The correct explanation requires quantum physics and too complicated to discuss here³. But regardless to the source of the attractive interaction the formation of pairs of electrons of opposite momenta means that the total energy of the pair will be smaller than the energy of two unpaired electrons. This creates a band gap of about 10^{-3} eV below \mathcal{E}_F . This effective energy gap in superconductors can be measured in microwave absorption experiments.

A band gap is implied by the very fact that the resistance is precisely zero. Charge carriers may only move through a crystal lattice without interacting with lattice vibrations and crystal defects if their energies are quantized and there are no available energy levels in the ranges needed for interactions.

The critical temperature for superconductivity must be a measure of the band gap, since the material could lose superconductivity if thermal energy could get charge carriers across the gap. The critical temperature depends upon isotopic mass. This supported that the superconducting transition involved some kind of interaction with the crystal lattice.

³It involves phonon transfer between electrons of the pair.

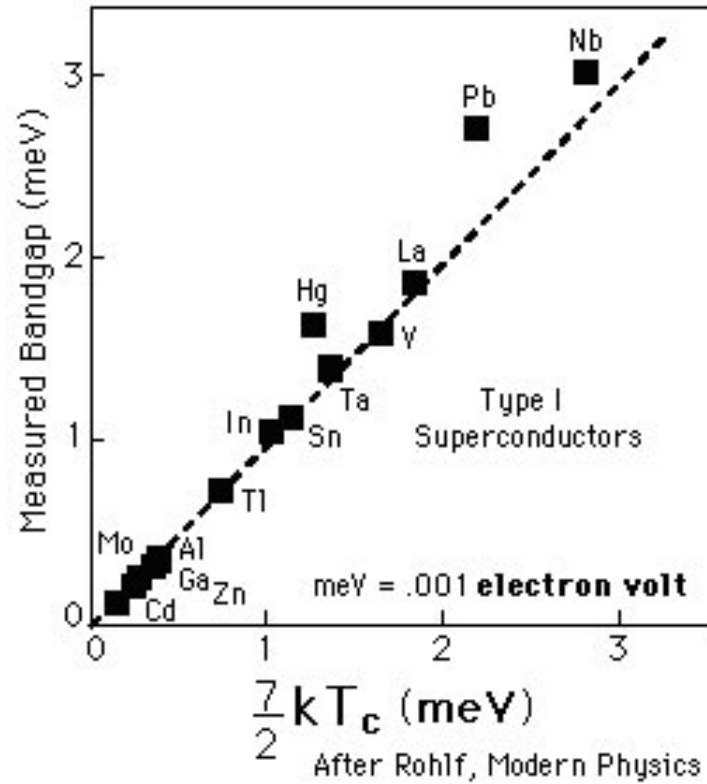


Figure 17.5: The measured bandgap in Type I superconductors. BCS prediction is $\mathcal{E}_g \sim 7/5k_B T_c$.

Although electrons (as every half spin particles) are fermions, the Cooper pairs they form have a total spin of 0 or 1 therefore behave like bosons. There are profound consequences of this fact, as no two electrons may be in the same quantum state, but any number of bosons can and will be.

Conduction in the superconducting state

According to the argument in Section 9.5 bosons in a system tend to *condensate* into the same quantum state. When no external electric field is applied the momentum of any Cooper pair will be 0 in which case their position is completely undetermined in the crystal. When we start a current by applying an electric field we force *all* Cooper pairs to move with the same \mathbf{k} vector. After switching off the field the Cooper pairs remain in motion, because they can only be slowed down collectively and at low temperatures there is not enough thermal energy to break the pairs.

The current in superconductors must flow in a thin layer of width Λ near the surface.

Λ is called the *London penetration depth*⁴ and can be calculated from

$$\Lambda = \sqrt{\frac{m_e c^2}{4\pi n_s(T) e^2}}$$

where $n_s(T)$ is the density of superconducting electrons.

High temperature superconductivity

Until 1986, physicists had believed that BCS theory, which forbade superconductivity at temperatures above about 30 K, is valid for all superconductors. But in 1986 Georg Bednorz and Karl Müller discovered *high-temperature superconductivity* in a lanthanum barium copper oxide. High temperature here means liquid nitrogen temperatures⁵ of about $100\text{K} = -173^\circ\text{C}$. This is good news, because liquid nitrogen can be produced cheaply on-site from air.

Many other cuprate superconductors have since been discovered, and the theory of superconductivity in these materials is one of the major outstanding challenges of theoretical condensed matter physics. In 2013 there is still no suitable theory which completely describes high temperature superconductivity.

Macroscopic quantum effects involving superconductors

Magnetic flux quantization

Place a ring from a superconductive material, not yet in the superconductive state, into a magnetic field. Cool it down below T_c . When it becomes superconducting the magnetic field is expelled from it by *supercurrents* that flow through the ring as we discussed. But a surprisingly unexpected fact is that the value of *the magnetic flux enclosed by the superconductive ring is quantized*. The flux quantum in SI units is

$$|\phi_B| = n \frac{h}{2e} = n \phi_0 = n 2.0679 \cdot 10^{-15} \text{Wb} \quad (17.1.4)$$

Such arrangement occurs in the normal state filaments of Type II superconductors which are subjected to a magnetic field between $H_{c,1}$, and $H_{c,2}$. The magnetic flux penetrates in discrete units while the bulk of the material remains superconducting.

⁴This can be deduced from the London equations of classical physics, named after F. and H London (1935)

⁵The boiling point of nitrogen is 77 K (-196°C) at atmospheric pressure

Josephson junction

Fig. 17.6 shows a structure called a *Josephson junction*. As you now know, because they are bosons, all the Cooper pairs in a superconductor can be described by a single wavefunction because all the pairs will have the same phase. This phase can be different in the two superconducting half of the junction. When the insulating layer is very thin Cooper pairs may tunnel through it without breaking up thus creating a continuous current. This is called *Josephson effect*⁶ and has four variants:

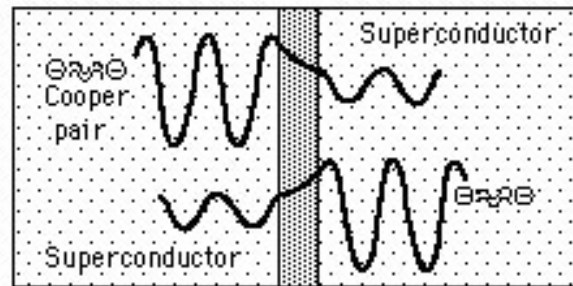


Figure 17.6: Schematics of a Josephson junction: two superconducting material separated by a thin insulating or normal state layer.

DC Josephson effect: *in the absence* of a voltage a current proportional to the sine of the phase difference across the insulator can flow through the junction.

AC Josephson effect: when a constant V voltage is applied the phase difference varies linearly in time and an alternating current flows through the junction with the frequency

$$\nu_{\text{Josephson}} = \frac{2e\Delta V}{h}$$

Because frequencies can be measured with great accuracy ($\sim 1/10^{-10}$) a Josephson junction device is the standard measure of voltage⁷.

The inverse AC Josephson effect: if the applied voltage is of the following form: $\phi(t) = \phi_0 + n\omega t + a\sin(\omega t)$, then for distinct DC voltages the junction carries a DC current and the junction behaves like an ideal frequency to voltage converter

⁶The Josephson effect was discovered in 1962 by Brian Josephson as a graduate student, while investigating what would happen if two superconducting metals were placed very close together without touching. He was awarded the Nobel price in physics for this discovery in 1973.

⁷The American National Institute of Standards and Technology (NIST) has produced a chip with 19000 series junctions to measure voltages on the order of 10 volts with this accuracy.

DC Josephson heat transfer: In 1965 it was proposed that the total heat flux through two parallel, paired Josephson junctions (SQUID - see below) whose half was heated up can be influenced by an applied magnetic field. This was only proved to be true in 2012.

Application of the Josephson effect includes:

- as the measure of voltage
- SQUIDs, or superconducting quantum interference devices (see below)
- "superconducting single-electron transistors"
- rapid single flux quantum (RSFQ) digital electronics
- Josephson junctions are integral in superconducting quantum computing as qubits
- Superconducting tunnel junction detectors (STJs) for use in astronomy and astrophysics in a few years.

Important 17.1.2. *Definition of the unit of voltage in SI used the Josephson effect between 1990 and 1997:*

The voltage on a Josephson junction is 1 V when the Josephson frequency is 483.6 GHz

Application of superconductors

- Superconducting magnets are some of the most powerful electromagnets known. They are used in MRI/NMR machines, mass spectrometers, MAGLEV trains⁸ and the beam-steering magnets used in particle accelerators (e.g. LHC)⁹
- They can also be used for magnetic separation, where weakly magnetic particles are extracted from a Background of less or non-magnetic particles, as in the pigment industries.
- The standard volt is now defined in terms of a Josephson junction oscillator.

⁸One, built in Japan in 2005, traveled at half the speed of sound.

⁹Type II superconductors such as niobium-tin and niobium-titanium are used to make the coil windings for superconducting magnets. These two materials can be fabricated into wires and can withstand high magnetic fields. Typical construction of the coils is to embed a large number of fine filaments (20 micrometers diameter) in a copper matrix. The solid copper gives mechanical stability and provides a path for the large currents in case the superconducting state is lost. These superconducting magnets must be cooled with liquid helium. Superconducting magnets can use solenoid geometries as do ordinary electromagnets. Most high energy accelerators now use superconducting magnets. The proton accelerator at Fermilab uses 774 superconducting magnets in a ring of circumference 6.2 kilometers.

- SQUIDs (superconducting quantum interference devices) are the most sensitive magnetometers known. There are DC and RF SQUIDs. RF SQUIDs use one Josephson junction, while a DC SQUID consists of two superconductors separated by thin insulating layers to form two parallel Josephson junctions. When a bias

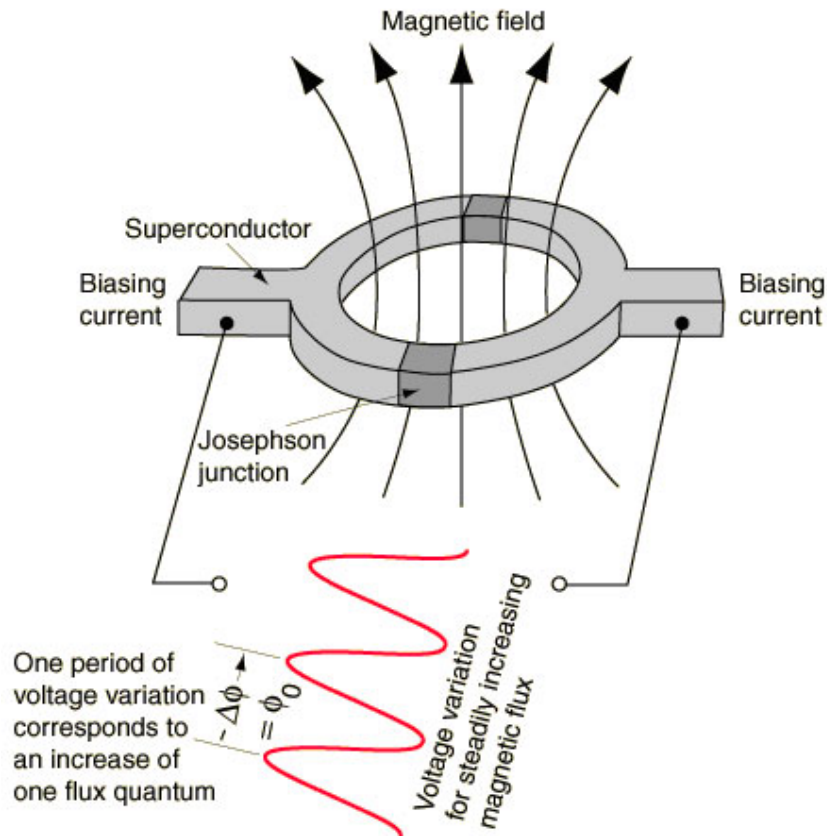


Figure 17.7: Workings of a DC SQUID.

current flows through the structure in the absence of an external field it will split equally at the two branches. But in a non zero external magnetic field the magnetic flux inside the loop must be an integer multiple of the flux quantum ϕ_0 . This is achieved by induced screening supercurrents in the loop as we discussed above, whose direction will depend on the magnitude of the external field, because if the flux from the external field is smaller than $\phi_0/2$ the screening current must cancel out the flux, while just above $\phi_0/2$ the screening current must increase it to get an integer multiple of ϕ_0 . Thus the screening current changes direction every time the external flux increases by half integer multiples of ϕ_0 . We therefore will measure an oscillating voltage across the loop.

The device may be configured as a magnetometer to detect incredibly small magnetic fields – small enough to measure the magnetic fields in living organisms. SQUIDs are used in scanning SQUID microscopes and magnetoencephalography¹⁰.

- Superconducting transmission lines (experimental): In prototype superconducting transmission lines at Brookhaven National Laboratory, 1000 MW of power can be transported within an enclosure of diameter 40 cm. This amounts to transporting the entire output of a large power plant on one enclosed transmission line.

¹⁰SQUIDs have been used to measure the magnetic fields in mouse brains to test whether there might be enough magnetism to attribute their navigational ability to an internal compass. Some data:

Threshold for SQUID	$10^{-14} T$
Human heart	$10^{-10} T$
Human brain	$10^{-13} T$

Chapter 18

Optical properties

18.1 Optical properties. X-ray emission and absorption.

18.1.1 X-ray emission

X-ray emission occurs during electronic transitions from an upper band to empty states in a core level band (very nearly atomic states). (C.f. Section 7.4.) The difference between atomic and interband X-ray transitions:

- Atomic: single line in emission spectrum (2 sharp atomic levels)
- Solids: emission band spectrum (source level can be any occupied level above E_g)

The shape of the emission curve is determined by the density of state function. For nearly free electrons this is well known:

$$I(\mathcal{E}) \propto \frac{dn(\mathcal{E})}{d\mathcal{E}} (= g(\mathcal{E})), \text{ and } g(\mathcal{E}) = \frac{8\pi\sqrt{2m_e^3}}{h^3} \sqrt{\mathcal{E}}$$

The edge of the spectra is at E_F , so measuring X-ray emission spectra is a way to measure E_F . Two types of X-ray emission spectroscopy is used¹: resonant inelastic X-ray emission spectroscopy (RIXS), in which the core electron is excited to a bound state in the conduction band and non-resonant X-ray emission spectroscopy (NXES), when the

¹Soft X-rays have different optical properties than visible light and therefore experiments must take place in ultra high vacuum, where the photon beam is manipulated using special mirrors and diffraction gratings. Gratings diffract X-ray photons of each wavelength present in the incoming radiation in a different direction, while the specific photon energy we wish to use to excite the sample with is selected by grating monochromators. Diffraction gratings are also used in the spectrometer to analyze the photon energy of the radiation emitted by the sample.

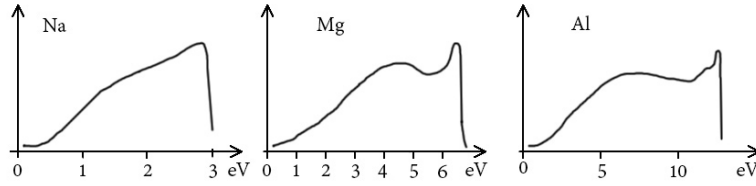


Figure 18.1: Schematic comparison of soft X-ray spectra of some metals. The cause of the peaks for Mg and Al is the band overlap: electrons may come from bands 3p and 3s.

core electron is excited to the continuum. Both involve the photonic promotion of a core level electron, and the measurement of the *fluorescence* (see below) that occurs as the electron relaxes into a lower-energy state.

Metal	Na	Mg	Al
$E_F[\text{eV}]$	3.12	7.3	11.9

Table 18.1: Fermi energies of some metals

18.1.2 X-ray absorption

The absorption of an X-ray photon by an atom results in the consequent emission of a photoelectron from the core level. The core hole created this way is filled in by an electron from another shell. The energy lost by this decaying electron either manifests itself in the emission of a fluorescent photon or by exciting a second outer shell electron (called an *Auger electron*) out of the atom². The directly emitted photoelectron, the fluorescent photon or the Auger electron (or any combination of these) are measured. The shape of the absorption spectra depends on the type of the solid and its chemical composition.

- Conductors with no band overlap:

²The latter is called the Auger effect, which is the base of Auger electron spectroscopy (AES).

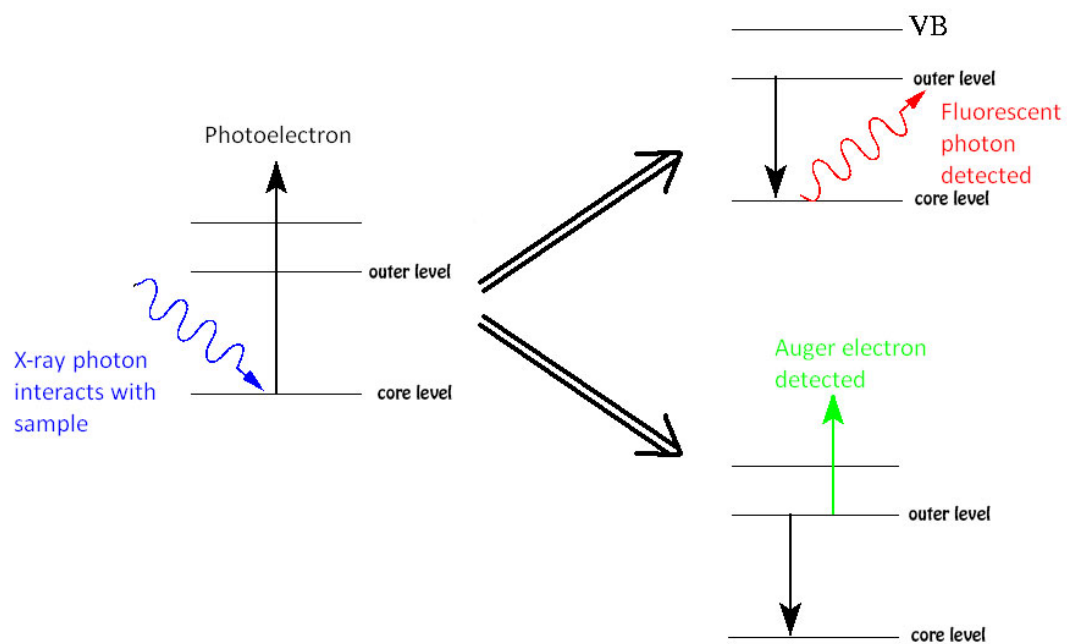
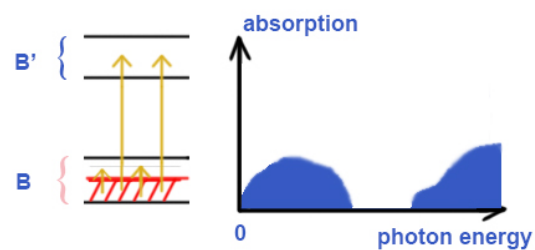
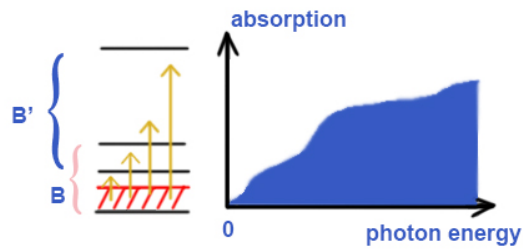


Figure 18.2: Possible processes in X-ray absorption spectroscopy



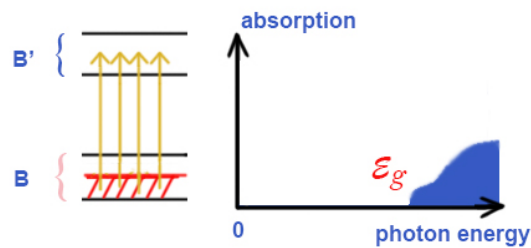
there are inter-band and intra-band transitions, resulting in two separate regions

- Conductors with band overlap:



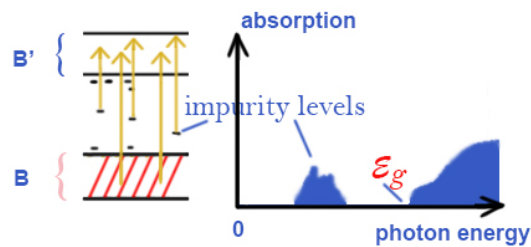
because there is no forbidden gap the spectrum is continuous.

- Insulators:



The minimum energy required is \mathcal{E}_g

- Intrinsic semiconductors: they have spectrum similar to insulators, except that because the band gap is smaller there is a small contribution from the valence band as well.
- Extrinsic (doped) semiconductors:



they have impurity levels in the band gap therefore transitions from the gap are also possible.

18.2 Emission and absorption of visible light by solids. Luminescence and phosphorescence

18.2.1 Absorption of visible light

Solids may or may not absorb light of a given frequency depending on their band structure. Absorption occurs when the photon in question is able to excite an electron in the solid and it is prohibited if no such transition is allowed, either because some selection rule prohibits it or the photon energy is smaller than the band gap.

The color of a particular solid is determined by its absorption, reflection and refraction characteristics³.

Important 18.2.1. *The energy range of light visible to the human eye is: 1.6 – 3.2 eV.*

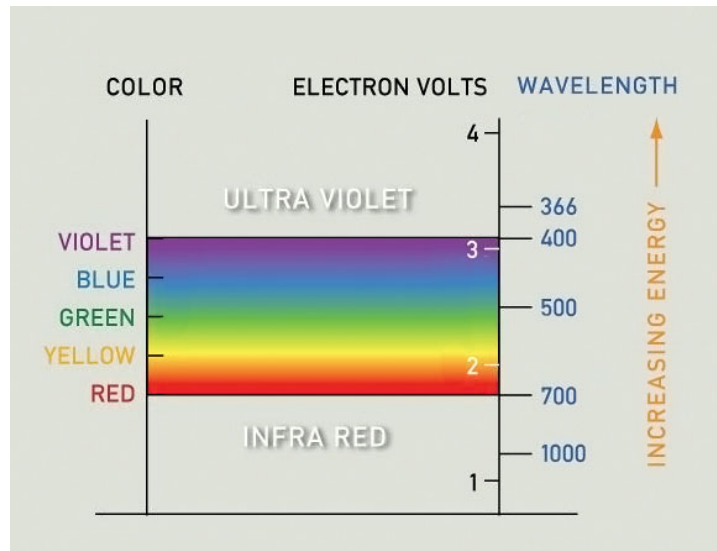


Figure 18.3: Colors and corresponding wavelengths and energies

It follows that insulators with a band gap larger than 3.2 eV are clear transparent materials, except when they contain many lattice defects. If, for instance, they contain impurities with impurity levels inside the band gap then electrons can be excited from the impurity levels to the next band or electrons can be excited from a lower band to these levels by visible light they become colored while either retaining or losing transparency, depending on the impurity concentration.

³The average human eye has only three color receptors (some women's eye have four), two of these (sensitive to red and green) with very close spectral responses, therefore the color of two solids with different absorption and reflection characteristics may look the same when illuminated by white light.

Example: impurity free crystalline form of Al_2O_3 (corundum) is transparent and colorless. When contaminated with impurities like substitutional Cr they may remain transparent but become colored. Transparent specimens are used as gems. The gems are called ruby if they are red and padparadscha if pink-orange. All other colored gems are called sapphire (e.g., "green sapphire")⁴.

Insulators that are semiconductors have a band gap smaller than 1.6 eV, consequently they are opaque.

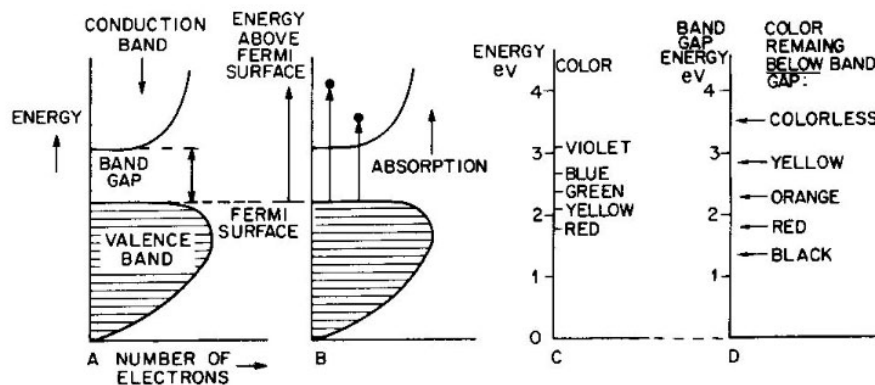


Figure 18.4: Color of insulators. A) schematic band diagram together with density of state, B) possible light absorption transitions, C) absorbed colors vs band gap width, D) observed color for various E_g values

If the efficiency of absorption and reflection (re-emission) is approximately equal at all optical energies, then all the different colors in white light will be reflected equally. This leads to the metallic silver color of polished iron and silver surfaces.

For most metals, a single continuous band extends through to high energies. The surface of a metal can absorb all wavelengths of incident light which excites electrons to a higher unoccupied energy level. When they fall back almost immediately most of the incident light is immediately re-emitted at the surface, creating the metallic luster we see in gold, silver, copper, and other metals. This is why most metals are white or silver⁵.

The efficiency of this emission process depends on selection rules. However, even when the energy supplied is sufficient, and a transition is permitted by the selection rules, this transition may not yield appreciable absorption. This can happen when the energy level accommodates only a small number of electrons.

⁴The red color of ruby comes from the absorption of green light at 561 nm ($\mathcal{E} = 2.21\text{eV}$)

⁵Gold is so malleable that it can be beaten into very thin foil less than 100 nm thick, revealing a bluish-green color when light is transmitted through it. Gold reflects yellow and red, but not blue or blue-green. The direct transmission of light through a metal in the absence of reflection is observed only in rare instances.

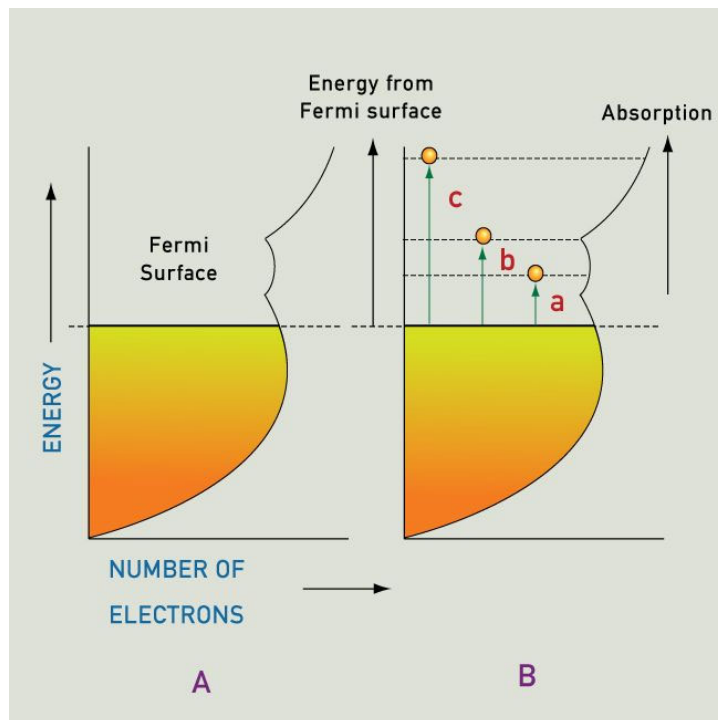


Figure 18.5: Absorption in metals. A) schematic band diagram together with density of state, B) light absorption transitions may occur at any visible frequencies.

Silver, gold and copper have similar electron configurations, but we perceive them as having quite distinct colors. This is explained in Appendix [23.13](#).

18.2.2 Luminescence and phosphorescence

Luminescence is emission of light by a substance not resulting from heat. Compare it with *incandescence* which is light emitted by a substance as a result of heating⁶. It can be caused by chemical reactions, electrical energy, subatomic motions, or stress on a crystal.

In luminescence an excited electron emits a photon when it returns to the ground state. Many processes are possible:

- direct transition to ground state

Occurs in a perfect lattice. There is a small time delay between excitation and

⁶Historically, radioactivity was thought of as a form of "radio-luminescence", although it is today considered to be separate since it involves more than electromagnetic radiation. The term 'luminescence' was introduced in 1888 by Eilhard Wiedemann.

recombination, because electrons and holes are moving in opposite directions with high mobility.

- transition through impurity levels

Electrons return from the conduction band to the valence band through impurity levels in the gap. I.e. the electrons from the impurity levels fall into the valence band, while electrons from the conduction band falls to the impurity levels left empty. These are small energy transitions in the IR region.

- electron transitions from one impurity level to an other one

$$\mathcal{E}_{lumin} < \mathcal{E}_{excitation}$$

- electron transition through traps

Traps are metastable states in the gap, from where transitions to the valence band or to impurity levels are prohibited. To get back to the valence band they first must somehow get back to the conduction band from which they decay by a luminescence process which requires a (relatively) long time. This is called *phosphorescence*.

Examples: *Zn S* - used in cathode ray tubes (excitations by electrons), *Na J* used in scintillation detectors (excitation by γ rays).

Chapter 19

Magnetism

19.1 Magnetic susceptibility

Any material placed into a magnetic field of field strength \mathbf{H} will interact with it. The best known such interaction is the case when a material becomes a permanent magnet, but this is not the most general form of *magnetism*.

From a macroscopic point of view the magnetic interaction is described by the Maxwell equation

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \quad (19.1.1)$$

where the quantity $\mu_0 \mathbf{M}$ is called *magnetic polarization*. For homogeneous and isotropic materials¹ put into weak magnetic fields in a good approximation

$$\mathbf{M} = \chi \mathbf{H} \quad (19.1.2)$$

The constant of proportionality being called the magnetic susceptibility. This definition works well for dia- and paramagnetism, but breaks down for ferro- or ferrimagnetism, because in those materials the connection between \mathbf{H} and \mathbf{B} is not linear. To simplify and unify the description a linear notation is used by making the magnetic susceptibility itself dependent on \mathbf{H} :

$$\mathbf{M} = \chi(\mathbf{H}) \mathbf{H} \quad (19.1.3)$$

¹Crystals are not always isotropic, therefore the definition of the susceptibility must be modified to cover them as well:

$$\mathbf{M} = \underset{=}{\chi} \cdot \mathbf{H}$$

where the magnetic susceptibility $\underset{=}{\chi}$ is a tensor, i.e. it is represented in any coordinate system by a 3×3 matrix whose components depends on the selection of the coordinate system. It follows that if a crystal is anisotropic the magnetic polarization is not necessarily parallel with \mathbf{H} .

Because (19.1.2) (and even the form of (19.1.3)) is linear \mathbf{H} can be expressed as

$$\mathbf{B} = \mu_0 \mu_r \mathbf{H} \quad (19.1.4)$$

where the μ_r *relative permeability* is expressed through the susceptibility by the equation:

$$\mu = \mu_0 \mu_r = \mu_0 (1 + \chi) \quad (19.1.5)$$

19.2 Types of magnetism

There are many types of magnetism in different materials:

Diamagnetism is caused by magnetic moments induced in every material by external magnetic fields. These diamagnetic moments have a magnetic field of opposite direction to the generating field due to Lenz's law and therefore, are repelled by the magnetic field. Although all materials have diamagnetic properties other types of magnetism are stronger hiding this common behavior. The diamagnetic moment is created by the external field (induced moment).

Diamagnets may be levitated in stable equilibrium in a magnetic field, with no power consumption.

The diamagnetic susceptibility is a small negative number: $-10^{-4} \lesssim \chi_{dia} \lesssim -10^{-5}$.

Paramagnetism occurs in materials with magnetic moments already present in them (e.g. the spin magnetic moment of unpaired electrons). These existing magnetic moments align with the field, giving the material a small magnetic moment, which is lost as soon as the magnetic field is switched off.

The paramagnetic susceptibility is positive and also small $10^{-1} \gtrsim \chi_{para} \gtrsim 10^{-3}$.)

Ferromagnetism of a material results in a permanent magnetic moment of either the whole of its volume or at least in regions of it. The unpaired electron magnetic moments align themselves not only as a response to external fields, but also because of a (non-magnetic) interaction with the magnetic moments of the other electrons. The magnetic polarization in ferromagnetic materials has a non-linear field dependence. Below a material specific temperature (called *Courier temperature*) the internal magnetism remains even after the field is switched off.

The ferromagnetic susceptibility $\chi(\mathbf{H}) \gg 1$, depends on the field strength and has a saturation value.

Antiferromagnetism A special form of ferromagnetism is antiferromagnetism, where the magnetic moments of the neighboring electrons are equal but opposites to each other. These materials have a zero net magnetic moment and are less common than ferromagnetic materials.

Ferrimagnetism is similar to ferromagnetism² in that respect that ferrimagnetic materials also keep their magnetic moment even in the absence of an external magnetic field. The magnetic moment of the neighboring valence electrons in them are opposites (like in antiferromagnets), but their magnitude is different, therefore they do not cancel each other out.

Superparamagnetism is the phenomena that suitably small ferro- or ferrimagnetic particles act like a single magnetic moment that is subject to Brownian motion. Their response to a magnetic field is qualitatively similar to the response of a paramagnet, but their susceptibility is much larger.

Although magnetism is strictly a quantum mechanical phenomenon³, formulas obtained using simple classical physical models to explain dia- and paramagnetic behavior are similar to those obtained by the rigorous use of quantum mechanics. Therefore we will try to use classical physical models to explain these phenomena. As for ferro- and ferrimagnetism no completely classical physical model would do.

19.3 Magnetism of free atoms.

Three factors affect the magnetic behaviour of free atoms:

- electron spin \rightarrow paramagnetism
- orbital moment \rightarrow paramagnetism
- change in the orbital moment caused by the external field \rightarrow diamagnetism

Examples:

H(1s)

orbital momentum is 0 so the source of the magnetic moment is the spin and the diamagnetic induced moment

He(1s²)

both orbital and spin momentum are 0 it only has diamagnetic induced moment

The total permanent magnetic moment of a completely filled shell including both orbital and spin momenta is 0.

²Conventionally ferrimagnetism and antiferromagnetism was considered as just some sub-cases of ferromagnetism.

³As *Bohr* and *van Leeuwen* proved, if we apply electrodynamics and classical mechanics together with thermodynamics and statistical physics consistently we find there can be no magnetism whatsoever!

Electrons in an atom or molecule are not the only particles that have magnetic moments because of their orbital and spin angular momentum. Protons and neutrons in the atomic nucleus also have spins and depending on the configuration may or may not have their own magnetic moments. Generally susceptibility from the nucleus is about 100 times smaller than that of the electrons.

19.4 Diamagnetism

Diamagnetism occurs because the external field alters the orbital velocity of electrons around their nuclei, thus changing the magnetic dipole moment. According to Lenz's law, the field of these electrons will oppose the magnetic field changes provided by the applied field.

In most materials diamagnetism is a weak effect, but in a superconductor a strong quantum effect repels the magnetic field entirely, apart from a thin layer at the surface.

In our classical physical model when $B = 0$ the electron moves around the nucleus in a classical orbit with a constant angular momentum. The magnetic moment of a current loop is equal to the current times the area of the loop, which in this case is $p_m = I A$. When an external \mathbf{B} field is turned on it exerts a torque on this magnetic moment:

$$\mathbf{T} = \mathbf{p}_m \times \mathbf{B}$$

which causes a *precession* of the angular momentum, as the magnetic moment and the angular momentum are coupled through

$$\mathbf{p}_m = \gamma \cdot \mathbf{L}$$

Here γ is the *gyromagnetic ratio* which is related to the *g-factor* g :

$$\gamma = g \frac{-e}{2m_e}$$

The angular frequency of the precession is given by the *Larmor formula*:

$$\omega = |\gamma| B = g \frac{e B}{2m_e} \quad (19.4.1)$$

where $g = 1$ in classical physics. Derivation of this formula is in Appendix 23.14.

If the atoms of the material have closed shells with a total of Z electrons on them then the total angular momentum and the coupled total magnetic moment of the atom without a magnetic field is 0, however the Larmor precession of \mathbf{L} gives rise to an additional

magnetic moment. The number of revolutions per unit time is $\omega/2\pi$, so the Z electrons of the atom present a loop current of⁴

$$I = -\frac{Ze^2B}{4\pi m_e} \quad (\text{here } g=1)$$

Suppose the field is aligned with the z axis. The average loop area can be given as $\pi \langle \rho^2 \rangle$ where $\langle \rho^2 \rangle$ is the mean square distance of the electrons perpendicular to the z axis. The magnetic moment μ of this current loop is therefore

$$p_m = -\frac{Ze^2B}{4m_e} \langle \rho^2 \rangle = -\frac{Ze^2B}{4m_e} (\langle x^2 \rangle + \langle y^2 \rangle)$$

For spherically symmetric charge distributions we may assume that the probability distribution of the three coordinates are independent and equal, so

$$\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \frac{1}{3} \langle \mathbf{r}^2 \rangle.$$

If n is the number of atoms in unit volume of the material, then the magnetic polarization of this material is⁵

$$\mathcal{M} = -\frac{nZe^2B}{6m_e} \langle r^2 \rangle$$

From the definition (19.1.2) of the magnetic susceptibility we arrive to the *Langevin formula* of the diamagnetic susceptibility of insulators and free atoms:

$$\chi = \frac{\mathcal{M}}{H} = \frac{\mu_0 \mathcal{M}}{B} = -\mu_0 \frac{NZe^2}{6m_e} \langle r^2 \rangle \quad (19.4.2)$$

This theory of diamagnetism however is not applicable to metals, as metals contain (quasi) free electrons⁶.

19.5 Pauli paramagnetism of metals

Paramagnetism is always connected to already existing permanent magnetic moments in materials. Strictly speaking paramagnetism occurs when the interaction between the

⁴The charge of an electron is $(-e)$.

⁵

$$\langle x^2 \rangle + \langle y^2 \rangle = \frac{2}{3} \langle r^2 \rangle$$

⁶Although the theory of the diamagnetism of an electron gas (*Landau diamagnetism*) is well known we do not deal with it here in detail, we will only use the result in the discussion of the (Pauli) paramagnetism of the electron gas.

magnetic moments is zero⁷ therefore without an external field the orientation of the magnetic moments is random. If the interaction is strong enough the material will be ferro-, antiferro- or ferrimagnetic⁸. Because the magnitude of the paramagnetic susceptibility is in the range⁹ $10^{-5} - 10^{-3}$, the paramagnetic polarization in an applied field is very small it requires a very sensitive device to be measured. Modern measurements on paramagnetic materials are often conducted with a SQUID magnetometer (see Section ??). A material may contain permanent magnetic moments if

- it is a metal which contains free electrons with their spin related magnetic moments. Examples: Al, Cs, Li, Mg, Na, W.
- it contains atoms, molecules or lattice defects with an odd number of electrons (total electron spin is not 0). The O_2 molecule is a good example.
- it contains atoms with unoccupied inner shells. Examples: K, Ca.

Non ferromagnetic metals are usually paramagnetic, because the wave functions of their s and p electrons are strongly delocalized, which usually leads to the pairing of electron spins therefore very weak magnetic moments. An exception is gold, which is diamagnetic, because in a magnetic field the diamagnetic moments from the electrons on its closed inner shells are larger than the paramagnetic moments of the delocalized electrons.

Metals contain (quasi) free electrons which have a spin-related magnetic moment of

$$\mu_S = \frac{-g\mu_B}{\hbar} \mathbf{S} = g \frac{-e}{2m_e} \mathbf{S} \quad (19.5.1)$$

where $g = 2$ and the $-$ sign is present because the charge of the electron is $(-e)$, and $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr-magneton (C.f. (6.3.1)).

In an external \mathbf{B} field the magnetic moment of the electron, like the spin, must be either parallel or anti-parallel to the magnetic field. Consequently the interaction energy is either positive or negative

$$\Delta\mathcal{E} = -\mu_S \cdot B = \pm \frac{g e}{4m_e} = \pm \frac{e}{2m_e} \quad (19.5.2)$$

where the $+$ sign corresponds to the anti-parallel ($B \uparrow \mu_S \downarrow$), the $-$ sign to the parallel ($B \uparrow \mu_S \uparrow$) orientation. Because the energy of the electrons depend on the orientation of their spins (the total energy will be $\mathcal{E}(B) = \mathcal{E}(B = 0) + \Delta\mathcal{E}$) this deforms their distributions. But the \mathcal{E}_F Fermi energy of electrons of both spin orientations must be the

⁷Or at least the interaction energy is smaller than the thermal energy.

⁸In this section when we use the term 'ferromagnetic' we mean any of these.

⁹may be as high as 10^{-1} for *synthetic paramagnets*

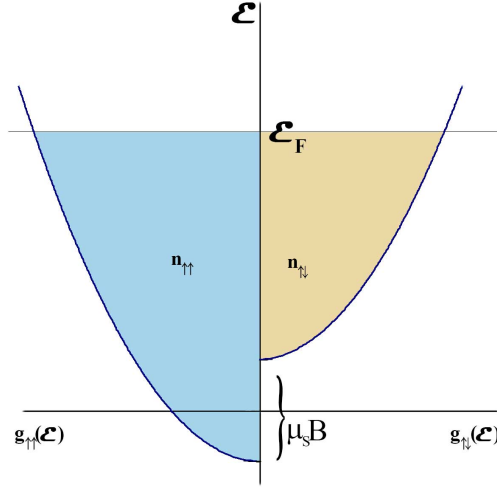


Figure 19.1: The formation of the electron (Pauli) paramagnetism

same. As a consequence the number density of electrons with parallel and ant-parallel spins denoted by $n_{\uparrow\uparrow}$ and $n_{\uparrow\downarrow}$ respectively must also differ.

The paramagnetic polarization \mathcal{M} from Appendix 23.15 (equation (23.15.4))

$$\mathcal{M} = \mu_S \cdot (n_{\uparrow\uparrow} - n_{\uparrow\downarrow}) = \mu_S^2 g(\mathcal{E}_F) B = \frac{3 n_{tot} \mu_S^2}{2 e \mathcal{E}_F} B \quad (19.5.3)$$

where n_{tot} is the electron density in the metal and the paramagnetic susceptibility of the electron gas is

$$\chi = \frac{3 n_{tot} \mu_S^2}{2 e \mathcal{E}_F} \quad (19.5.4)$$

However the external B field modifies the spatial movement of the electrons as well, which gives rise to a diamagnetic momentum. Without going into details we just use the result:

$$\mathcal{M}_{dia} = -\frac{1}{3} \mathcal{M}_{para} \quad (19.5.5)$$

The total susceptibility therefore positive and its magnitude is

$$\chi = \frac{n_{tot} \mu_S^2}{e \mathcal{E}_F} \quad (19.5.6)$$

As we see the susceptibility is independent of the temperature. This follows from the Pauli principle: only electrons in the vicinity of the Fermi energy can change their energy

as a response to the \mathbf{B} field. The number of these electrons is proportional to $\frac{k_B T}{\mathcal{E}_F}$ while the difference of the magnetic moments in is proportional to $\frac{\mu_S B}{k_B T}$. The resulting magnetic moment \mathcal{M} being the product of these factors is temperature independent. But the paramagnetic moment from independent atomic moments is temperature dependent.

19.6 Paramagnetism of independent atomic moments

Paramagnetism may not only arise due to the collective system of electrons as in metals, but also as a result of independent atomic (orbital momentum and electron spin related) moments being oriented into the direction of the external field. In this case the occupation numbers of the two non-degenerate levels produced by the split of the twice degenerated ones may be calculated using the Boltzmann-factor. Let us denote the total electron density with n_0 and the electron density on the two non-degenerate levels with n_1 and n_2 . Then for a single level in thermal equilibrium the well known *Zeeman splitting* occurs

$$\frac{n_1}{n_0} = \frac{e^{\mu_S B/k_B T}}{e^{\mu_S B/k_B T} + e^{-\mu_S B/k_B T}} \quad (19.6.1)$$

$$\frac{n_2}{n_0} = \frac{e^{-\mu_S B/k_B T}}{e^{\mu_S B/k_B T} + e^{-\mu_S B/k_B T}} \quad (19.6.2)$$

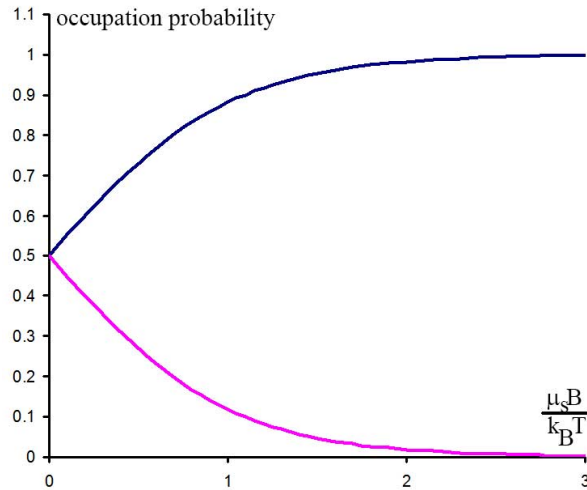


Figure 19.2: Paramagnetic moment of independent electrons as a function of $\mu_S B/k_B T$.

The total magnetic moment of a unit volume then

$$\mathcal{M} = (n_1 - n_2) \cdot \mu_S = n_0 \cdot \frac{e^{\mu_S B/k_B T} - e^{-\mu_S B/k_B T}}{e^{\mu_S B/k_B T} + e^{-\mu_S B/k_B T}} = n_0 \cdot \mu_S \cdot \tanh \frac{\mu_S B}{k_B T} \quad (19.6.3)$$

When $\mu_S B \ll k_B T$ the hyperbolic tangent function may be approximated by its argument, and the magnetic polarization becomes

$$\mathcal{M} = \chi_p B \approx n_0 \cdot \mu_S \cdot \frac{\mu_S B}{k_B T} = \frac{C \cdot B}{T} \quad (19.6.4)$$

where

$$C = \frac{n_0 \cdot \mu_S^2}{k_B} \quad (19.6.5)$$

is called the Curie-constant. *This is the Curie-law of paramagnetism.* Curie's law is valid under the commonly encountered conditions of low magnetization, but does not apply in the high-field/low-temperature regime where saturation of magnetization occurs ($\mu_S B \approx k_B T$) and magnetic dipoles are all aligned with the applied field. When all of the dipoles are aligned, increasing the external field will not increase the total magnetization since there can be no further alignment.

Let us denote the maximum of the z component of the total angular momentum (both the orbital momentum and spin) with $J \cdot \hbar$ and the total magnetic moment with μ . The degeneracy of the original single energy level in this case¹⁰ is $2J + 1$. If we calculate \mathcal{M} again in the limit of $\mu B \ll k_B T$ we find a similar formula as in (19.6.4)¹¹:

$$\mathcal{M} = \frac{n_0 \cdot J(J+1) \mu^2 B}{3 k_B T} = \frac{C \cdot B}{T} \quad (19.6.6)$$

where μ is the magnetic moment associated to the total angular momentum¹²

$$C = \frac{n_0 \cdot J(J+1) \mu^2}{3 k_B} \quad (19.6.7)$$

(19.6.6) shows that the paramagnetic susceptibility is influenced strongly by the total J angular momentum:

$$\chi = \frac{n_0 \cdot J(J+1) \mu^2}{3 k_B T} \quad (19.6.8)$$

¹⁰e.g. for the inner unfilled levels of Pd

¹¹The derivation of this formula which explains the appearance of all the factors is somewhat complicated so we do not present it here. Those interested may refer to <http://en.wikipedia.org/wiki/Paramagnetism>.

¹² $\mu = g_J \cdot \mu_B$, where μ_B is the (6.3.1) Bohr-magneton.

19.7 Ferromagnetism

Ferromagnetic materials also contain constant magnetic moments, like paramagnetic materials do. However in contrast with the paramagnetic behavior ferromagnetism is a collective phenomena. In ferromagnetic materials permanent magnetism may be observed when all of the elementary moments are oriented in the same direction, even without any external magnetic field. This is called *spontaneous magnetization*. Even when a ferromagnetic material seemingly does not possess a permanent magnetic moment we find that regions of it do. These regions are called *magnetic domains* and are separated by relatively stable *domain walls*. Domain sizes range between 0.1 to several mm. In each of these domains all magnetic moments are parallel, but the orientation of the magnetic moments of these domains are random because it is energetically favorable, therefore no macroscopic magnetic moment is observed.

When an external \mathbf{H} field is applied to a ferromagnetic material with unordered domains the domain structure changes. At lower field strengths this change is reversible, after switching the field off again the original structure will reassert itself. Irreversible

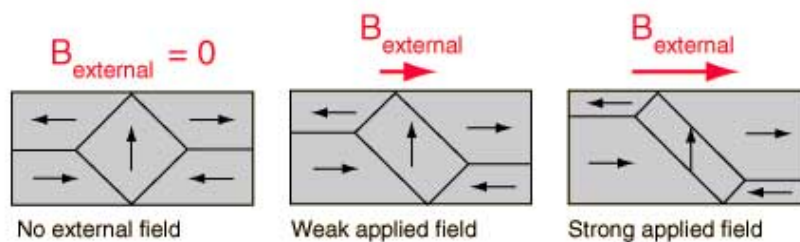


Figure 19.3: Schematics of ferromagnetic magnetization

change occurs when the magnetic field is strong enough to make domain walls move and rotate through lattice defects. As the \mathbf{H} field changes, the magnetization too changes in thousands of tiny discontinuous jumps as the domain walls suddenly "snap" past defects. This is called the *Barkhausen effect*, which is the direct evidence of the existence of ferromagnetic domains. It can be observed by winding a coil around the material in which these sudden jumps induce electric pulses. After amplification the pulses will be audible as a series of clicks. The material will not return to its macroscopically non-magnetic state even after we switch off the magnetic field, because the domain walls cannot move past the defects without an external energy source, consequently a non zero \mathcal{M}_{rem} magnetization remains in it.

This non-ground state of aligned domains is metastable and can persist for long periods of time. Some samples of magnetite collected from the sea floor, have maintained their magnetization for millions of years.

Increasing the external magnetic field strength turns more and more domains to be parallel with it. There exists a saturation field strength \mathbf{H}_{sat} when all magnetic moments point the same direction. Further increase in the magnetic field cannot produce larger magnetization. Due to the irreversible changes to return to the original non-magnetic state we need to apply again an external field in a direction opposite to the field which created the magnetized state. Therefore the $\mathcal{M}(H)$ curve (Fig. 19.4) shows *hysteresis*.

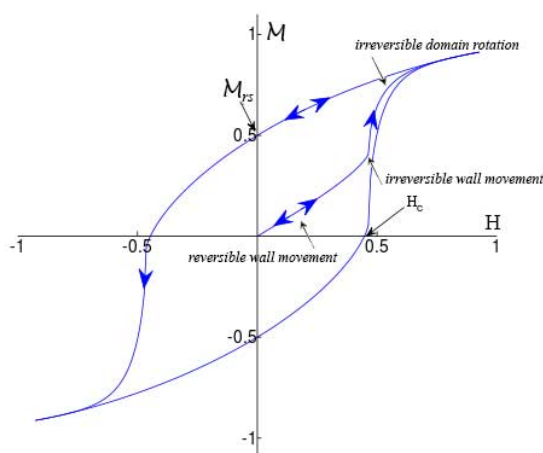


Figure 19.4: Magnetization \mathcal{M} against magnetic field H . Units used are the saturation field H_s and saturation magnetization \mathcal{M}_s . Starting at the origin, the upward curve is the *initial magnetization curve*. The downward curve after saturation, along with the lower return curve, form the main loop. The intercepts H_c and \mathcal{M}_{rs} are the *coercivity* and *saturation remanence*.

Ferromagnetism is a property related not only to the chemical composition of a material, but of its crystalline structure and microscopic organization¹³.

This ordered state may not be explained by classical magnetic interaction between the constant magnetic moments from two reasons:

¹³Some metal alloys, called Heusler alloys are ferromagnetic although their constituents are not themselves ferromagnetic. On the other hand some alloys, like certain types of stainless steel, are non-magnetic, but are composed almost exclusively of ferromagnetic metals.

Non-crystalline ferromagnetic metallic alloys may be produced by very rapid quenching (cooling) of a liquid alloy. Their advantage is their nearly isotropic properties, this results in low coercivity, low hysteresis loss, high permeability, and high electrical resistivity. One such typical material is a transition metal-metalloid alloy, made from about 80% transition metal (usually Fe, Co, or Ni) and a metalloid component (B, C, Si, P, or Al) that lowers the melting point. A relatively new class of exceptionally strong ferromagnetic materials are the rare-earth magnets. They contain lanthanide elements that are known for their ability to carry large magnetic moments in well-localized f-orbitals.

1. because the magnetic dipol–dipol interaction energy

$$\mathcal{E} = -\frac{\mu_0}{4\pi} \left(\frac{3(\boldsymbol{\mu}_1 \mathbf{r})(\boldsymbol{\mu}_2 \mathbf{r})}{r^5} - \frac{\boldsymbol{\mu}_1 \boldsymbol{\mu}_2}{r^3} \right)$$

is in the range of 10^{-4} eV which is too small compared to the $k_B T$ thermal energy¹⁴, which means the ordering interaction may not be a classical physical magnetic one.

2. classical dipole-dipole interaction would turn neighboring moments not the same but in opposite directions.

Still we may describe this behavior by introducing an internal *exchange (magnetic) field*.

The ordering interaction is the consequence of the Pauli principle. The total state function of the system of electrons must be antisymmetric. When electron spins are parallel (spin part of the wave function is symmetric) the corresponding spatial wave function must be antisymmetric, which corresponds to electrons that are further apart. This reduces the electrostatic energy of the electrons when their spins are parallel compared to their energy when the spins are anti-parallel, so the parallel-spin state is more stable. This difference in energy is called the *exchange energy* (C.f. Section 6.7).

We may assume the internal exchange field B_i is proportional to the magnetization \mathcal{M} itself which it creates

$$B_i = \lambda \mathcal{M}$$

and using the paramagnetic susceptibility formula (see (19.6.4)) $\chi_p = C/T$ we deduced above:

$$\mathcal{M} = \chi_p \cdot (B + B_i) = \chi_p \cdot (B + \lambda \mathcal{M}) = \frac{C}{T} \cdot (B + \lambda \mathcal{M})$$

After reordering

$$\mathcal{M} = \frac{C}{T - \lambda C} B \tag{19.7.1}$$

$$\chi = \frac{C}{T - \lambda C} = \frac{C}{T - T_c} \tag{19.7.2}$$

This is the *Curie-Weiss law*¹⁵. C is the material specific Curie constant and the quantity T_c is called the *Curie temperature* aka Curie point. From (19.7.2) it follows that as the T temperature approaches T_c from above the susceptibility approaches infinity. Above T_c the susceptibility is positive, i.e. the material becomes paramagnetic. Although this model may loose validity near T_c , experimental facts indicate that real ferromagnetic materials obey this law with a good accuracy and ferromagnetic ordering is lost above T_c , and paramagnetic behavior takes over.

¹⁴At 300K it $k_B T = 0.0258 \text{ eV}$.

¹⁵More accurate models predict a dependence on the 1.33 power of the denominator

Material	T_c (K)
Co	1388
Fe	1043
MnBi	630
Ni	627
MnSb	587
CrO2	386
MnAs	318
Gd	292
Dy	88
EuO	69

Table 19.1: Curie temperatures of some crystalline materials(Source:Wikipedia)

Below the Curie temperature *spontaneous magnetization* occurs¹⁶. The resulting *saturation magnetization* of the domains that form can be calculated from a formula similar to (19.6.3):

$$\mathcal{M} = n_0 \cdot \mu \cdot th \frac{\mu B_i}{k_B T} = n_0 \cdot \mu \cdot th \frac{\mu \lambda \mathcal{M}}{k_B T}$$

Substituting $\lambda = T_c/C$ and using (19.6.5)

$$\frac{\mathcal{M}}{n_0 \mu} = th \frac{\mu T_c \mathcal{M}}{C k_B T} = th \frac{\mathcal{M}}{n_0 \mu} \frac{T_c}{T}$$

\mathcal{M} could be calculated from this numerically or graphically.

Example 19.1. *The susceptibility of a ferromagnetic material is $\chi = 0.0116$ at $T = 1100$ K and $\chi = 0.0042$ at $T = 1200$ K. What material is it? of this material?* **Solution** Calculate its Curie temperature first. Using (19.7.2) for both susceptibility we get two equations for the two unknown, from which both T_c and C (and

¹⁶This is an example of *spontaneous symmetry breaking*: above the Curie temperature the state of the system is symmetric, and this symmetry breaks as the non-symmetric spontaneous magnetization occurs

therefore λ too) can be determined

$$\begin{aligned}\chi_1 &= \frac{C}{T_1 - T_c} &\Rightarrow & C = \chi_1 \cdot (T_1 - T_c) \\ \chi_2 &= \frac{C}{T_2 - T_c} &\Rightarrow & C = \chi_2 \cdot (T_2 - T_c) \\ \chi_1 \cdot (T_1 - T_c) &= \chi_2 \cdot (T_2 - T_c) \\ T_c &= \frac{\chi_1 \cdot T_1 - \chi_2 \cdot T_2}{\chi_1 - \chi_2} = 1043 \text{ K}\end{aligned}$$

Using Table 19.1 we find the material is iron. Furthermore from the equations above we get the value of $C = 0.66$ too

19.8 Antiferromagnets

In antiferromagnetic materials the magnetic moments are ordered similarly as in ferromagnets, but the neighboring magnetic moments are oriented in opposite directions to each other. Therefore the magnetization in these materials vanishes. But this is only valid at low temperatures. Antiferromagnetic materials also have a critical temperature called the *Néel temperature* above which the antiferromagnetic ordering is lost and the material becomes paramagnetic. In contrast, to the transition between the ferromagnetic to the paramagnetic phases where the susceptibility diverges the magnetic susceptibility of an antiferromagnetic material typically shows a maximum at the Néel temperature. Examples of antiferromagnetic materials are: hematite, Cr, iron manganese (*FeMn*), *NiO*.

19.9 Ferrimagnetism

The oldest known magnetic substance magnetite (iron(II,III) oxide: Fe_3O_4) is a *ferrimagnetic material*¹⁷. In a crystal lattice with a basis a part of the basis in every cell form a *sublattice*. In a ferrimagnetic material the magnetic moments of the different sublattices of different materials or ions (such as Fe^{2+} and Fe^{3+}) are antiparallel and are unequal which results in spontaneous magnetization. Ferrimagnetic materials are the ferrites, (composed of iron oxides and other elements such as aluminum, cobalt, nickel, manganese and zinc), and magnetic garnets (silicate minerals, yttrium iron garnet or YIG).

¹⁷Magnetite was originally classified as a ferromagnet before Neel's discovery of ferrimagnetism and antiferromagnetism in 1948.

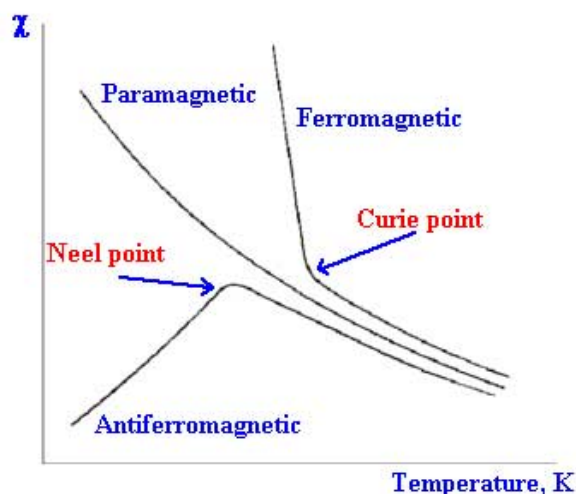


Figure 19.5: Comparison of the temperature dependence of para-, ferro- and antiferromagnetic susceptibilities

Like ferromagnetic materials ferrimagnetic ones have spontaneous magnetization below the Curie temperature, and are paramagnetic (show no magnetic order) above this temperature¹⁸.

Ferrimagnetic materials have high resistivity. This is an advantage as it prohibits the appearance of *eddy currents* which lead to significant heating losses in metallic ferromagnetic materials. Ferrimagnetic materials absorb long wavelength electromagnetic radiation which is a unique property as metals reflect, insulators transmit those. Ferrite crystals are a major ingredient in snoopers paint, which makes stealth airplanes undetectable by radar. They also have anisotropic properties. The anisotropy is induced by an external applied field, which if it aligns with the magnetic dipoles causes a net magnetic dipole moment and causes the magnetic dipoles to precess at the Larmor frequency (see Section 19.4). For instance a microwave signal circularly polarized in the same direction as this precession strongly interacts with the magnetic dipole moments; when it is polarized in the opposite direction the interaction is very low¹⁹.

¹⁸However, there is sometimes a temperature below the Curie temperature at which the two sublattices have equal moments, resulting in a net magnetic moment of zero; this is called the magnetization compensation point. Furthermore, ferrimagnets may also exhibit an angular momentum compensation point at which the angular momentum of the magnetic sublattices is compensated. This compensation point is a crucial point for achieving high speed magnetization reversal in magnetic memory devices.

¹⁹When the interaction is strong, the microwave signal can pass through the material. This directional property is used in the construction of microwave devices like isolators, circulators and gyrators. Ferrimagnetic materials are also used to produce optical isolators and circulators. See <http://en.wikipedia.org/wiki/Ferrimagnetism>

Material	T_c (K)
MnO	116
MnS	160
MnTe	307
MnF2	67
FeF2	79
FeCl2	24
FeO	198
CoCl2	25
CoO	291
NiCl2	50
NiO	525
Cr	308

Table 19.2: Neel temperatures of some crystalline materials(Source:Wikipedia)

Chapter 20

Dielectric properties of solids

In insulators there are no movable charge carriers therefore no electric current flows when an external electric field is applied to them. But the unmovable charges (e.g. ion cores and valence electrons) are affected by the external field and they shift from their original positions forming electric *dipoles*. This phenomena is called *dielectric polarization* or simply *polarization*. Polarization is characterized by the *polarization density* or *polarizability* vector \mathbf{P} which is the density of the electric dipole moment of the material¹.

It is also possible that the material was *polar*, i.e. already contained molecules with non zero dipole moments and the external \mathbf{E} field tries to rotate these into the same direction. However independent of the concrete method of polarization the field created or rotated dipole moments create an electric field in the opposite direction of \mathbf{E} thus weakening it inside the material.

The electric susceptibility χ_e of a dielectric material is a measure of how easily it polarizes in response to an electric field. In many cases polarizability is proportional to the field²

$$\mathbf{P} = \epsilon_o \chi_e \mathbf{E} \quad (20.0.1)$$

The well known macroscopic formula connects the *electric displacement* \mathbf{D} , the external field \mathbf{E} and the polarization \mathbf{P}

$$\mathbf{D} = \epsilon_o \mathbf{E} + \mathbf{P} \quad (20.0.2)$$

In homogeneous and isotropic materials χ_e is a scalar and

$$\mathbf{D} = \epsilon_o (1 + \chi_e) \mathbf{E} = \epsilon_r \epsilon_o \mathbf{E} \quad (20.0.3)$$

¹In the simplest case this is the electric dipole moment of the unit volume.

² $\epsilon_o = 8.854187817620 \cdot 10^{-12} \text{ F/m}$ is the electric permittivity of free space which has a *defined value*

$$\epsilon_o = \frac{1}{\mu_o c^2}$$

because both $\mu_o = 4 \pi \cdot 10^{-7} \text{ H/m}$. and $c = 299\,792\,458 \text{ m/s}$ are defined values.

In non-isotropic (non-cubic) crystals χ_e is a tensor, represented by a matrix.

The different kinds of polarization are

- Induced polarization - polarization of non-polar molecules/atoms (electron polarization, displacement or ionic polarization)

The external field creates dipole moments inside the material either by deforming the electron shell of atoms or moving the atoms of molecules apart.

- orientation polarization - polarization of polar molecules

The external field only aligns existing dipole moments.

20.1 Induced polarization

Fig. 20.1 shows the schematics of electron polarization. For ionic polarization the same schematics holds just the constant dipole moment is created by displacing the atoms of a molecule from each other by the external field. In both cases we may use a harmonic oscillator model for the dipoles. The interaction energy around the minimum can be

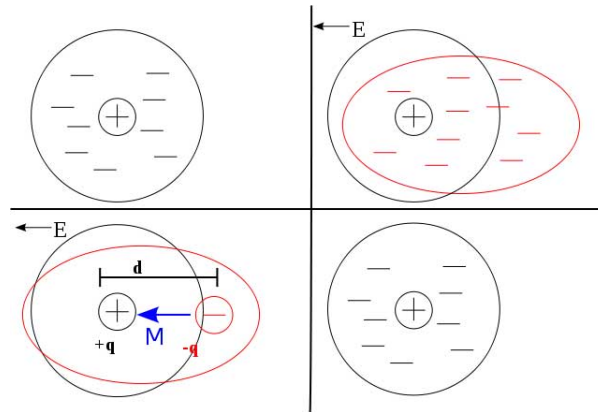


Figure 20.1: Classical physical model of electron polarization. The external $\mathbf{E}(t)$ field shifts the center of charges of the ion cores and valence electrons. This charge distribution can be reduced to that of a single dipole using the superposition principle. \mathbf{M} is the dipole moment vector: $M = q d$.

approximated by a harmonic potential. For electron polarization this is in which the electron moves. As the mass of the electron is much smaller than the mass of the ion core we may consider the ion core immobile and deal with the electron only. For ionic polarization this harmonic potential is a good approximation around the equilibrium ion distance. In the following we use the example of electron polarization, but the same arguments could be used for ionic polarization too.

Electronic polarization

When a time dependent electric field $\mathbf{E}(t)$ is applied to an atom the Newton's equation of classical physics for the movement of the (center of mass of the) electron is described by a damped and driven harmonic oscillator. In 1 dimension:

$$m \ddot{x} + m \omega_0^2 x - \gamma \dot{x} = -e E \quad (= -e E_0 e^{i\omega t}) \quad (20.1.1)$$

Here ω_0 is the angular frequency of the atomic oscillator and γ is the damping constant.

The electric field $\mathbf{E} = (\mathbf{E}_{loc} =) \mathbf{E}_{aver} + \mathbf{E}_{pol}$ is the field felt by the atom. In addition to the average \mathbf{E}_{aver} field the atomic/molecular dipoles feel the field \mathbf{E}_{pol} of the neighboring dipoles as well. We will determine this local \mathbf{E}_{loc} field later.

When the field is static ($E(t) = const$), $\dot{x} = 0$ and $\ddot{x} = 0$ and the equilibrium distance of the electron and the ion core is

$$x = \frac{-e E}{m \omega_0^2}$$

where in solids m is the effective mass of the electron, from which the (static) electric dipole moment of the atom is

$$p = -e x = \frac{e^2 E}{m \omega_0^2} = \alpha \epsilon_o E \quad (20.1.2)$$

where we introduced the atomic polarizability α

$$\alpha = \frac{e^2}{\epsilon_o m \omega_0^2} \quad (20.1.3)$$

The polarization density of N such atoms (N is the atom density) then

$$\mathcal{P} = N p = N \alpha \epsilon_o E \quad (20.1.4)$$

and the susceptibility is

$$\chi_e = \frac{\mathcal{P}}{\epsilon_o E} = N \alpha = N \frac{e^2}{\epsilon_o m \omega_0^2} \quad (20.1.5)$$

When the external field is not static the solution of (20.1.1) has the form

$$\mathbf{r}(t) = \mathbf{r}_0 e^{i\omega t}$$

After substitution we arrive to the formula

$$\mathbf{r} = \frac{e}{m} \cdot \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma} \mathbf{E} \quad (20.1.6)$$

The induced dipole moment of the atom then simply

$$\mathbf{p}_{ind} = e \cdot \mathbf{r} = \frac{e^2}{m} \cdot \frac{1}{\omega_0^2 - \omega^2 + i\omega\gamma} \mathbf{E} \quad (20.1.7)$$

Since this is a complex quantity, in this case the atomic polarizability will also be complex:

$$\alpha_{ind} = \frac{e^2}{\epsilon_o m} \cdot \frac{1}{\omega_0^2 - \omega^2 + i \omega \gamma} \quad (20.1.8)$$

If we introduce the *plasma frequency*

$$\omega_p \equiv \sqrt{\frac{N e^2}{m \epsilon_o}} \quad (20.1.9)$$

$$\alpha_{ind} = \frac{\omega_p^2}{N} \cdot \frac{1}{\omega_0^2 - \omega^2 + i \omega \gamma} \quad (20.1.10)$$

The effects of a complex polarizability will be dealt later in Section 20.4.

Ionic polarization

This occurs in ionic solids such as sodium chloride etc. Ionic solids possess net dipole moment even in the absence of external electric field. But when the external electric field is applied the separation between the ions further increases. Hence the net dipole moment of the material also increases.

To get the formula for displacement polarizability in general and not only for electrons substitute the q charge of the ions into every equations in which we used e so far.

20.2 Orientation polarization

When the molecules of the solid have constant dipole moments a homogeneous external \mathbf{E} field³ creates a torque which tries to rotate the dipoles into the direction of the field. Thermal vibrations act against this, consequently only part of the molecular dipoles will be rotated into the direction of the field. Let us select a coordinate system whose positive z-direction is parallel with \mathbf{E} and use spherical polar coordinates (ρ, θ, φ) . Then for the polarization density and the susceptibility in homogeneous an isotropic material⁴ we obtain the following formulas:

$$\mathbf{P} = \frac{N p_e^2}{3 k_B T} \mathbf{E} \quad (20.2.1)$$

$$\chi = \frac{N p_e^2}{3 k_B T} \quad (20.2.2)$$

³ Although denoted by \mathbf{E} this is not the external electric field but the *local electric field* which incorporates the external field and the fields of the neighboring dipoles.

⁴e.g. cubic crystals

It follows that the temperature dependence of the orientation susceptibility is $\propto T$, which is the *Courier law*. In the high field limit all dipoles are turned into the direction of the field while in the small field limit the polarization density is proportional to the field⁵:

$$\lim_{E/T \rightarrow \infty} \mathcal{P} = N p_e \cdot (1 - 0) = N p_e \quad \text{high field limit}$$

$$\lim_{E/T \rightarrow 0} \mathcal{P} = \frac{N p_e \alpha}{3} = \frac{N p_e^2}{3 k_B T} E \quad \text{low field limit}$$

The local \mathbf{E}_{loc} field in uniformly polarized homogeneous isotropic (cubic) solids

As we said earlier the local electric field \mathbf{E}_{loc} which creates or orients the electric dipoles is not the same as the average \mathbf{E} field in the dielectric, because the atomic/molecular dipoles feel the field of the neighboring dipoles as well. In an isotropic uniformly polarized material every atom/molecule finds itself surrounded by the other atoms/molecules a spherical symmetric way. When a uniform average \mathbf{E} field is present all of them will be uniformly polarized. To calculate \mathbf{E}_{loc} we must sum up the electric field of all of these at the position of our selected atom. This would require complicated calculations. Instead of this we may use a simplified model.

Let us take a spherical plug of polarized material surrounding our selected atom out of the dielectric. This will result in the appearance of inhomogeneous surface charges on the surface of the hole this creates. These surface charges will create the same field at the position of our atom (origin of the hole) as the original material without the hole did.

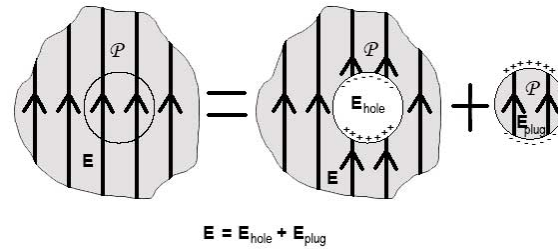


Figure 20.2: The field at any point of the dielectric can be considered as a sum of the field in a spherical hole (i.e. the local field) plus the field of the spherical plug.

$$\mathbf{E} = \mathbf{E}_{hole} + \mathbf{E}_{plug}$$

The local field is the field in the hole:

$$\mathbf{E}_{loc} \equiv \mathbf{E}_{hole} = \mathbf{E} - \mathbf{E}_{plug}$$

⁵Detailed derivation of the polarization for this case is in Appendix 23.16

where \mathbf{E}_{plug} is the electric field inside a uniformly polarized sphere, which is

$$\mathbf{E}_{plug} = -\frac{\mathbf{P}}{3\epsilon_o} \quad (20.2.3)$$

Therefore the local field, which is larger than \mathbf{E} is:

$$\mathbf{E}_{loc} = \mathbf{E} + \frac{\mathbf{P}}{3\epsilon_o} \quad (20.2.4)$$

Those interested in the derivation will find it in Appendix [23.17](#).

The Clausius-Mosotti formula

The fact that the local field is on one hand responsible for the polarization and on the other hand at the same time it depends on the polarization too leads to the reformulation of the susceptibility. Substituting the local field [\(20.2.4\)](#) into the electronic polarization formula [\(20.1.4\)](#) containing the atomic polarizability α

$$\mathbf{P} = N p = N \alpha \epsilon_o \left(E + \frac{\mathbf{P}}{3\epsilon_o} \right)$$

$$\mathbf{P} = \frac{N \alpha}{1 - \frac{N \alpha}{3}} \epsilon_o \mathcal{E} \quad (20.2.5)$$

$$\chi_e = \frac{N \alpha}{1 - \frac{N \alpha}{3}} \quad (20.2.6)$$

20.3 Solid Dielectrics

There are special solid dielectric materials in which there exists a permanent built-in polarization even in the absence of an external electric field. For instance wax contains long molecules having a permanent dipole moment. If you melt some wax and apply a strong electric field on it when it is a liquid then cool it down the (partial) ordering of the permanent dipoles stays that way when the liquid freezes. Such a solid is called an *electret*. An electret is the electrical analog of a ferromagnet only much less useful. The air always contains free charges which are attracted to the surface and which neutralize the polarization charges.

Not only electrets may contain permanent polarization but there exist such crystalline materials too. Normally this effect is also unnoticed from the same reason. However when these permanent dipole moments change external fields appear. This change may be caused by

- thermal expansion - this is called *pyroelectricity*
- mechanical stress - this is called *piezoelectricity*

Ferroelectric crystals like BaTiO₃ also have built-in permanent dipole moment. But if we increase the temperature even a tiny bit they loose this permanent moment. However in cubic crystals in which the moments can be oriented into any direction all of them can change at the same time when the external field changes and we get a large effect.

20.4 Application of the oscillator model

In section 20.1 we calculated the complex polarizability as a function of the frequency of the external field. This describes propagation of electromagnetic waves in dielectric materials. If we substitute (20.1.8) in the (20.2.6) Clausius-Mosotti formula we may calculate the real and imaginary part of the permittivity. To simplify the formulas let us introduce two frequencies: the plasma frequency ω_p and the resonance frequency ω_1

$$\omega_p^2 \equiv \frac{N q^2}{\epsilon_o} m_e \quad (20.4.1)$$

$$\omega_1^2 \equiv \omega_o^2 - \frac{N q^2}{\epsilon_o} \quad (20.4.2)$$

With these the real (ϵ') and imaginary (ϵ'') parts of the permittivity are

$$\epsilon'(\omega) = 1 + \omega_p^2 \frac{\omega_1^2 - \omega^2}{(\omega_1^2 - \omega^2)^2 + \omega^2 \gamma^2} \quad (20.4.3)$$

$$\epsilon''(\omega) = \omega_p^2 \frac{\gamma \omega}{(\omega_1^2 - \omega^2)^2 + \omega^2 \gamma^2} \quad (20.4.4)$$

The real part of the complex *refractive index* (or *index of refraction*)

$$\hat{n} \equiv n - i n \kappa = \sqrt{\epsilon}$$

describes refraction, while the imaginary part is responsible for absorption:

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E} e^{i\omega(t - \hat{n}r/c)} = \\ &= (\mathbf{E} e^{-n\kappa r/c}) e^{i\omega(t - nr/c)} \end{aligned} \quad (20.4.5)$$

The real (not complex) refractive index n and absorption coefficient κ can be calculated from

$$\epsilon'(\omega) = n^2 - n^2 \kappa \quad (20.4.6)$$

$$\epsilon''(\omega) = 2 n^2 \kappa \quad (20.4.7)$$

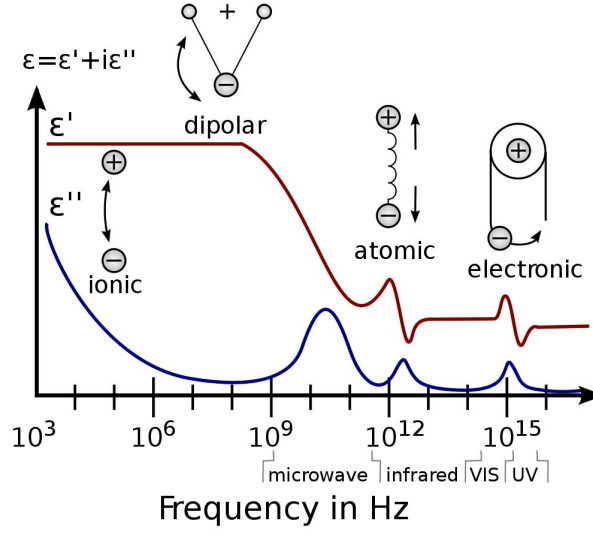


Figure 20.3: Dielectric responses in different frequency ranges

or

$$\epsilon'(\omega) = n^2 - n^2 \kappa \quad (20.4.8)$$

$$\epsilon''(\omega) = 2 n^2 \kappa \quad (20.4.9)$$

Usually the real refractive index $n \geq 1$. This is nice, because by definition (C.f. (20.4.5))

$$n = \frac{c_m}{c}$$

where c is the speed of light in vacuum and c_m is the *phase velocity* of light in the material, but in some cases (e.g. near to resonant frequencies or for X-rays) n may be smaller than 1 giving a phase velocity which is larger than c . This is possible as the phase velocity is not the velocity of the energy or information of light, which are given by the propagation speed which is never larger than c . As an example, water has a refractive index of $1 - 2.6 \cdot 10^{-7} = 0.99999974$ at a photon energy of 30 keV (0.04 nm wavelength)⁶.

⁶Recent research has also demonstrated the existence of the negative refractive index, which can occur if permittivity and permeability have simultaneous negative values. This can be achieved with specially crafted periodic structures which are called *metamaterials*. The resulting negative refraction (i.e., a reversal of Snell's law) offers the possibility (but not the reality - as of 2012) of the superlens (lens whose resolution go beyond the diffraction limit) and other exotic phenomena.

Conduction electrons in metals

The conduction (Bloch) electrons in metals can move freely therefore they are not bounded by harmonic forces, so the angular frequency ω_o in (20.1.10) is 0. This gives

$$\hat{\alpha} = \frac{\omega_p^2}{N} \cdot \frac{1}{-\omega^2 + i\omega\gamma} \quad (20.4.10)$$

Introducing the average collision time τ with $\gamma = 1/\tau$, the complex susceptibility and permittivity are using (20.1.9) and (16.1.1) are

$$\hat{\chi} = \epsilon_o \cdot \frac{\omega_p^2}{-\omega^2 + i\omega\gamma} \quad (20.4.11)$$

$$\hat{n}^2(=\hat{\epsilon}) = 1 + \frac{\omega_p^2 \tau \epsilon_o}{i\omega(1 + i\omega\tau)} \quad (20.4.12)$$

From (20.4.12) we see⁷ that when $\omega < \omega_p$ then \hat{n} is complex with a large imaginary part so the attenuation of the wave is large, while in the region $\omega > \omega_p$ \hat{n} is real, i.e. the metal becomes transparent. In most metals, the plasma frequency is in the ultraviolet, making them shiny (reflective) in the visible range. Some metals, such as copper and gold, have electronic interband transitions in the visible range, whereby specific light energies (colors) are absorbed, yielding their distinct color. In semiconductors, the valence electron plasma frequency is usually in the deep ultraviolet which is why they too are reflective.

20.5 Non-linear effects

At very high electric field strengths the polarization density is not linear with the field strength. In most cases the susceptibility can be expanded in a power series, which for isotropic materials can be written as⁸:

$$\mathbf{P} = \epsilon_o \chi^{(1)} \cdot \mathbf{E} + \epsilon_o \chi^{(2)} \cdot \mathbf{E}^2 + \epsilon_o \chi^{(3)} \cdot \mathbf{E}^3 + \dots \quad (20.5.1)$$

⁷(20.4.12) may also be expressed with the conductivity σ , which is related to the plasma frequency by the formula $\omega_p^2 \tau = \sigma/\epsilon_o$ as

$$\hat{n}^2(=\hat{\epsilon}) = 1 + \frac{\sigma/\epsilon_o}{i\omega(1 + i\omega\tau)}$$

⁸For non-isotropic crystals a more complicated formula must be used. The j-th component of the polarization density is calculated from

$$\mathbf{P}_j = \epsilon_o \chi_{j,k}^{(1)} \cdot \mathbf{E}_k + \epsilon_o \chi_{j,k,l}^{(2)} \cdot \mathbf{E}_k \cdot \mathbf{E}_l + \epsilon_o \chi_{j,k,l,m}^{(3)} \cdot \mathbf{E}_k \cdot \mathbf{E}_l \cdot \mathbf{E}_m + \dots$$

where we used the common notation in which a summation must be performed to all of the indices occurring twice in a product.

In this formula the numbers in braces are the order of approximation. From the possible consequences⁹ we only show you one: frequency doubling.

If the electric field is in the form

$$E(t) = E_o \cdot e^{i\omega t - kr}$$

then the second term in (20.5.1) gives:

$$\mathbf{P} = \epsilon_o \chi^{(2)} E_o \cdot e^{i(\omega t - kr)} \cdot E_o \cdot e^{i(\omega t - kr)} = \epsilon_o \chi^{(2)} E_o^2 \cdot e^{i(2\omega t - 2kr)}$$

so the polarization contains a term which oscillates twice the frequency of the incoming light. Such frequency doubling is used e.g. in green laser pointers.

⁹ $\chi^{(2)}$ is responsible for frequency doubling, the Pockels effect used in nanosecond optical shutters, $\chi^{(3)}$ corresponds to the Kerr effect, which is a change in the refractive index of a material in response to an applied electric field.

Chapter 21

Appendices

Chapter 22

Quantum Mechanics

22.1 Measurement of the electromagnetic spectrum by a spectrometer

White light can be broken up unto a colored band by an optical prism or an optical grating.

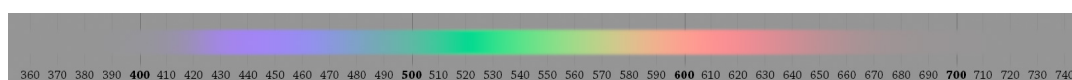


Figure 22.1: Visible spectrum of a white light source. The numbers mean the wavelengths in nm.

The reason for this is that the diffraction angle of the light depends on the wavelength (or frequency), which determines the color of the light. The light intensity at different wavelengths depends on the characteristics of the light source and the medium the light travels through. This wavelength - intensity (or frequency - intensity) relation is called the *spectrum* of the light.

Generally the properties (intensity, polarization state, etc) of light or any other form of electromagnetic radiation depend on the wavelength (or equivalently on the frequency). The corresponding $I(\lambda)$ or $I(\nu)$, etc functions are called the *spectrum* of the electromagnetic wave. In practice the independent variable used can also be the wave number ($k = \frac{2\pi}{\lambda}$) or any quantity which is directly proportional to the energy. ¹

¹The term *electromagnetic spectrum* means the range of all possible electromagnetic radiation, while the *electromagnetic spectrum of an object* means the characteristic distribution of electromagnetic radiation emitted or absorbed by the object.

Devices that measure this wavelength dependence are called *spectrometers*, *spectrophotometers*, *spectrographs*, *spectroscopes* or *spectral analyzers*.

Spectrophotometers measure the *absolute* light intensity as a function of λ or ν . The majority of spectrophotometers are used in spectral regions near, or in the visible spectrum. Other devices measure intensities *relative* to the spectrum of some standard.

Spectroscopes are often used in astronomy and some branches of chemistry. Early spectroscopes were simply prisms with graduations marking wavelengths of light. Modern spectroscopes generally use a diffraction grating, a movable slit, and some kind of photodetector, all automated and controlled by a computer. In microwave and radio frequencies the spectrum analyzer is a closely related electronic device.

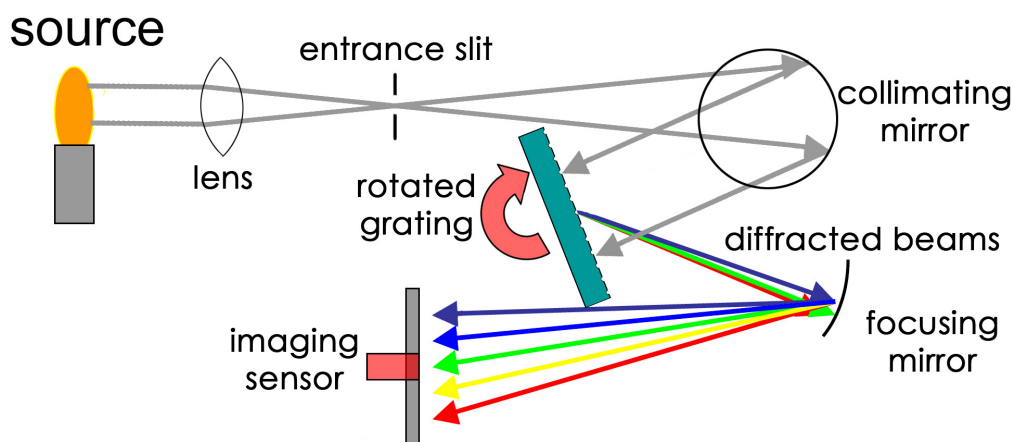


Figure 22.2: Schematics of a spectrometer. A parallel beam of light is projected to a reflective optical grating which diffracts different wavelengths to different directions and the photodetector measures the intensity of the light in a particular dimension. Rotating the grating changes what wavelength is diffracted to the detector.

Fig. 22.2 shows the schematic design of a spectrometer working in the optical range. The *reflective grating* nowadays produced by holographic techniques. It consists of closely spaced parallel lines (their period d is the *grating constant*) of varying reflectivity.

The operation of a reflective grating is shown in Fig 22.3. If the grating is perpendicular to the incoming parallel ray of light (the *angle of incidence*, i.e. the angle between the light ray and the surface normal, $\alpha_0 = 0$) and it is of wavelength λ , part of the light is reflected straight back, and in addition we can observe intensity maxima in the diffracted light waves in directions α_m , in which the condition

$$d \sin \alpha_m = m \lambda, \quad m = 1, 2, \dots \quad (22.1.1)$$

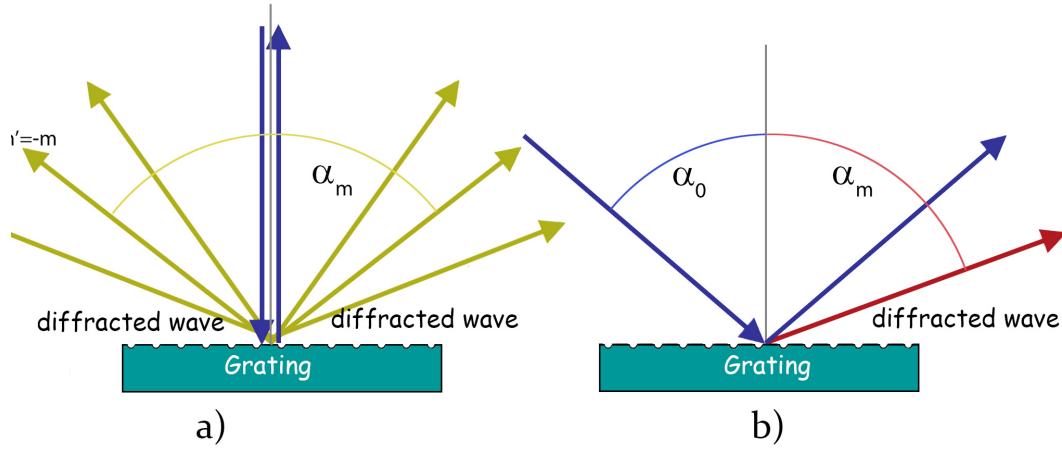


Figure 22.3: Diffraction on a reflective optical grating. The arrows denote the traveling direction of beams of light. a) Diffraction when the angle of incidence $\alpha_0 = 0$. Only four reflected beams, and the angle for the m -th and $m' = -m$ -th order are shown. b) One diffracted beam when the angle of incidence $\alpha_0 \neq 0$. The number m can be both positive and negative resulting in diffracted waves at both sides of the zero order beam.

is true. Here m is the diffraction order. If the angle of incidence is $\alpha_0 \neq 0$ the grating equation becomes:

$$d(\sin \alpha_0 + \sin \alpha_m) = m \lambda \quad m = 1, 2, \dots \quad (22.1.2)$$

For $m = 0$ it describes the *specular reflection* and is called the zero order. Fig. 22.4 shows what happens when the light diffracted is a mixture of multiple (discrete) wavelengths (colors).

22.2 The spread of a wave packet in time

In the double slit experiment when we put a detector to one of the slits to determine which slit the electron went through the interference pattern vanishes. Why is it so?

For the sake of simplicity in the following we will usually confine the description to one dimension. In one dimension (3.4.1) becomes

$$\mathcal{P}(x, dx) = |C \cdot \psi(x, t)|^2 \cdot dx \quad (22.2.1)$$

The shape of the wave function (wave packet) of an electron changes in time. If at $t = 0$ the shape of the wave packet was a sharply localized Gaussian function, e.g.

$$\psi(x, 0) = C e^{-\frac{x^2}{2\Delta x_0}} \cdot e^{-ikx}$$

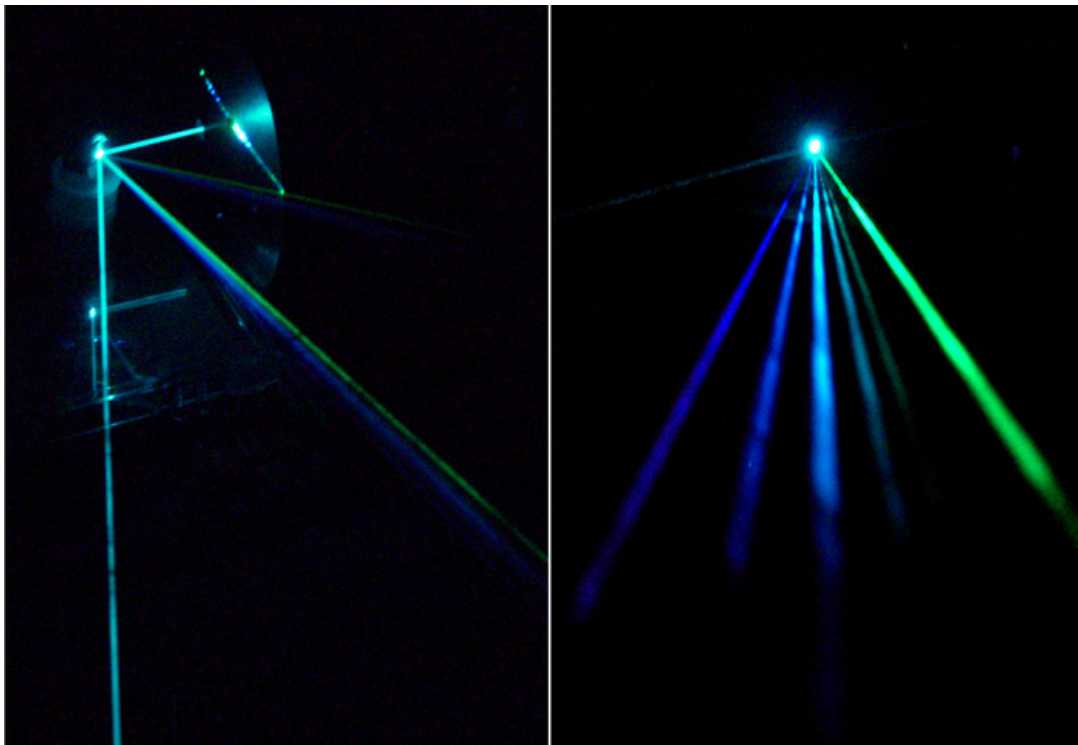


Figure 22.4: An argon laser beam consisting of multiple colors (wavelengths) strikes a silicon diffraction mirror grating and is separated into several beams, one for each wavelength. The wavelengths are (left to right) 458nm, 476nm, 488nm, 497nm, 502nm, 515nm.

where Δx is small with

$$|\psi(x, 0)|^2 = |C|^2 e^{-\frac{x^2}{\Delta x}}$$

then at any $t > 0$ time the width of the wave packet will grow according to the formula²

$$\Delta x(t) = \sqrt{\frac{(\Delta x_0)^2 + (\hbar t/m)^2}{\Delta x_0}}$$

In the light of the Heisenberg uncertainty relations this means that the wave packet will spread over time even without an external electric field. In Fig. 22.5 4 stages of this spreading is shown. At the same time because no external field is acting on the electrons

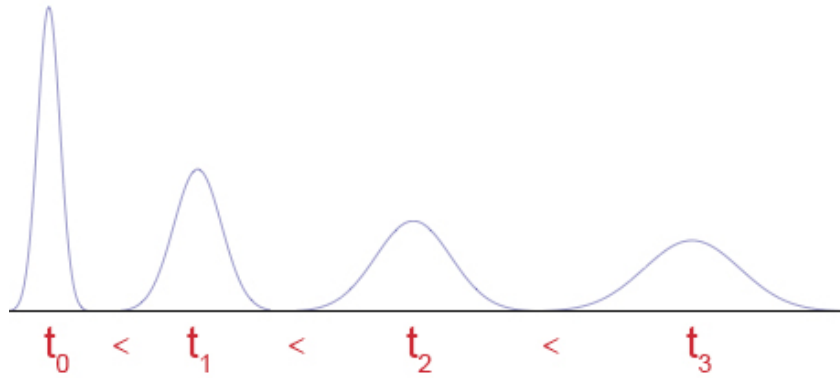


Figure 22.5: The spread of the wave function of an electron

the momentum uncertainty does not change, only the relative phases of the constituent waves change, therefore the product of uncertainties $\Delta x \cdot \Delta p$, which at $t = 0$ was exactly $\hbar/2$ also increases:

$$\Delta x(t) \cdot \Delta p > \frac{\hbar}{2} \quad \text{if } t > 0$$

This spread of the wave function would continue indefinitely long, unless the electron enters in an interaction with any other particle (e.g. collides with a photon). In the interaction the position uncertainty of the electron shrinks and the wave function becomes sharply localized again. This fast shrinking of the wave function cannot be described by quantum mechanics.

²This can be derived with a long and tedious although straightforward manner with (22.4.3) and (22.4.4) and taking the absolute square of the $\psi(x, t)$ function.

22.3 Derivation of the Compton formula

A photon with wavelength λ collides with an electron e in an atom, which is treated as being at rest. The collision causes the electron to recoil, and a new photon γ' with wavelength λ' emerges at angle θ from the photon's incoming path as shown in Fig 22.6 a). According to the conservation laws both the energy and the momentum must be preserved in the collision:

$$\mathcal{E}_e + \mathcal{E}_{ph} = \mathcal{E}'_e + \mathcal{E}'_{ph} \quad (22.3.1)$$

$$0 + \mathbf{p}_{ph} = \mathbf{p}'_e + \mathbf{p}'_{ph} \quad (22.3.2)$$

where $\mathcal{E}_e = m_e c^2$ and \mathcal{E}'_e are the energy of the electron before and after the collision and p_e is the momentum of the electron after the collision, while $\mathcal{E}_{ph} = h\nu$ and $\mathcal{E}'_{ph} = h\nu'$ are the energies, \mathbf{p}_{ph} and \mathbf{p}'_{ph} are the momenta of the incoming and outgoing photons respectively. From special relativity we know that the magnitude of the momentum of a photon is

$$p_{ph} = \frac{\mathcal{E}}{c}$$

and the connection between the energy and the momentum of the electron is³

$$\mathcal{E}^2 - p^2 c^2 = m^2 c^4$$

The energy loss of the photon then

$$h(\nu - \nu') = \sqrt{p_e'^2 + m^2 c^4} - m^2 c^2 \quad (22.3.3)$$

Solving for the term concerning the post-collision momentum of the electron gives

$$p_e'^2 = (h\nu + m_e c^2 - h\nu') \quad (22.3.4)$$

On the other hand from the conservation of momentum

$$\mathbf{p}'_e = \mathbf{p}_{ph} - \mathbf{p}'_{ph} \quad (22.3.5)$$

therefore

$$\begin{aligned} p_e'^2 &= (\mathbf{p}_{ph} - \mathbf{p}'_{ph}) \cdot (\mathbf{p}_{ph} - \mathbf{p}'_{ph}) = \\ &= p_{ph}^2 + p_{ph}'^2 - 2 p_{ph} p_{ph}' \cos\theta = \\ &= h^2 \cdot (\nu^2 + \nu'^2 - 2\nu\nu' \cos\theta) \end{aligned}$$

³Substituting $p_e = 0$ gives the well known relation of $\mathcal{E} = m c^2$

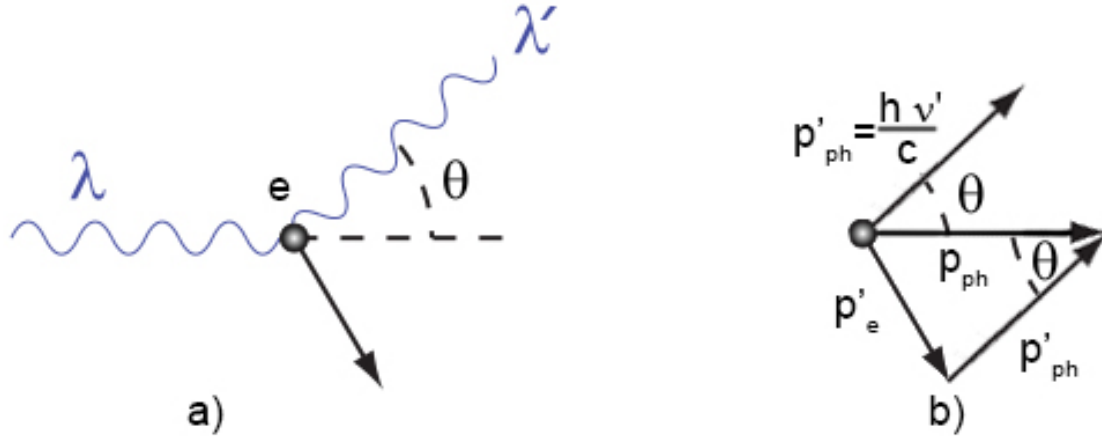


Figure 22.6: Dynamics of the Compton effect. a) visualization, b) momentum diagram for the calculation

Substituting this formula for $p_e'^2$ into the left hand side of (22.3.4) and dividing both sides with $2h\nu\nu'c$

$$\frac{c}{\nu'} - \frac{c}{\nu} = \frac{h}{m_e c} (1 - \cos\theta) \quad (22.3.6)$$

Then the final result using that $\lambda = c/\nu$

$$\lambda' - \lambda = \frac{h}{m_e c} (1 - \cos\theta), \quad (22.3.7)$$

22.4 Uncertainty relations for a wave packet

Localized electronic states are described by a wave packet. According to the Fourier-theorem any function, therefore our wave packet as well, can be written as a sum of an infinite number of harmonic functions. For periodic functions this is a sum of an infinite series, for non-periodic functions it is an integral. For the sake of simplicity we will stay in one dimension. therefore our wave packet will be a linear combination of waves of the form

$$u(x, t) = \mathcal{A}(k) e^{i(\omega(k)t - kx)},$$

where the wave with $k > 0$ travels in the positive, the ones with $k < 0$ in the negative x direction with a k dependent *phase velocity* of

$$c := v_{ph} = \frac{\omega(k)}{k} \quad (22.4.1)$$

The $\omega(k)$ function for a particle⁴

$$\omega(k) = \frac{\mathcal{E}_{kin}}{\hbar} = \frac{\hbar k^2}{2m_e}$$

It follows that the *group velocity* of a constituent wave

$$v_g = \frac{d\omega(k)}{dk} = \frac{\hbar k}{m_e} \quad (22.4.2)$$

depends on the actual k : different harmonic components of the originally localized wave function travel with different velocities their relative phases change in time. As a result the wave packet becomes wider and wider, it disperses. The (3.2.1d) relation is a dispersive one.

The group velocity of electromagnetic waves in vacuum is the same as their phase velocity, therefore an electromagnetic wave packet can only disperse if the medium is dispersive. *Material waves* do not need a dispersive medium to spread over time, they do it in vacuum too.

Our wave packet therefore is written as

$$\psi(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{A}(k) e^{i(\omega(k)t - kx)} dk, \quad (22.4.3)$$

where $\mathcal{A}(k)$ is the k dependent amplitude.

You can recognize that (22.4.3) is the Fourier transform of the function⁵

$$\mathcal{A}(k) e^{i\omega(k)t}$$

⁴using the de Broglie relations (3.2.1c) and (3.2.1d)

⁵There are some conventions regarding the Fourier transform. We use the one called unitary angular frequency convention.

In one dimension the $\xi(x)$ Fourier transform of the function $f(k)$ is denoted with $\mathcal{F}(f(k))$ its inverse with $\mathcal{F}^{-1}(\xi(x))$:

$$\begin{aligned} \xi(x) &:= \mathcal{F}(f(k)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(k) e^{-ikx} dk, \text{ and} \\ \mathcal{F}^{-1}(\xi(k)) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \xi(x) e^{ikx} dx \end{aligned}$$

The $\mathcal{A}(k)$ amplitude function then can be calculated by the inverse Fourier transform at $t = 0$:

$$\mathcal{A}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x, 0) e^{ikx} dx \quad (22.4.4)$$

As an example let us suppose that at $t = 0$ the wave packet describing an electron localized around the origin has a Gaussian shape

$$\psi(x, 0) = e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-ik_o x}$$

where σ is a positive real number, whose square is twice the width of the wave packet:

$$2(\Delta x)^2 \equiv \sigma^2. \quad (22.4.5)$$

The first exponential is the envelope which determines the localized shape of the function, the second one with the imaginary exponent describes the motion of the center of the wave packet with a constant momentum of $p_o = \hbar k_o$. This second part when combined with $e^{i\omega t}$ is what makes the function a wave. Substituting into (22.4.4)

$$\begin{aligned} \mathcal{A}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2} - ik_o x} e^{ikx} dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2} + i(k - k_o)x} dx \end{aligned}$$

$\mathcal{A}(k)$ can be obtained from a table of Fourier integrals:

$$\mathcal{A}(k) = \frac{1}{\sqrt{2\pi}} \sigma e^{-\frac{\sigma^2 (k - k_o)^2}{2}} \quad (22.4.6)$$

Substituting back into (22.4.3) we can calculate the shape of the wave function at an arbitrary time.

The $\mathcal{A}(k)$ function describes the localization of the K values of the wave packet around k_o . Because it also has a Gaussian shape it may also be written as

$$\begin{aligned} \mathcal{A}(k) &= \frac{1}{\sqrt{2\pi}} \sigma e^{-\frac{(k - k_o)^2}{2\sigma_k^2}}, \text{ where} \\ \sigma_k &= \frac{1}{\sigma} \end{aligned}$$

and because (c.f. (22.4.5))

$$2 (\Delta k)^2 \equiv \sigma_k^2$$

the product of the square of the two widths

$$\sigma^2 \cdot \sigma_k^2 = 4 \cdot \Delta x \cdot \Delta k \quad \text{from where}$$

$$\Delta x \cdot \Delta k = \frac{1}{2},$$

and using the (3.2.1c) de Broglie relations

$$\Delta x \cdot \Delta p = \frac{\hbar}{2} \quad (22.4.7)$$

A more general derivation would prove that for any functional shape of a wave packet the product of σ^2 and σ_k^2 cannot be less than 1, but it can be larger than that, therefore

$$\Delta x \cdot \Delta p \geq \frac{\hbar}{2} \quad (22.4.8)$$

22.5 The linear harmonic oscillator - Analitical solution

The Schrödinger equation of the linear harmonic oscillator is

$$-\frac{\hbar^2}{2m} \frac{d^2 \varphi}{dx^2} + \frac{1}{2} m \omega^2 x^2 \varphi = \mathcal{E} \varphi$$

Reorder the equation first:

$$\frac{d^2 \varphi}{dx^2} + \frac{2m}{\hbar^2} \left(\mathcal{E} - \frac{1}{2} m \omega^2 x^2 \right) \varphi = 0 \quad (22.5.1)$$

With introducing

$$\xi := \sqrt{\frac{m\omega}{\hbar}} x \quad \text{and} \quad k := \frac{2\mathcal{E}}{\hbar\omega}$$

the equation takes the form

$$\varphi'' + (k - \xi^2) \varphi = 0 \quad (22.5.2)$$

If $x \rightarrow \infty$ then $\xi \rightarrow \infty$. In this case k is negligible compared to ξ and for very large x , i.e. very large ξ we have an *asymptotic* equation:

$$\varphi''(\xi) - \xi^2 \varphi(\xi) = 0, \quad \text{when } x \rightarrow \infty$$

The solution of this asymptotic equation is

$$\varphi(\xi) = e^{\pm \frac{1}{2} \xi^2}$$

For physical solutions the total probability and correspondingly the scalar product of φ with itself must be finite, which requires that $\lim_{x \rightarrow \infty} \varphi(x) = 0$. This is equivalent to $\lim_{\xi \rightarrow \infty} \varphi(\xi) = 0$. This allows only the asymptotic solution with the negative sign in the exponent. Therefore try the solution for the non asymptotic equation in the form:

$$\varphi(\xi) = u(\xi) e^{-\frac{1}{2} \xi^2} \quad (22.5.3)$$

The derivatives then (now ' will denote the derivative with respect to ξ : $\varphi'' \equiv \frac{d\varphi}{d\xi}$)

$$\begin{aligned} \varphi' &= (u' - \xi u) e^{-\frac{1}{2} \xi^2} \\ \varphi'' &= (u'' - \xi u' - u - \xi u' + \xi^2 u) e^{-\frac{1}{2} \xi^2} \end{aligned}$$

With this after dividing the equation with the non zero exponential factor (22.5.2) becomes:

$$u'' - 2\xi u' + (k-1)u = 0 \quad (22.5.4)$$

We will try to satisfy this equation with a polynomial

$$u(\xi) = \sum_{r=0}^{\infty} c_r \xi^r \quad (22.5.5)$$

After substitution:

$$\sum_{r=0}^{\infty} [(r+2)(r+1)c_{r+2} - 2c_r + (k-1)c_r] \xi^r = 0$$

This equation can only be true for every possible value of ξ if the term inside the square brackets is itself 0 for every r :

$$(r+2)(r+1)c_{r+2} - 2c_r + (k-1)c_r = 0$$

$$c_{r+2} = \frac{2r+1-k}{(r+2)(r+1)} c_r \quad \text{where } r = 0, 1, 2, \dots \quad (22.5.6)$$

Because the index jumps by 2 we may select both coefficients c_0 and c_1 independently⁶. Therefore the sum (22.5.5) may be separated to two sums, one for the indexes 0,2,4, etc, the other one for 1,3,5, etc. The first one is an even function of ξ the second one is an odd function of it.

φ is	c_0	c_1	$\varphi(0)$	$\varphi'(0)$
even	1	0	1	0
odd	0	1	0	1

Still it would be impossible to calculate all coefficients. But we know (see (22.5.3)) that how φ should look like for large ξ (asymptotically). We also know that φ must be square-integrable. See if we can satisfy this condition.

For large ξ (22.5.6) simplifies to

$$c_r + 2 \approx \frac{2}{r} c_r \quad \text{so}$$

$$u \approx \sum_{r=0}^{\infty} \frac{(2\xi^2)^r}{r!} = e^{2\xi^2}$$

But then (22.5.3) becomes:

$$\varphi = e^{2\xi^2} e^{-\frac{1}{2}\xi^2} = e^{\frac{3}{2}\xi^2}$$

which is *not* square-integrable. Therefore we must not allow an infinite number of non-zero c_r coefficients in our result. Let us suppose therefore that in (22.5.6) there is an index n for which $c_{n+2} = 0$, while $c_n \neq 0$:

$$\frac{2n+1-k}{(n+2)(n+1)} = 0$$

from which

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

i.e.

$$c_{r+2} = \frac{2(r-n)}{(r+2)(r+1)} c_r \quad \text{and}$$

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right) = h\nu \left(n + \frac{1}{2} \right) \quad (22.5.7)$$

⁶The reason behind this is simply that (22.5.4) is a second order differential equation whose general solution requires two constants.

The polynomial (see (22.5.5))

$$H_n(\xi) = \sum_{r=0}^n c_r \xi^r \quad \text{where} \quad (22.5.8)$$

$$c_{r+2} = \frac{2(r-n)}{(r+2)(r+1)} c_r$$

is called the *Hermite polynomial*⁷ The $\varphi_n(x)$ eigenfunctions then

$$\varphi_n(x) = C_n H_n \left(\sqrt{\frac{m\omega}{\hbar}} x \right) e^{-\frac{1}{2} \frac{m\omega^2}{\hbar} x^2}, \quad (22.5.9)$$

where C_n is the normalization constant. The first 4 Hermite polynomials are

$$\begin{aligned} H_0 &= 1 & H_1 &= x \\ H_2 &= 1 - 2x^2 & H_3 &= x - \frac{2}{3}x^3 \\ H_4 &= 1 - 4x^2 + \frac{1}{3}x^4 \end{aligned}$$

22.6 The linear harmonic oscillator - Ladder operators

To determine the eigenvalues of the linear harmonic oscillator we have to solve the corresponding eigenvalue equation (the 'hat' (^) symbol denotes an operator):

$$\hat{\mathcal{H}}\varphi(\mathbf{r}) = \mathcal{E}\varphi(\mathbf{r}) \quad (22.6.1)$$

where

$$\hat{\mathcal{H}} = \frac{\hat{\mathbf{p}}^2}{2m_e} + \frac{1}{2}m\omega^2\hat{\mathbf{r}}^2 \quad (22.6.2)$$

In the Schrödinger picture of quantum mechanics the operators $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ are:

$$\hat{\mathbf{r}} = \mathbf{r} \cdot \quad (22.6.3)$$

$$\hat{\mathbf{p}} = \frac{\hbar}{i} \nabla \quad (22.6.4)$$

⁷Unfortunately there are at least two other definitions for the Hermite polynomials. The one we selected here is the one best suited to our derivation.

For the sake of simplicity in one dimension:

$$\hat{\mathcal{H}} = \frac{\hat{p}^2}{2m_e} + \frac{1}{2}m_e\omega^2\hat{x}^2 \quad (22.6.5)$$

The corresponding Schrödinger equation then becomes

$$-\frac{\hbar^2}{2m_e} \frac{d^2\varphi(x)}{dx^2} + \frac{1}{2}m_e\omega^2 x^2 \varphi(x) = \mathcal{E}\varphi \quad (22.6.6)$$

The solution to this equation is far from simple. Furthermore using the Schrödinger equation suggests that quantum mechanical eigenvalue problems are differential equations. In fact, however, this is not true. To illustrate this we will solve (22.6.5) in one dimension and determine the E eigenvalues without using any differential calculus, even without any advanced mathematics whatsoever (unless of course the algebraic calculations with non-commuting quantities is considered „advanced” mathematics)!

We will only use (algebraic) operator equations so we will not use the mathematical form of operators \hat{x} and \hat{p} . But for our purposes (22.6.5) is not enough we need something more from quantum mechanics, namely the commutativity relation of quantum mechanical operators, called the commutator. The definition of the commutator of any two quantum mechanical operator $\hat{\mathcal{A}}, \hat{\mathcal{B}}$ is:

$$[\hat{\mathcal{A}}, \hat{\mathcal{B}}] := \hat{\mathcal{A}}\hat{\mathcal{B}} - \hat{\mathcal{B}}\hat{\mathcal{A}} \quad (22.6.7)$$

The commutator may or may not be 0. When the commutator of two operator is $\pm i\hbar$ they are called canonically conjugate operators. The operators \hat{x} and \hat{p} are canonically conjugate quantities, because from (22.6.3):

$$[\hat{x}, \hat{p}] = i\hbar \quad (22.6.8)$$

The *state of the electron* is determined by the $\varphi(x)$ wave function. If we had known the wave function we also knew the solution of (22.6.5). in the followings we will determine the energy eigenvalues *without* determining the wave function itself. Because we know the solution of the classical mechanical problem of the linear harmonic oscillator we can easily see that (22.6.5) may be written in a more symmetrical form by introducing to new operators $\hat{\mathcal{P}}$ and $\hat{\mathcal{Q}}$ with equations:

$$\hat{\mathcal{X}} := \sqrt{\frac{m_e\omega}{\hbar}} \hat{x} \quad \text{and} \quad \hat{\mathcal{P}} := \sqrt{\frac{1}{m_e\omega\hbar}} \hat{p} \quad (22.6.9)$$

we can see by substitution into (22.6.5) and (22.6.8):

$$\frac{\hbar\omega}{2} (\hat{\mathcal{P}}^2 + \hat{\mathcal{Q}}^2) \varphi(x) = E\varphi(x) \quad (22.6.10)$$

$$[\hat{\mathcal{Q}}, \hat{\mathcal{P}}] = i \quad (22.6.11)$$

Those who know the result we are seeking may find (22.6.10) interesting. Unfortunately (22.6.10) is not much simpler than the original equation was. \hat{Q} and \hat{P} are just intermediate forms to make the following calculations easier. If we were using (commuting) complex numbers and not non-commuting operators, then inside the braces in (22.6.10) we may recognize the product of a sum and a subtraction⁸:

$$Q^2 + P^2 = (Q + iP)(Q - iP)$$

Therefore let us introduce the following two new operators:

$$\hat{a} := \frac{\hat{Q} + i\hat{P}}{\sqrt{2}} \quad \text{and} \quad \hat{a}^+ := \frac{\hat{Q} - i\hat{P}}{\sqrt{2}} \quad (22.6.12)$$

Substituting into (22.6.11) it is easy to see that

$$[\hat{a}, \hat{a}^+] = 1 \quad (22.6.13)$$

and because

$$\hat{Q} = \frac{1}{\sqrt{2}}(\hat{a} + \hat{a}^+) \quad (22.6.14)$$

$$\hat{P} = \frac{1}{i\sqrt{2}}(\hat{a} - \hat{a}^+) \quad (22.6.15)$$

so in (22.6.10)

$$\begin{aligned} & \frac{\hbar\omega}{4} ((\hat{a} + \hat{a}^+)(\hat{a} + \hat{a}^+) - 2(\hat{a} - \hat{a}^+)(\hat{a} - \hat{a}^+)) = \\ & \frac{\hbar\omega}{4} (\hat{a}\hat{a} + \hat{a}\hat{a}^+ + \hat{a}^+\hat{a} + \hat{a}^+\hat{a}^+ - \hat{a}\hat{a} + \hat{a}\hat{a}^+ + \hat{a}^+\hat{a} - \hat{a}^+\hat{a}^+) = \\ & \frac{\hbar\omega}{4} (2\hat{a}\hat{a}^+ + 2\hat{a}^+\hat{a}) \end{aligned}$$

Using (22.6.13) $\hat{a}\hat{a}^+ = \hat{a}^+\hat{a} + 1$

$$\frac{\hbar\omega}{4} (2\hat{a}\hat{a}^+ + 2\hat{a}^+\hat{a}) = \frac{\hbar\omega}{2} (2\hat{a}^+\hat{a} + 1)$$

therefore

$$\hat{\mathcal{H}} \equiv \hbar\omega(\hat{a}^+\hat{a} + \frac{1}{2}) \quad \text{so} \quad (22.6.16)$$

$$\hbar\omega(\hat{\mathcal{N}} + \frac{1}{2})\varphi(x) = E\varphi(x) \quad (22.6.17)$$

⁸This formula is not instantly usable for us, because \hat{Q} and \hat{P} are non-commuting operators, but may point us into the right direction.

where we introduced the notation:

$$\hat{\mathcal{N}} := \hat{a}^+ \hat{a}$$

Because of the connection between $\hat{\mathcal{H}}$ and $\hat{\mathcal{N}}$ both operators have the same eigenfunctions. Therefore if we determine the eigenvalues for $\hat{\mathcal{N}}$ we will get the eigenvalues of $\hat{\mathcal{H}}$ too. Let φ an eigenfunction of $\hat{\mathcal{N}}$ with eigenvalue λ :

$$\hat{\mathcal{N}}\varphi = \lambda\varphi \quad (22.6.18)$$

Multiplying both sides with \hat{a}^+ from the left and using the definition of $\hat{\mathcal{N}}$:

$$\hat{a}^+(\hat{a}^+\hat{a})\varphi = \lambda\hat{a}^+\varphi \quad (22.6.19)$$

because λ is a number it commutes with \hat{a}^+ .

Using (22.6.13)

$$\hat{a}^+(\hat{a}^+\hat{a}) = \hat{a}^+(\hat{a}\hat{a}^+ - 1) = (\hat{a}^+\hat{a} - 1)\hat{a}^+ = (\hat{\mathcal{N}} - 1)\hat{a}^+$$

after reordering

$$\hat{\mathcal{N}}(\hat{a}^+\varphi) = (\lambda + 1)(\hat{a}^+\varphi) \quad (22.6.20)$$

i.e. if φ_λ is an eigenfunction of $\hat{\mathcal{N}}$ with the eigenvalue λ then $\varphi_{\lambda+1} := \hat{a}^+\varphi_\lambda$ is also either an eigenfunction of $\hat{\mathcal{N}}$ with the eigenvalue $(\lambda + 1)$, or it may only differ from $\varphi_{\lambda+1}$ by a constant factor. If we can find any of the eigenfunctions of $\hat{\mathcal{N}}$ then we can construct an other eigenfunction by applying the operator \hat{a}^+ to it. This way we can construct an infinite series of eigenfunctions and eigenvalues. On a similar way by multiplying both sides with \hat{a} from the left:

$$\begin{aligned} \hat{a}(\hat{a}^+\hat{a})\varphi &= \lambda\hat{a}\varphi \\ \hat{a}(\hat{a}^+\hat{a}) &= (\hat{a}\hat{a}^+)\hat{a} = (\hat{a}^+\hat{a} + 1)\hat{a} = (\hat{\mathcal{N}} + 1)\hat{a} \\ \hat{\mathcal{N}}(\hat{a}\varphi_\lambda) &= (\lambda - 1)(\hat{a}\varphi_\lambda) \end{aligned} \quad (22.6.21)$$

i.e. if φ_λ is an eigenfunction of $\hat{\mathcal{N}}$ with the eigenvalue λ then $\varphi_{\lambda-1} := \hat{a}\varphi_\lambda$ is also either an eigenfunction of $\hat{\mathcal{N}}$ with the eigenvalue $(\lambda - 1)$, or it may only differ from $\varphi_{\lambda-1}$ by a constant factor. As before we can construct an infinite series of eigenfunctions each with an eigenvalue 1 less then the previous one. But this poses a problem. From (22.6.17)

$$E = \hbar\omega\left(\lambda + \frac{1}{2}\right)$$

The infinitely decreasing series of λ values correspond to an infinitely decreasing series of energies, which may lead even to impossible negative energy values, *unless* there exists a φ_0 eigenfunction with an eigenvalue of $\lambda = 0$:

$$\hat{\mathcal{N}}\psi_0 = 0\psi_0 \quad \text{and} \quad \psi_0 \neq 0$$

Therefore the possible λ values are

$$\lambda = 0, 1, 2, \dots$$

and then (replacing λ with the usual n):

$$E = \hbar\omega\left(n + \frac{1}{2}\right) \quad n = 0, 1, 2, 3, \dots \quad (22.6.22)$$

22.7 1 dimensional potential well

$$V(x) = \begin{cases} V_0, & |x| \geq \frac{L}{2} \\ 0, & |x| < \frac{L}{2} \end{cases} \quad (22.7.1)$$

Again we will solve the Schrödinger equation piecewise and use boundary conditions to connect the pieces.

The solutions in the three regions are similar to the wave functions we used section 3.5:

$$\begin{aligned} \varphi_I(x) &= A e^{iqx} + B e^{-iqx} \\ \varphi_{II}(x) &= C e^{ikx} + D e^{-ikx} \\ \varphi_{III}(x) &= E e^{iqx} + F e^{-iqx} \end{aligned} \quad (22.7.2)$$

where

$$\begin{aligned} q &= \sqrt{\frac{2m(\mathcal{E} - V_0)}{\hbar}} \quad \text{and} \\ k &= \sqrt{\frac{2m\mathcal{E}}{\hbar}} \end{aligned}$$

The boundary conditions for the continuity of the wave function and its derivative:

$$\begin{aligned} \varphi_I\left(-\frac{L}{2}\right) &= \varphi_{II}\left(-\frac{L}{2}\right) \\ \varphi_{II}\left(\frac{L}{2}\right) &= \varphi_{III}\left(\frac{L}{2}\right) \\ \varphi'_I\left(-\frac{L}{2}\right) &= \varphi'_{II}\left(-\frac{L}{2}\right) \\ \varphi'_{II}\left(\frac{L}{2}\right) &= \varphi'_{III}\left(\frac{L}{2}\right) \end{aligned}$$

These boundary conditions present only 4 equations for the 6 unknowns A, B, C, D, E and F . That is either we can assign any values for two of the parameters or there must

exist other conditions for the wave function. Again we must distinguish between the two cases where the total energy of the particle is larger than V_0 or smaller than V_0 .

In the first case we must decide from where the particle comes from. If it comes from the left then $F = 0$ and we may set A to any value, if it arrives from right then $A = 0$ and the value of F is arbitrary. In any case the particle will not be trapped in the potential well, and this is the behavior we would expect from a classical particle too. There will be no constraints for the possible energy values of these *unbounded* states. We say that the *energy spectrum* of unbounded states is continuous and not quantized.

More interesting is the case when $\mathcal{E} < V_0$. Then q becomes imaginary

$$q \equiv i\alpha = \sqrt{\frac{2m(V_0 - \mathcal{E})}{\hbar}}$$

and the exponent of the wave functions in φ_I and φ_{III} will be real. Because the wave function must be square-integrable only the exponentially decreasing terms can differ from 0, i.e.

$$\begin{aligned} \varphi_I(x) &= B e^{\alpha x} & x &\leq -\frac{L}{2} \\ \varphi_{II}(x) &= C e^{i k x} + D e^{-i k x} & -\frac{L}{2} &\leq x \leq \frac{L}{2} \\ \varphi_{III}(x) &= E e^{-\alpha x} & x &\geq \frac{L}{2} \end{aligned} \quad (22.7.3)$$

We reduced the number of unknowns to 4 and we have 4 independent homogeneous equations. For such a system non zero solutions may only exist if the determinant of the equation is 0 and in that case an infinite number of solutions exist. This is good news, because it makes possible to select one of the unknowns in a way that the wave functions is normalized!

As it turns out this is one of those occasions when working with sine and cosine functions is easier than with complex exponential ones, therefore we will write

$$\varphi_{II}(x) = G \sin k x + H \cos k x \quad |x| \leq \frac{L}{2} \quad (22.7.4)$$

The two new constants are $G = C + D$ and $H = C - D$ ⁹. The 2 boundary conditions at

⁹Using the well known formula: $e^{i\alpha} = \cos \alpha + i \sin \alpha$ this is easy to prove.

$-L/2$ and the other two at $L/2$ with these then are:

$$\begin{aligned} B e^{-\alpha L/2} &= G \sin(-k L/2) + H \cos(-k L/2) \\ \alpha B e^{-\alpha L/2} &= -k G \cos(-k L/2) - k H \sin(-k L/2) \\ \text{and} \\ E e^{-\alpha L/2} &= G \sin(k L/2) + H \cos(k L/2) \\ -\alpha E e^{-\alpha L/2} &= -k G \cos(k L/2) - k H \sin(k L/2) \end{aligned}$$

On the right hand side of these equations we see a sum of two *independent* functions, one is even $\varphi^{(1)}(-x) = \varphi^{(1)}(x)$ the other one is odd $\varphi^{(2)}(-x) = -\varphi^{(2)}(x)$. The probability density is proportional with $|\varphi(x)|^2$ so both of these are acceptable. (See Fig 3.15) We can satisfy the boundary conditions with either of them. The boundary conditions for the even *cos* function are

$$B e^{-\alpha L/2} = H \cos(-k L/2) \quad (22.7.5a)$$

$$\alpha B e^{-\alpha L/2} = -k H \sin(-k L/2) \quad (= +H \sin(k L/2)) \quad (22.7.5b)$$

$$E e^{-\alpha L/2} = H \cos(k L/2) \quad (22.7.5c)$$

$$-\alpha E e^{-\alpha L/2} = -k H \sin(k L/2) \quad (22.7.5d)$$

If we are not interested in the wave functions themselves, but want to determine the possible energy eigenvalues only, then we get rid of the unknowns by either dividing (22.7.5b) with (22.7.5a), or (22.7.5d) with (22.7.5c):

$$\alpha = k \tan \frac{k L}{2} \quad (22.7.6a)$$

On a similar way for the odd *sin* function we obtain:

$$\alpha = -k \cot \frac{k L}{2} \quad (22.7.6b)$$

Both α and k is related to the \mathcal{E} energy. From their definitions it is easy to see that

$$\alpha^2 + k^2 = \frac{2 m V_0}{\hbar^2} \quad (22.7.7)$$

As you can see despite its simplicity this is a potential for which no analytical solution exists. We may either use graphical or numerical methods to determine the energy values. If we draw the $\alpha(k)$ function from (22.7.6a) (or from (22.7.6b)) and the (22.7.7) curve in the same $k - \alpha$ coordinate system (see Fig. 22.7) then their intersection provides the possible energy values. The number of the possible energy levels in this case is finite.

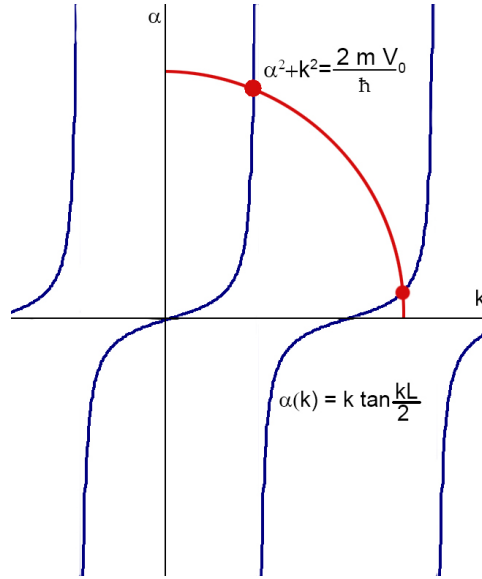


Figure 22.7: Solution of (22.7.7)

22.8 Derivation of Perturbation theory formulas

First we try to write the unknown solution as a linear combination of the known eigenfunctions of the *unperturbed* Schrödinger equation:

$$\psi(x, t) = \sum_n C_n(t) \varphi_n(x) e^{-\frac{i}{\hbar} \mathcal{E}_n t}$$

where the $C_n(t)$ coefficients are time dependent complex numbers (functions). Details of the calculation are in Appendix 22.8 Substitute $\psi_n(x, t)$ into the Schrödinger equation:

$$\begin{aligned} \sum_n C_n(t) e^{-\frac{i}{\hbar} \mathcal{E}_n t} \left(\left[-\frac{\hbar^2}{2m} \frac{d^2 \varphi_n}{dx^2} + V \varphi_n \right] + K \right) &= \\ &= \sum_n i \hbar \frac{d(C_n(t) e^{-\frac{i}{\hbar} \mathcal{E}_n t})}{dt} \varphi_n \end{aligned}$$

The expression in the square bracket is the left hand side of the stationary Schrödinger equation for φ_n , therefore it is equal to \mathcal{E}_n . After reordering this gives:

$$\sum_n \left(C_n(t) e^{-\frac{i}{\hbar} \mathcal{E}_n t} (\mathcal{E}_n + K) \varphi_n(x) - i \hbar \frac{dC_n(t) e^{-\frac{i}{\hbar} \mathcal{E}_n t}}{dt} \varphi_n(x) \right) = 0$$

The term containing the derivative with respect to time is

$$i \hbar \frac{dC_n(t) e^{-\frac{i}{\hbar} \mathcal{E}_n t}}{dt} \varphi_n(x) = \left(i \hbar \frac{dC_n(t)}{dt} + C_n(t) \mathcal{E}_n \right) e^{-\frac{i}{\hbar} \mathcal{E}_n t} \varphi_n(x)$$

Therefore

$$\sum_n \left(C_n(t) K(x, t) - i \hbar \frac{dC_n(t)}{dt} \right) \varphi_n(x) e^{-\frac{i}{\hbar} \mathcal{E}_n t} = 0$$

Now we can use the orthogonality of the eigenfunctions $\varphi_n(x)$. Multiply this equation with $\left(\varphi_m(x) e^{-\frac{i}{\hbar} \mathcal{E}_m t} \right)^* = \varphi_m^*(x) e^{+\frac{i}{\hbar} \mathcal{E}_m t}$ and integrate for the whole space:

$$\int_{-\infty}^{\infty} \varphi_m^*(x) e^{+\frac{i}{\hbar} \mathcal{E}_m t} \left[\sum_n \left(C_n(t) K(x) - i \hbar \frac{dC_n(t)}{dt} \right) \varphi_n(x) e^{-\frac{i}{\hbar} \mathcal{E}_n t} \right] dx = 0$$

The order of the summation and integration can be exchanged

$$\begin{aligned} \sum_n \left[\int_{-\infty}^{\infty} \varphi_m^*(x) e^{+\frac{i}{\hbar} \mathcal{E}_m t} \left(C_n(t) K(x) - i \hbar \frac{dC_n(t)}{dt} \right) \varphi_n(x) e^{-\frac{i}{\hbar} \mathcal{E}_n t} dx \right] &= 0 \\ \sum_n \left[C_n(t) e^{+\frac{i}{\hbar} (\mathcal{E}_m - \mathcal{E}_n) t} \int_{-\infty}^{\infty} \varphi_m^*(x) K(x) \varphi_n(x) dx \right] &- \\ - \sum_n \left[i \hbar \frac{dC_n(t)}{dt} e^{+\frac{i}{\hbar} (\mathcal{E}_m - \mathcal{E}_n) t} \int_{-\infty}^{\infty} \varphi_m^*(x) \varphi_n(x) dx \right] &= 0 \end{aligned}$$

Because φ_n and φ_m are orthogonal for $n \neq m$ only the term where $m = n$ remains and the second sum evaluates to

$$\sum_n \left[i \hbar \frac{dC_n(t)}{dt} e^{+\frac{i}{\hbar} (\mathcal{E}_m - \mathcal{E}_n) t} \int_{-\infty}^{\infty} \varphi_m^*(x) \varphi_n(x) dx \right] = i \hbar \frac{dC_m(t)}{dt},$$

therefore

$$\frac{dC_m(t)}{dt} = -\frac{i}{\hbar} \sum_n \left[C_n(t) e^{+\frac{i}{\hbar} (\mathcal{E}_m - \mathcal{E}_n) t} \int_{-\infty}^{\infty} \varphi_m^*(x) K(x) \varphi_n(x) dx \right] \quad (22.8.1)$$

The integral in this equation is called the m, n -th *matrix element* of the potential $K(x, t)$ and is denoted with $K_{mn}(t)$:

$$K_{mn}(t) := \int_{-\infty}^{\infty} \varphi_m^*(x) K(x, t) \varphi_n(x) dx$$

Introducing $\omega_{mn} = (\mathcal{E}_m - \mathcal{E}_n)/\hbar$ the equation for the time dependence of the coefficients $C_n(t)$ is

$$\frac{dC_m(t)}{dt} = -\frac{i}{\hbar} \sum_n K_{mn}(t) C_n(t) e^{i\omega_{mn}t}$$

The equation may be solved by *successive approximation*. For a sufficiently small $\Delta\tau$ interval

$$\frac{dC_m(t)}{dt} \approx \frac{C_m(t + \Delta\tau) - C_m(t)}{\Delta\tau}$$

so

$$C_m(t + \Delta\tau) \approx C_m(t) - \frac{i}{\hbar} \sum_n K_{mn}(t) e^{i\omega_{mn}t} C_n(t) \Delta\tau$$

Therefore an approximation of the coefficient C_m at t is

$$C_m(t) \approx C_m(0) - \frac{i}{\hbar} \sum_n \int_0^t K_t(\tau) e^{i\omega_{mn}\tau} C_n(\tau) d\tau$$

where for the sake of clarity the variable under the integral is denoted by τ instead of t . To turn this an interpolation formula first we arbitrarily select the initial C_n values. Denote them with $C_n^{(0)}(t)$. Substituting these functions into the formula above we can determine the first approximation of the coefficients:

$$C_m^{(1)}(t) = C_m^{(0)}(0) - \frac{i}{\hbar} \sum_n \int_0^t K_t(\tau) e^{i\omega_{mn}\tau} C_n^{(0)}(\tau) d\tau$$

In the next step we substitute $C_n^{(1)}$ back in the equation and get the next approximation $C_m^{(2)}$. Continuing this process in the r -th step of the approximation we get

$$C_m^{(r)}(t) = C_m^{(r-1)}(0) - \frac{i}{\hbar} \sum_n \int_0^t K_{mn}(\tau) e^{i\omega_{mn}\tau} C_n^{(r-1)}(\tau) d\tau \quad (22.8.2)$$

22.9 The operator of the angular momentum and its z component in spherical polar coordinates

Determine the representation of the \hat{L}_z and \hat{L}^2 operators in spherical polar coordinates!
Solution In Cartesian coordinates

$$\hat{L}_z \varphi = \hat{x} \hat{p}_y - \hat{y} \hat{p}_x = \frac{\hbar}{i} \left(x \frac{\partial \varphi}{\partial y} - y \frac{\partial \varphi}{\partial x} \right) \quad (22.9.1)$$

Let us separate the \hbar constant

$$\hat{L}_z \varphi = \hbar \hat{l}_z, \text{ where } \hat{l}_z = -i \left(x \frac{\partial \varphi}{\partial y} - y \frac{\partial \varphi}{\partial x} \right)$$

The connections between Cartesian and spherical polar coordinates are

$$\begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{aligned} \quad (22.9.2)$$

The Cartesian x, y and z can be thought as functions of r, θ and ϕ , therefore we may use the chain rule of differentiation when we calculate the derivatives of $\varphi(x, y, z) = \varphi(r, \theta, \phi)$ with respect to the spherical polar coordinates:

$$\begin{aligned} \frac{\partial \varphi}{\partial r} &= \frac{\partial \varphi}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial \varphi}{\partial y} \frac{\partial y}{\partial r} + \frac{\partial \varphi}{\partial z} \frac{\partial z}{\partial r} \\ \frac{\partial \varphi}{\partial \phi} &= \frac{\partial \varphi}{\partial x} \frac{\partial x}{\partial \phi} + \frac{\partial \varphi}{\partial y} \frac{\partial y}{\partial \phi} + \frac{\partial \varphi}{\partial z} \frac{\partial z}{\partial \phi} \\ \frac{\partial \varphi}{\partial \theta} &= \frac{\partial \varphi}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial \varphi}{\partial y} \frac{\partial y}{\partial \theta} + \frac{\partial \varphi}{\partial z} \frac{\partial z}{\partial \theta} \end{aligned}$$

Calculate the derivatives of x, y and z with respect to r, θ and ϕ :

$$\begin{aligned} \frac{\partial x}{\partial r} &= \sin \theta \cos \phi = \frac{x}{r}, & \frac{\partial y}{\partial r} &= \sin \theta \sin \phi = \frac{y}{r}, & \frac{\partial z}{\partial r} &= \cos \theta = \frac{z}{r}, \\ \frac{\partial x}{\partial \phi} &= -r \sin \theta \sin \phi = -y, & \frac{\partial y}{\partial \phi} &= r \sin \theta \cos \phi = y, & \frac{\partial z}{\partial \phi} &= 0, \\ \frac{\partial x}{\partial \theta} &= r \cos \theta \cos \phi = \frac{\cos \theta}{\sin \theta} x = \frac{x}{\tan \theta}, & \frac{\partial y}{\partial \theta} &= r \cos \theta \sin \phi = \frac{y}{\tan \theta}, \\ \frac{\partial z}{\partial \theta} &= -r \sin \theta = -\frac{\sin \theta}{\cos \theta} z = -z \tan \theta, \end{aligned}$$

and substitute back into (22.9.2):

$$\begin{aligned}\frac{\partial \varphi}{\partial r} &= \frac{x}{r} \frac{\partial \varphi}{\partial x} + \frac{y}{r} \frac{\partial \varphi}{\partial y} + \frac{z}{r} \frac{\partial \varphi}{\partial z} \\ \frac{\partial \varphi}{\partial \phi} &= -y \frac{\partial \varphi}{\partial x} + x \frac{\partial \varphi}{\partial y} \\ \frac{\partial \varphi}{\partial \theta} &= \frac{x}{\tan \theta} \frac{\partial \varphi}{\partial x} + \frac{\partial \varphi}{\partial y} \frac{y}{\tan \theta} + \frac{\partial \varphi}{\partial z} \tan \theta\end{aligned}\tag{22.9.3}$$

Comparing the second equation with (22.9.1) we see that

$$\hat{L}_z = \frac{\hbar}{i} \frac{\partial}{\partial \phi}\tag{22.9.4}$$

Using the equations above we can also calculate the whole 3D Laplace operator, which in Cartesian coordinates is

$$\Delta \equiv \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\tag{22.9.5}$$

in spherical polar coordinates. The result:

$$\Delta = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2}.\tag{22.9.6}$$

The sum of the second and third parts contains the operator of the square of the length of the angular momentum:

$$\frac{1}{\hbar^2 r^2} \hat{L}^2 \equiv \frac{1}{r^2} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right].\tag{22.9.7}$$

22.10 Russel-Sounders (LS) and jj coupling of angular momenta. Effects on the electronic structure of atoms

We have learned about spin-orbit coupling in a hydrogen atom, where the total angular momentum was the sum of L and S. In other atoms the calculation of the total angular momentum is harder. Angular momentum coupling in atoms is of importance in atomic spectroscopy. Angular momentum coupling of electron spins is of importance in quantum chemistry.

In light atoms ($Z < 30$) the spins and the orbital momenta are interacting with themselves, i.e. spin-spin ($s-s$) and orbit-orbit ($L-L$) interactions are

larger than the individual spin-orbit ($s - L$) interactions, therefore we can combine the spins into a total spin and the L 's into a total orbital angular momentum with the formulas

$$\mathbf{S} = \sum_k \mathbf{S}_k, \quad \mathbf{L} = \sum_m \mathbf{L}_m$$

The interaction between the resulting \mathbf{L} and \mathbf{S} is called *Russell–Saunders coupling* or *L-S coupling*¹⁰. The total angular momentum \mathbf{J} then can be calculated as

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \quad (22.10.1)$$

This approximation is good as long as no large external magnetic fields are present, which would decouple these two momenta¹¹. Because the complete wave function must be antisymmetric not all combinations of L and S are possible. The ground state term symbol is predicted by Hund's rules

In heavier atoms with bigger nuclear charges spin-orbit coupling is frequently as large as or larger than the spin-spin or orbit-orbit interactions. In those cases not the total spins and orbital momenta, but the individual total angular momenta $\mathbf{J}_k = \mathbf{L}_k + \mathbf{S}_k$ must be combined and the resulting total angular momentum is

$$\mathbf{J} = \sum_k \mathbf{J}_k = \sum_k (\mathbf{L}_k + \mathbf{S}_k)$$

This is called *jj coupling*.

The effect of the L-S and jj couplings differ because the interaction energies between the momenta are proportional to the product of the corresponding angular momenta and the proportionality factors for the different couplings are different:

$$\Delta\mathcal{E}_{L-S} = C_1 \left(\sum_k \mathbf{L}_k \right) \left(\sum_k \mathbf{S}_k \right) + \{small\ terms\} \quad (22.10.2)$$

$$\Delta\mathcal{E}_{jj} = D_1 \left(\sum_k (\mathbf{L}_k + \mathbf{S}_k) \right) \left(\sum_k (\mathbf{L}_k + \mathbf{S}_k) \right) + \{small\ terms\} \quad (22.10.3)$$

Both of these lead to a complicated energy level structure, because there are many possible combinations of L and S for the same principal quantum number n .

¹⁰Named after Henry Norris Russell, 1877-1957 a Princeton Astronomer and Frederick Albert Saunders, 1875-1963 a Harvard Physicist and published in *Astrophysics Journal*, 61, 38, 1925.

¹¹This is called the Paschen-Back effect.

22.11 Other type of hybridization: sp^2 and sp .

We show only two other possible hybrids here: sp^2 and sp . In sp^2 one s and two p orbitals are combined to create three hybridized wave functions while the fourth electron goes into the p_z state, as shown in Fig. 22.8:

$$\psi_1 = \frac{1}{\sqrt{3}}(s + \sqrt{2}p_x) \quad (22.11.1a)$$

$$\psi_2 = \frac{1}{\sqrt{3}} \left(s - \frac{1}{\sqrt{2}}p_x + \sqrt{\frac{3}{2}}p_y \right) \quad (22.11.1b)$$

$$\psi_3 = \frac{1}{\sqrt{3}} \left(s - \frac{1}{\sqrt{2}}p_x - \sqrt{\frac{3}{2}}p_y \right) \quad (22.11.1c)$$

$$\psi_4 = p_z. \quad (22.11.1d)$$

This is the hybridization that occurs in ethylene, which is $H_2C = CH_2$, where the double bond between the carbon atoms is formed by one sp^2 hybrid from each carbon atom forming a σ bond and the overlapping p_z orbitals form a π bond, again as a resemblance of the π orbitals in diatomic molecules. The

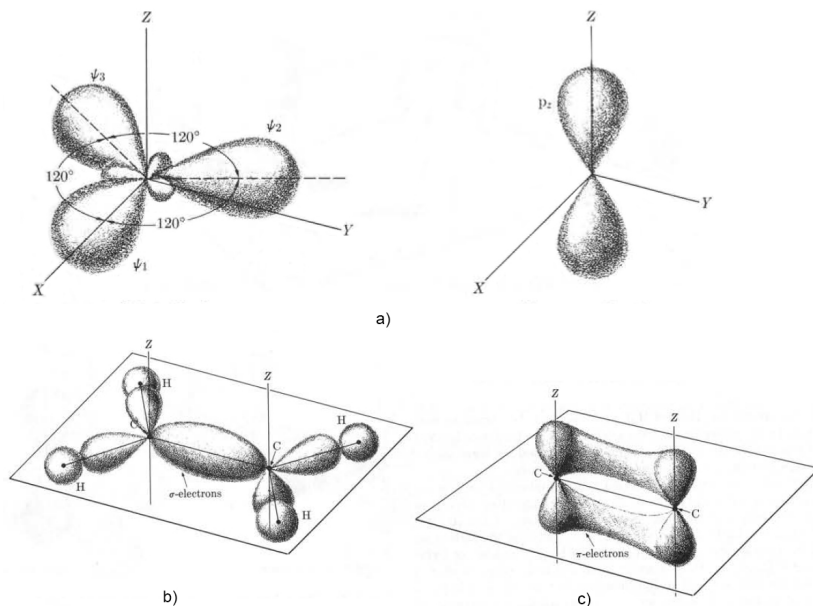


Figure 22.8: a) the sp^2 electron orbitals in carbon, b) σ and c) π bonds in ethylen.

hydrogen atoms are connected to the remaining sp^2 hybrids, therefore this molecule has a planar structure.

The last one of the hybridization we present here is the sp hybridization, found for instance in acetylene ($HC \equiv CH$). The wave functions are:

$$\begin{aligned}\psi_1 &= p_x \\ \psi_2 &= p_y \\ \psi_3 &= s + p_z \\ \psi_4 &= s - p_z,\end{aligned}$$

where the last two are the sp hybrids. The triple bond results from overlapping sp wave functions (σ bond) and the two p_x and p_y bonds creates two π bonds. The hydrogen atoms are bonded by the remaining sp bonds all together forming a linear molecule. These bonds are in Fig .22.9.

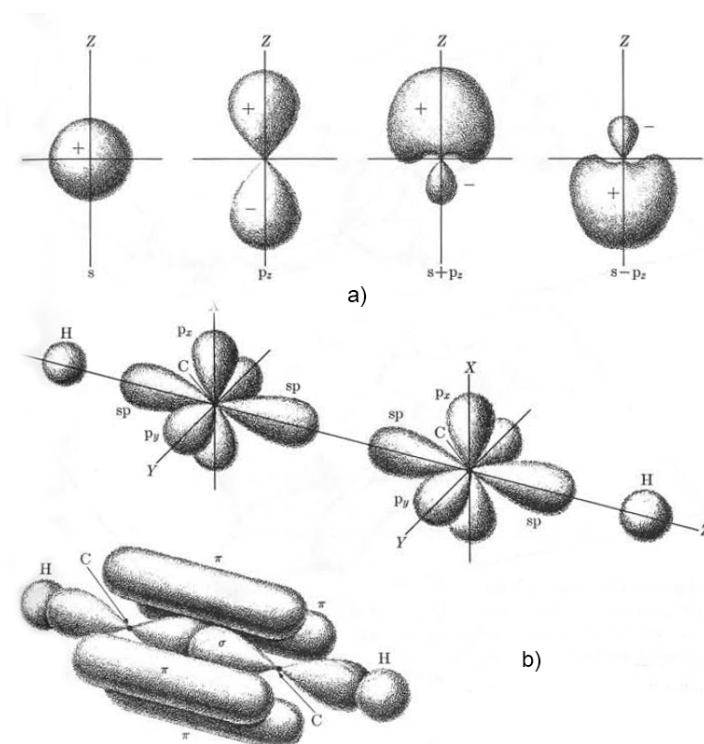
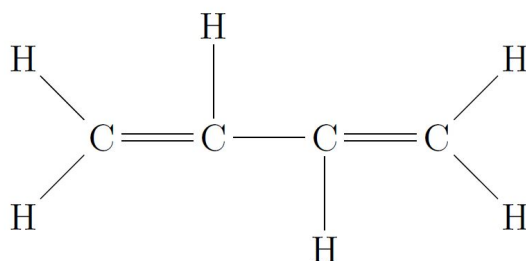


Figure 22.9: a) The wave functions of the sp hybridization - b) and in the acetylene molecule.

22.12 Conjugated molecules

Organic chemistry deals with molecules containing carbon. Not only it is an interesting area in its own right but it also the base of our existence. Organic molecules contain single and (or) multiple bonds between carbon atoms. We will discuss a class of molecules in which there is a single bond between two double bonds (in other words single and multiple bonds alternate between the carbon atoms), these are called *conjugated molecules*¹². *Conjugation* is the overlap of one p-orbital with another across an intervening σ bond¹³.

Our first example is butadiene (C_4H_6) which can also be written as



The carbon atoms along the chain are bound by σ bonds using sp^2 hybrid wave functions. The four p_z electrons form π bonds along the chain, but in a special way. Instead of being localized in particular regions of the molecule as electrons in the σ bonds, these π bonding electrons can move along the molecule. The electron wave functions in butadiene are in Fig. 22.10

The π electrons introduce a certain rigidity into the molecular structure. If we now look at the four π wave functions, which, like before are created as a linear combination of the four p_z atomic wave functions. This linear combination must be either symmetric or anti-symmetric relative to the center of the molecule as shown in Fig. 22.11

Each energy level in Fig. 22.11 a) can accept two electrons with opposite spins. Molecular orbital ψ_1 is of the bonding type for each carbon atom, while ψ_2 is bonding for the pairs 1–2 and 3–4 and anti-bonding for the pair 2–3. This is the reason for the dip in the probability distribution at the center of the molecule. Therefore the bond strength is smaller between the C atoms 2–3 than between the other pairs.

It is easy to generalize these thoughts for other *polyenes*, which are conjugate compounds of $2n$ carbon atoms. The bonding between the carbon atoms

¹²If there is no other bond between two double bonds the molecule has *cumulated* bonds, while if the double bonds are separated by two or more single bonds it is called *unconjugated*.

¹³In larger atoms d-orbitals can be involved.

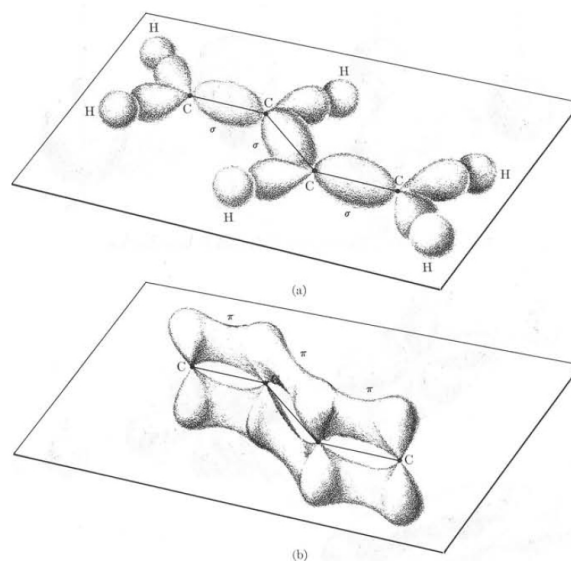


Figure 22.10: Electron distributon in butadiene. a) localized σ bonds, b) unlocalized π bonds.

would be written in the classical valence model as $\dots - -C == C - -C == C - -C == \dots$. Again in addition to the σ bonds between pairs of carbon atoms we have $2n$ π electrons spread along the molecule. In this case there are $2n$ closely spaced energy levels available. These could accomodate $4n$ electrons, but we only have $2n$. Hence in the ground state only the lower half of the energy levels are occupied which leads to an easily excitable system. It can easily exited e.g. by light with frequency corresponding to $\Delta\mathcal{E}/\hbar$. These molecules absorb light at selected frequencies, i.e. they show color.

Our final example is the conjugate molecule of benzene (C_6H_6), or graphically

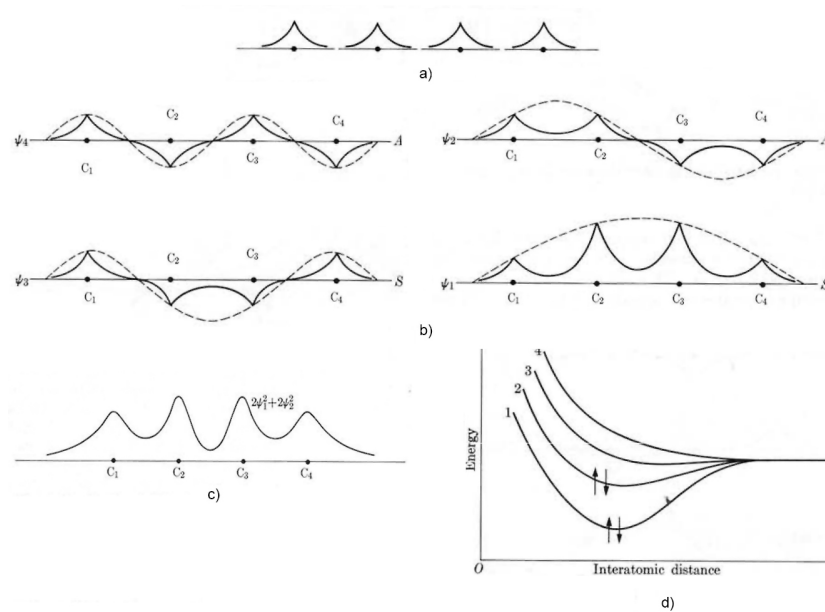
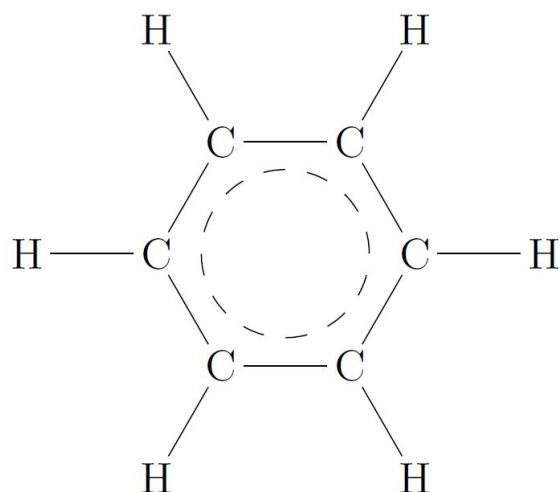


Figure 22.11: Schematic molecular orbitals for π electrons in butadiene. a) The atomic wave functions when the atoms are far apart. “ $C_1...C_4$ ” are the carbon atoms. b) Molecular orbitals, “A” - antisymmetric, “S” - symmetric wave function c) total probability density along the axis, d) electronic potential energy vs distance.



As seen in Fig. 22.12 the carbon atoms sit in the vertices of a hexagon and joined by σ bonds using sp^2 hybrid wave functions along each $C - C$ line and the hydrogen atoms are attached to the remaining sp^2 orbitals. There are also 6 electrons, one from each carbon atom in the p_z orbitals, which are perpendicular to the plane of the molecule. These π electrons move freely along the hexagon like a closed current loop. This is the reason for the strong diamagnetism of benzene and other cyclic conjugate molecules.

22.13 Calculating the maximum probability partition of the Maxwell-Boltzmann distribution

The formula whose maximum we want to calculate is (9.2.3)

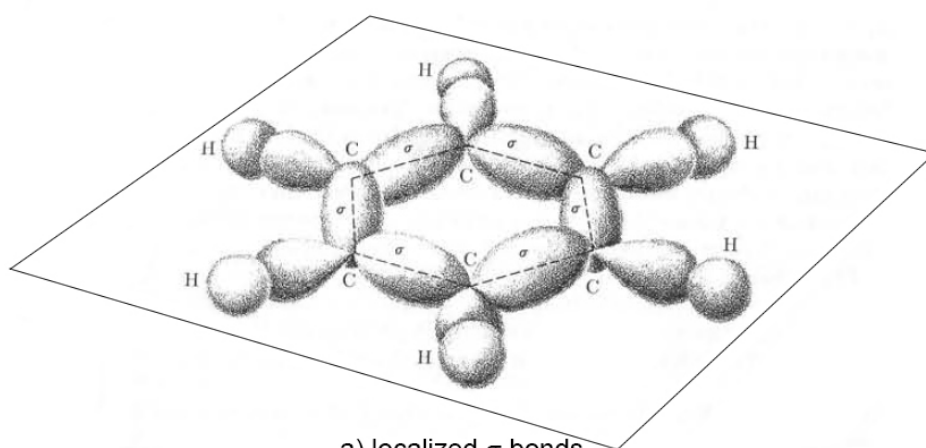
$$\mathcal{P}(\{n_i\}) = \prod_i \frac{g_i^{n_i}}{n_i!} \quad (22.13.1)$$

Usually a function $f(n_1, n_2, \dots)$ may have an extremum where

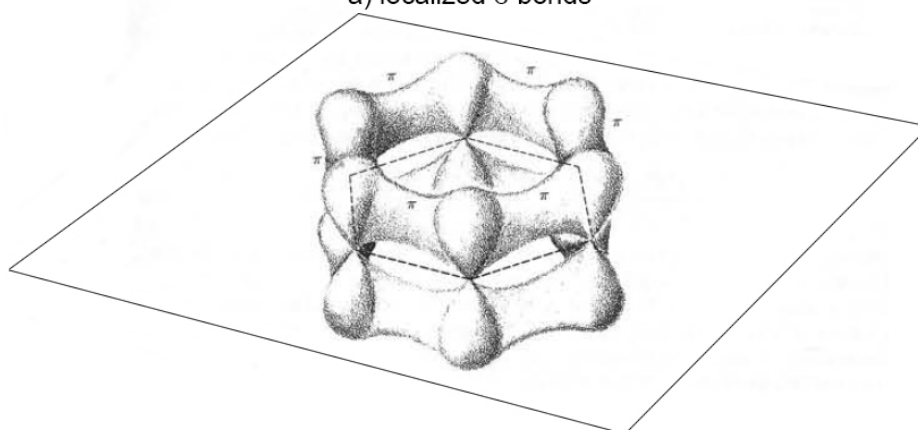
$$\frac{\partial f}{\partial n_i} = 0 \quad \text{for all values of } i$$

But this is not an ordinary maximum calculation for three reasons: the function is a product of many factors, it contains a factorial and the maximum is not unconditional.

To get rid of the first problem let us observe that the logarithm of a product is a sum and because the logarithm function is monotone f and $\ln f$



a) localized σ bonds



b) unlocalized π bonds

Figure 22.12: Benzene molecular orbitals

has maximums at the same $\{n_i\}$ position. We can take the logarithm of \mathcal{P} as it never become 0 or negative.

This helps to solve the second problem too because the factorial may be approximated by the analytical *Stirling formula*¹⁴:

$$\ln(n!) \approx \left(n + \frac{1}{2}\right) \ln n - n + \ln \sqrt{2\pi} \quad (22.13.2)$$

In practice only the following simplified form is used:

$$\ln(n!) \approx n \ln n - n. \quad (22.13.3)$$

This simplified formula gives results which differs from the real value by a factor of about 2.5, but even so it is be good enough for our purpose because the derivative of any additional constant is 0.

The third problem that there are additional conditions can be overcome by the *method of Lagrange multipliers*. In this method the conditions are reordered to one side of the equal sign and the resulting zero valued expressions multiplied by yet unknown constants, called *Lagrange multipliers*, are added to the function whose maximum (or minimum) we want to find and we search the maximum (or minimum) of the result. This will give us equations for the unknown Lagrange multipliers.

The conditions which must be fulfilled are:

$$\begin{aligned} N &= \sum_i n_i \\ \mathcal{E} &= \sum_i n_i \mathcal{E}_i \end{aligned}$$

which can be rearranged:

$$N - \sum_i n_i = 0 \quad (22.13.4)$$

$$\mathcal{E} - \sum_i n_i \mathcal{E}_i = 0 \quad (22.13.5)$$

¹⁴The Stirling formula is a good approximation for the factorial of large numbers. Even for small ones like 10 it gives results of the correct order of magnitude:

$$\begin{aligned} 10! &= 3,628,800 & \Rightarrow & \ln 10! = 15.1044125730755153 \\ (10 + 0.5) * \ln 10 - 10 + \ln \sqrt{2\pi} &= 15.0960820096421524 & \Rightarrow & 10! \approx 3,598,696 \end{aligned}$$

the error is only 0.8% of the real value. For larger numbers the accuracy increases, e.g. $\ln(100!) = 363.7393756$, while the Stirling formula gives 363.73854224, which is only a factor of 1.000834 times smaller than the real value.

Using α and β as Lagrange multipliers we must determine the extremum of the function:

$$f(n_1, n_2, \dots) = \ln \mathcal{P} + \alpha \left(N - \sum_i n_i \right) + \beta \left(\mathcal{E} - \sum_i n_i \mathcal{E}_i \right)$$

The logarithm of \mathcal{P} is

$$\ln \mathcal{P} = \ln \left(\prod_i \frac{g_i^{n_i}}{n_i!} \right) \approx \sum_i (n_i \ln g_i - n_i \ln n_i + n_i) \quad (22.13.6)$$

from which

$$f(n_1, n_2, \dots) = \alpha N + \beta \mathcal{E} + \sum_i n_i \ln g_i - n_i \ln n_i + n_i - (\alpha + \beta \mathcal{E}_i) n_i \quad (22.13.7)$$

The position of the maximum is determined by

$$\frac{\partial f}{\partial n_i} = \ln g_i - \ln n_i - (\alpha + \beta \mathcal{E}_i) = 0, \quad i = 1, 2, \dots, n \quad (22.13.8)$$

The solution is

$$\begin{aligned} \ln n_i &= \ln g_i - (\alpha + \beta \mathcal{E}_i), & \text{or} \\ n_i &= g_i e^{-\alpha - \beta \mathcal{E}_i} \end{aligned} \quad (22.13.9)$$

If we substitute n_i back into $\ln \mathcal{P}$ and assume that the total number of particles is large: $N \gg 1$ we get

$$\ln \mathcal{P} = \alpha N + \beta \mathcal{E} \quad (22.13.10)$$

We are almost done! Introduce a new variable $S \equiv k_B \ln \mathcal{P}$, express \mathcal{E} from this formula and see how much would \mathcal{E} change if both the number of particles and S would change infinitesimally, i.e. take the difference of both sides:

$$d\mathcal{E} = \frac{1}{k_B \beta} dS - \frac{\alpha}{\beta} dN \quad (22.13.11)$$

I hope you recognize this formula has the same the form as the *second law of thermodynamics*:

$$d\mathcal{E} = T dS + \mu dN \quad (22.13.12)$$

if you substitute $k_B T$ for β and μ for $(-\alpha k_B T)$, where μ is the *chemical potential*. From (22.13.9)

$$n_i = g_i e^{-\alpha - \beta \varepsilon_i} = g_i e^{-\frac{\varepsilon_i - \mu}{k_B T}} = g_i \frac{1}{e^{\frac{\varepsilon_i - \mu}{k_B T}}}$$

Because the total number of particles is

$$N = \sum_i n_i = \sum_i g_i e^{-\alpha - \beta \varepsilon_i} = e^{-\alpha} \sum_i g_i e^{-\beta \varepsilon_i}$$

therefore

$$e^{-\alpha} = \frac{N}{\sum_i g_i e^{-\beta \varepsilon_i}}$$

The factor $e^{-\alpha} = e^{\mu/k_B T}$ is called the *absolute activity*.

Introducing the Z partition function with

$$Z = \sum_i g_i e^{-\beta \varepsilon_i} \quad (22.13.13a)$$

the final result becomes

$$n_i = \frac{N}{Z} g_i e^{-\beta \varepsilon_i} \quad (22.13.13b)$$

22.14 Superfluidity in helium 4.

For instance *helium-4* (${}^4\text{He}$) have zero spin, therefore it is a boson¹⁵, At temperatures less than 2.17 K (the *lambda point*) it becomes a new kind of fluid, now known as a *superfluid*¹⁶. *Superfluid helium*, also called *He-II* has many unusual properties, including zero viscosity, the ability to flow without dissipating energy. This is so because all helium-4 atoms tend to be in the same quantum state characterized by the magnitude and direction of their velocity. As a result liquid ${}^4\text{He}$ cannot be kept in an open container as it will flow along its surface and out of it with a high velocity ($\approx 20\text{cm/s}$). Many ordinary liquids like alcohol or petroleum creep up solid walls because of *surface tension*, but this is constricted by their non-zero viscosity.

We can explain this phenomena on statistical ground:

¹⁵The more common helium-3, is a fermion.

¹⁶This was discovered in 1938 by Pyotr Kapitsa, John Allen and Don Misener.

Example 22.1. *Let us consider very low temperatures where the average thermal energy will only allow the occupation of the lowest two quantum states. In this case the probability of occupation will be larger for the lower state. If a single boson has a p probability to be in the higher state, then, because any number of bosons can be excited there with the same probability, the probability for n bosons being in the higher energy state will be proportional to p^n . The proportionality constant C is determined from the condition that the sum of the probabilities of all possible distributions must be 1:*

$$\sum_{n=1}^N C p^n = 1$$

For large N (or more accurately when $N \rightarrow \infty$)

$$\sum_{n=1}^{\infty} C p^n = C \sum_{n>0} p^n = C \frac{p}{1-p} = 1 \quad \Rightarrow \quad C = \frac{1-p}{p}$$

from the sum of the infinite geometric series. The average number of particles in the upper state for large N is then approximately

$$\langle n \rangle = \sum_{n>0} n C p^n = C \frac{p^2}{(1-p)^2} = \frac{p}{1-p}$$

Which is negligible compared to N if $p \ll 1$, therefore almost all particles will be in the lowest lying energy state.

Chapter 23

Solid State Physics

23.1 The origin of van der Waals forces

The fact that the atoms do not possess a constant electric dipole moment only means that the *time averaged dipole moment* is 0, but they may have non vanishing electric dipole moments at any instance. Consider two atoms at a distance r . Although the average moment is 0 the p_1 instantaneous non 0 electric dipole moment of atom 1 creates an E electric field at the other atom. This will induce a dipole moment in atom 2:

$$p_2 = \alpha E \sim \frac{\alpha p_1}{r^3} \quad (23.1.1)$$

where α is the polarizability of the atom. The interaction energy between p_1 and p_2 lowers the total energy of the system by¹:

$$\left\langle \frac{p_1 p_2}{r^3} \right\rangle \sim \alpha E \sim \left\langle \frac{\alpha p_1^2}{r^6} \right\rangle = \frac{\alpha \langle p_1^2 \rangle}{r^6} \quad (23.1.2)$$

Although $\langle p_1 \rangle (= \langle p_2 \rangle) = 0$, $\langle p_1^2 \rangle \neq 0$

23.2 Examps of Bravais lattices

Important 23.2.1. *Drawings which depict the geometry of the crystals (like the one in Fig. 11.13 usually shows small balls with connecting lines. These lines neither represent the chemical bonds between the atoms in the crystal (not even for monatomic lattices) nor do they show the real size of the atoms.*

¹Exact theory: must consider the interaction between groups of three or more atoms.

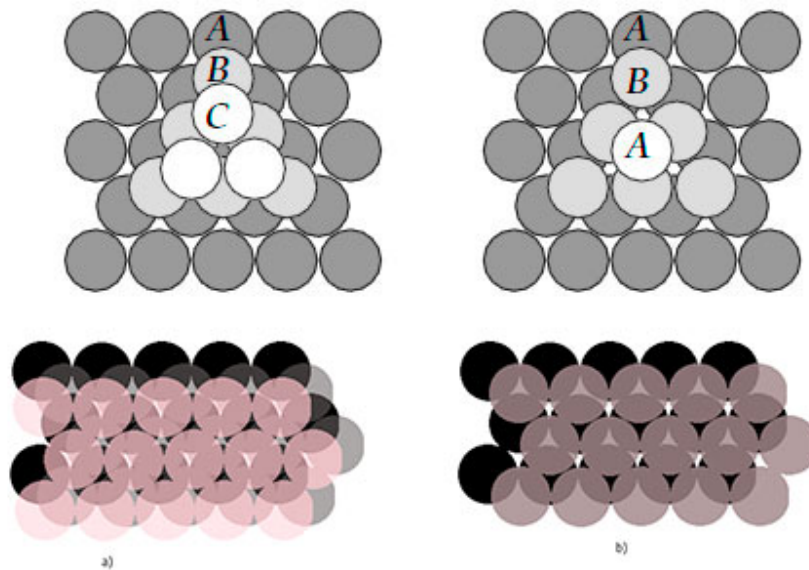


Figure 23.1: The two possible packing of spherical atoms in the densest formation

Examples:

- Close packing structures. If we consider a crystal with atoms represented by solid spheres then there are two different possible choices to select for the densest packing: we can view both structures as planes of spheres closely packed in two dimensions, which gives a hexagonal lattice; for close packing in three dimensions (Fig. 23.1) the successive planes must be situated so that a sphere in one plane sits at the center of a triangle formed by three spheres in the previous plane. There are two ways to form such a stacking of hexagonal close-packed planes: ...ABCABC..., and ...ABABAB..., where A, B, C represent the three possible relative positions of spheres in successive planes according to the rules of close packing, as illustrated in Fig. 23.1. The first sequence corresponds to the *fcc* (face centered cubic) lattice, the second to the *hcp* (hexagonal close packing) lattice. Elements which crystallize in monatomic fcc are: Ba, Cr, Co, Fe, K, Li, Mo, Na, Nb, Rb, Ta, V, W Elements with hexagonal close packed (hcp) crystal structure: Be, Cd, Ce, α -Co, Dy, Er, Gd, He (2K!), Hf, Ho, La, Lu, Mg, Nd, Os, Pr, Re, Ru, Ru, Sc, Tb, Ti, Tl, Tm, Y, Zn, Zr
- As a special variation of close packing the diamond lattice (or the zincblende lattice) that consists of two interpenetrating face centered cubic Bravais

lattice, displaced along the body diagonal of the cubic cell by one quarter the length of the diagonal. (The same lattice therefore may be considered an fcc lattice with a diatomic basis.) Elements with diamond

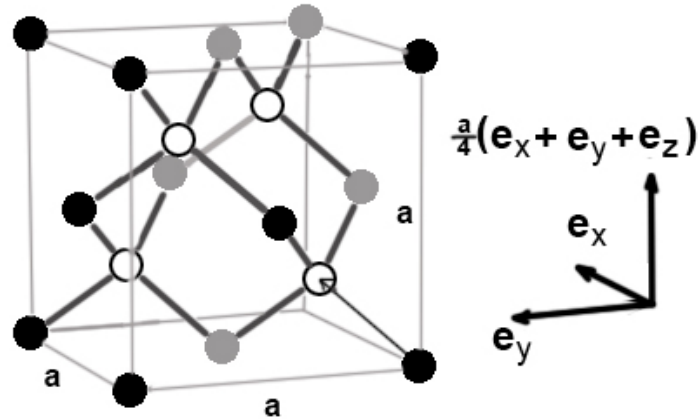


Figure 23.2: The diamond structure. The sides of the confining cell is drawn together with the bonds.

crystal structure: C, Si, Ge, α -Sn (grey)

- Another special variation of the close packing structure is the wurtzite lattice that consists of two interpenetrating hcp lattices.
- The bcc structure is almost close packing
Elements which crystallize in monatomic bcc are: Ar (at 4.2K!) , Ag, Al, Au, Ca, Ce, β -Co, Cu, Ir, Kr, La, Ne, Ni, Pb, Pd, Pr, Pt, δ -Pu, Rh, Sc, Sr, Th, Xe (at 58 K!), Yb

23.3 X-ray diffraction methods Laue-, rotating crystal and Debye-Scherrer methods.

A simple geometrical construction due to Ewald helps to visualize the possibilities.

The Ewald Construction

An incident wave vector will lead to a Bragg reflection if and only if the tip of the wave vector lies on a plane in the k-space. Since the set of lattice

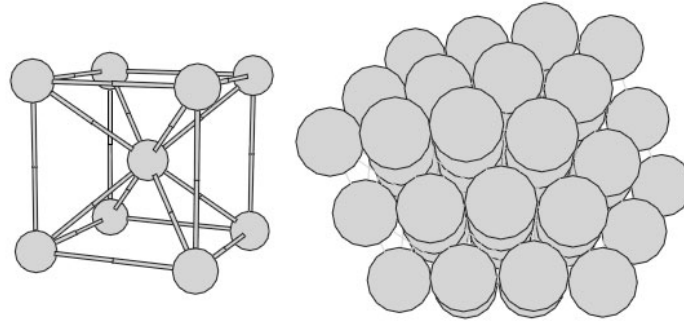


Figure 23.3: Left: one atom and its eight neighbors in the body-centered cubic (bcc) lattice; the size of the spheres representing atoms is chosen so as to make the neighbors and their distances apparent. Right: a portion of the three-dimensional bcc lattice; the size of the spheres is chosen so as to indicate the almost close-packing nature of this lattice.

planes is discrete this condition cannot be fulfilled with an arbitrarily fixed k , i.e. with arbitrarily fixed X-ray wavelength and arbitrarily fixed incident direction, so in general there will be no diffraction peaks at all.

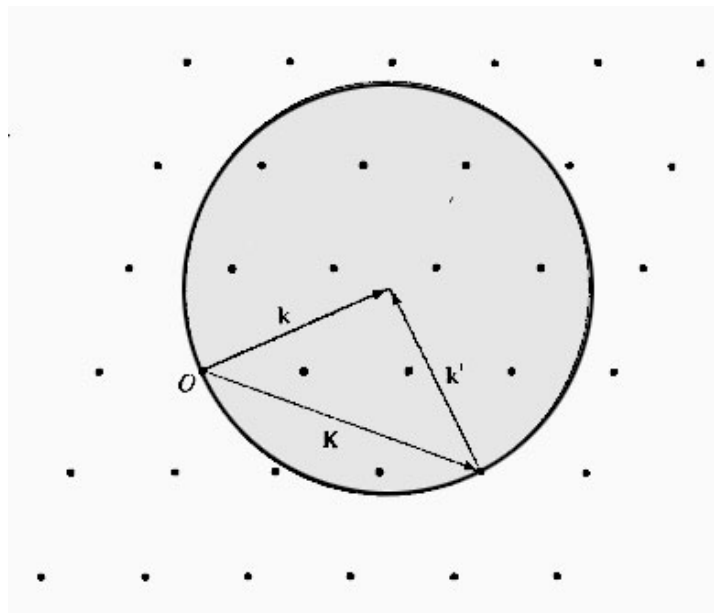


Figure 23.4: Ewald construction on a plane of the reciprocal lattice

To get a diffraction picture therefore we must vary k by either varying the

wavelength or varying the relative orientation of the crystal and the incident wave.

For elastic scattering the length of k and k' is the same and their difference K must be a vector of the reciprocal lattice, therefore both of the endpoints of K must lie on a point of the reciprocal lattice.

Let us take the incident k vector (Fig. 23.4). Put its origin in any point of the reciprocal lattice. The endpoint of this vector usually does not point to any other point of the reciprocal lattice. Draw a sphere with a radius of k around the endpoint. If there will be any other reciprocal lattice points on this sphere they all represent a possible Bragg peak. This sphere is called the *Ewald sphere*.

This construction suggests some methods to study the structure of crystals.

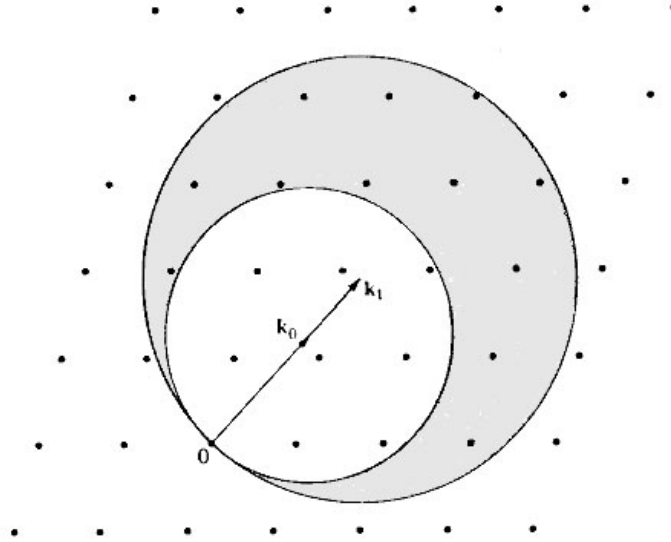
23.3.1 The Laue Method

The Laue method is mainly used to determine the orientation of large *single crystals* whose structure is known. E.g. if the incident direction lies along a symmetry axis of the crystal, the pattern of spots produced by the Bragg-reflected rays will have the same symmetry. Because solid state physicist generally do study known crystal structures, the Laue method is probably the one of greatest practical interest.

White radiation of wavelengths between λ_{min} and λ_{max} and of a *fixed direction* is reflected from, or transmitted through, a *single crystal of fixed orientation*. For many k 's there will be planes in the crystal which satisfy the Bragg-condition and produce constructive interference.

The Bragg angle is fixed for every set of planes in the crystal. Each set of planes picks out and diffracts the particular wavelength from the white radiation that satisfies the Bragg law for the values of d and Θ involved. The diffracted beams lie on the surface of imaginary cones and form arrays of spots. A sheet film perpendicular to the incident beam records these spots, on curves. Each curve therefore corresponds to a different wavelength. Experimental variations of the Laue method :

Crystal orientation is determined from the position of the spots. Each spot can be indexed, i.e. attributed to a particular plane, using special charts. The Laue technique can also be used to assess crystal perfection from the size and shape of the spots. If the crystal has been bent or twisted in any way, the spots become distorted and smeared out.



6

Figure 23.5: Section of the Ewald sphere for the Laue method. The smaller circle corresponds to $k_{min} = 2\pi/\lambda_{min}$, the larger one to $k_{max} = 2\pi/\lambda_{max}$, and the points within the shaded area corresponds the reciprocal vectors that gives the observable Bragg peaks.

23.3.2 The Rotating Crystal Method

In the rotating crystal method, a single crystal is mounted with an axis normal to a monochromatic x-ray beam. A cylindrical film is placed around it and the crystal is rotated about the chosen axis. As the crystal rotates, sets of lattice planes will at some point make the correct Bragg angle with the monochromatic incident beam, and at that point a diffracted beam will be formed. The reflected beams are located at discrete positions corresponding to the Laue condition on the surface of imaginary cones. When the film is laid out flat, the diffraction spots lie on horizontal lines. The main application of the rotating crystal method is the determination of unknown crystal structures.

23.3.3 The Debye-Scherrer Powder method

The powder method is used to determine the value of the lattice parameters accurately. Lattice parameters are the magnitudes of the unit vectors a_1 , a_2 and a_3 which define the unit cell for the crystal. If a monochromatic x-ray beam is directed at a single crystal, then only one or two diffracted beams

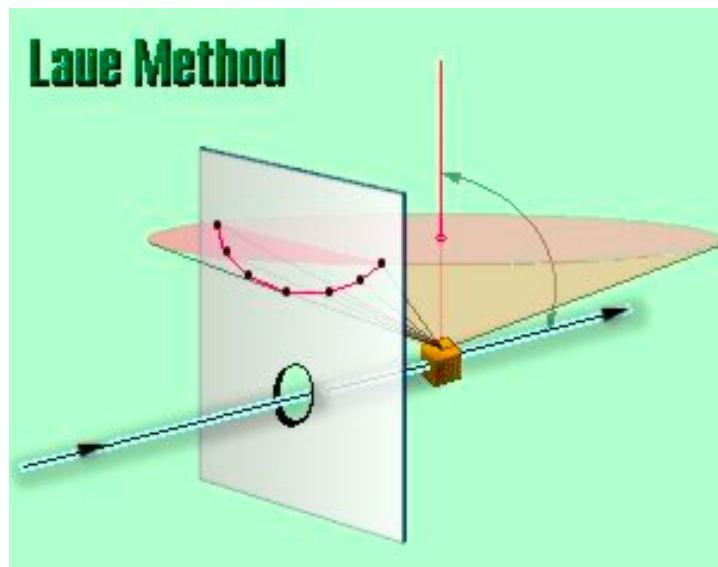


Figure 23.6: *Back-reflection Laue* In the back-reflection method, a photographic sheet film is placed between the x-ray source and the crystal. The beams which are diffracted in a backward direction are recorded. One side of the cone of Laue reflections is defined by the transmitted beam. The film intersects the cone, with the diffraction spots generally lying on an hyperbola.

may result. If the sample consists of some tens of randomly orientated single crystals, the diffracted beams are seen to lie on the surface of several cones. The cones may emerge in all directions, forwards and backwards. For a sample of some hundreds of crystals (i.e. a powdered sample) the diffracted beams form continuous cones. For every set of crystal planes, by chance, one or more crystals will be in the correct orientation to give the correct Bragg angle to satisfy Bragg's equation. Every crystal plane is thus capable of diffraction. Each diffraction line is made up of a large number of small spots, each from a separate crystal. Each spot is so small as to give the appearance of a continuous line. If the crystal is not ground finely enough, the diffraction lines appear speckled.

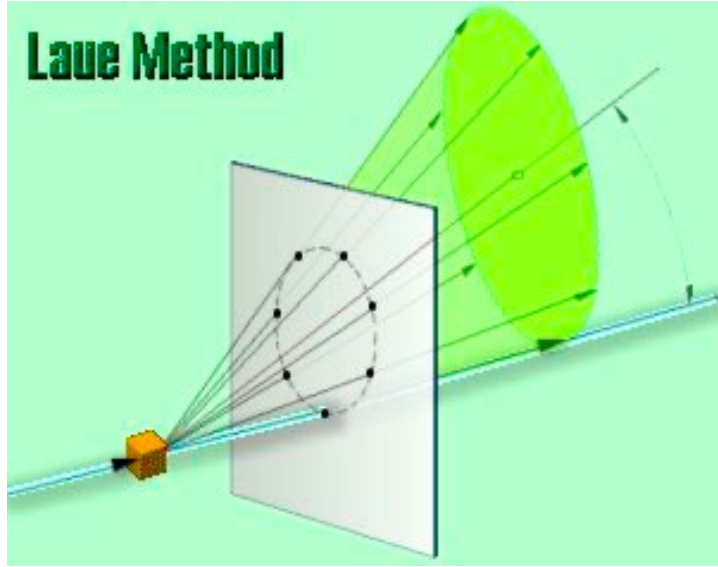


Figure 23.7: *Transmission Laue* In the transmission Laue method, the film is placed behind the crystal to record beams which are transmitted through the crystal. The film intersects the cone, with the diffraction spots generally lying on an ellipse.

23.4 Classical linear chain models of lattice vibrations

23.4.1 Single atomic linear chain

The equation of motion for the n -th atom inside an N atom linear chain (see (13.1.1)):

$$M \frac{d^2 u_n}{dt^2} = \beta(u_{n+1} - u_n) - \beta(u_n - u_{n-1}) = \beta(u_{n+1} - 2u_n + u_{n-1}) \quad (23.4.1)$$

Using the *Born - von Karman periodic boundary condition*:

$$u_{N+1} = u_1$$

and the ansatz of

$$u_n = u_o e^{\pm i(\omega t + k n a)} \quad (23.4.2)$$

The periodic boundary condition then requires that

$$e^{\pm i k N a} = 1$$

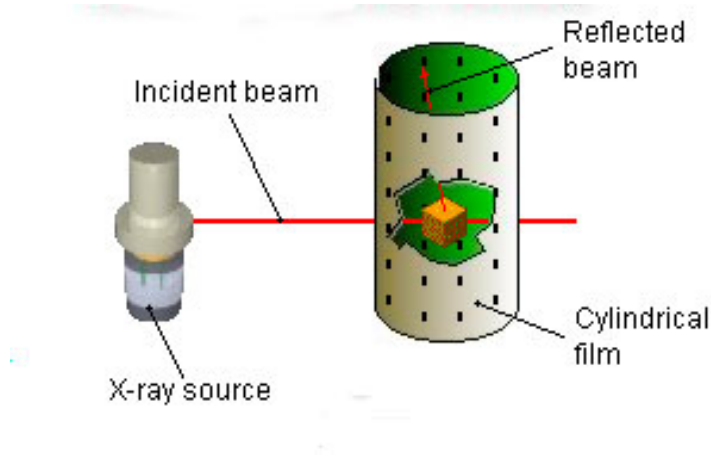


Figure 23.8: The Rotating Crystal Method

We know that $e^{i2\pi n} = 1$, where $n = 0, \pm 1, \pm 2, \dots$ i.e. the possible values for k are

$$k = \frac{2\pi}{a} \frac{n}{N} \quad \text{where } n \text{ is an integer} \quad (23.4.3)$$

Substituting (23.4.2) into (23.4.1):

$$-M\omega^2 u_o e^{\pm i(\omega t + k n a)} = \beta \left(u_o e^{\pm i(\omega t + k(n+1)a)} - 2u_o e^{\pm i(\omega t + k n a)} + u_o e^{\pm i(\omega t + k(n-1)a)} \right)$$

canceling the common $u_o e^{\pm i(\omega t + k n a)}$ factor:

$$-M\omega^2 = \beta (e^{ika} + e^{-ika} - 2)$$

(Both sign selection would result in this formula.) So the final result is:

$$\omega(k) = 2 \sqrt{\frac{\beta}{M}} \sin \frac{1}{2} k a \quad (23.4.4)$$

The general solution for the time dependent position excursion of the n -th atom will be a linear combination of solutions of the form of (23.4.2):

$$u_n(t) = \sum_k (\xi_k e^{i(\omega t + k n a)} + \xi_k^* e^{-i(\omega t + k n a)}) \quad (23.4.5)$$

This form ensures the reality of the solution and the $2N$ parameters, required for the general solution of N 2nd order differential equations, are the real

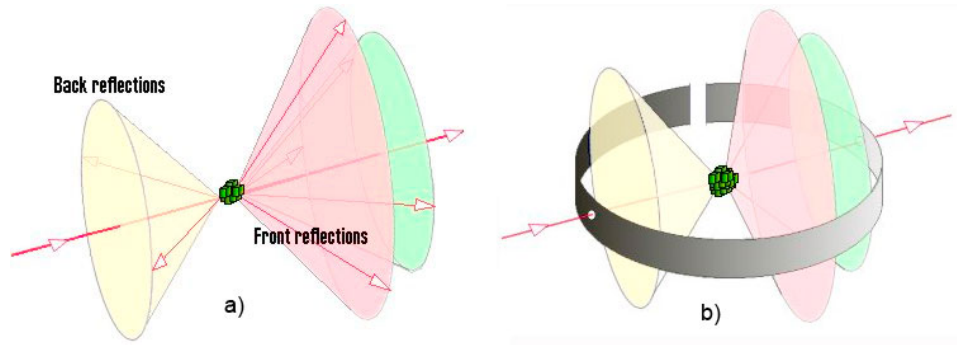


Figure 23.9: *The Powder Method* A circle of film is used to record the diffraction pattern as shown. Each cone intersects the film giving diffraction lines. The lines are seen as arcs on the film.

and imaginary parts of ξ_k . Because the summation goes over all positive and negative values of k (23.4.5) may be rewritten:

$$\begin{aligned} u_n(t) &= \sum_{k=-\pi/a}^{\pi/a} (\xi_k e^{i(\omega t + k n a)} + \xi_k^* e^{-i(\omega t + k n a)}) = \\ &= \sum_{k=-\pi/a}^{+\pi/a} \xi_k e^{i(\omega t + k n a)} + \sum_{k=+\pi/a}^{-\pi/a} \xi_{-k}^* e^{-i(\omega t - k n a)} \end{aligned}$$

If $\xi_{-k}^* = \xi_k$ then

$$\sum_{k=+\pi/a}^{-\pi/a} \xi_{-k}^* e^{-i(\omega t - k n a)} = \sum_{k=-\pi/a}^{+\pi/a} \xi_k e^{i(\omega t + k n a)},$$

therefore

$$\begin{aligned} u_n(t) &= \sum_k (\xi_k e^{i\omega t} e^{ikna} + \xi_{-k}^* e^{-i\omega t} e^{-ikna}) = \\ &= \sum_k (\xi_k e^{i\omega t} + \xi_{-k}^* e^{-i\omega t}) e^{ikna} \end{aligned}$$

Introducing new coefficients with

$$\chi_k^*(t) \equiv \xi_k (e^{i\omega t} + e^{-i\omega t})$$

$u_n(t)$ will be real and may be written in the form

$$u_n(t) = \sum_k \chi_k(t) e^{ikna} \quad (23.4.6)$$

23.4.2 Diatomic linear chain.

The equations for a diatomic linear chain with different mass atoms are

$$M_1 \frac{d^2 u_n}{dt^2} = \beta ((v_n - u_n) - (u_n - v_{n-1})) \quad (23.4.7)$$

$$M_2 \frac{d^2 v_n}{dt^2} = \beta ((u_{n+1} - v_n) - (v_n - u_n)) \quad (23.4.8)$$

Try the solutions in the form

$$u_n = u_k e^{i(\omega t + k n a)} \quad (23.4.9)$$

$$v_n = v_k e^{i(\omega t + k n a)} \quad (23.4.10)$$

Substituting into (23.4.7) and (23.4.8) then rearranging (to the right side) and canceling the common $e^{i(\omega t + k n a)}$ factors, we arrive at the following pair of equations:

$$0 = (M_1 \omega^2 - 2\beta) u_k + \beta(1 + e^{-ika}) v_k \quad (23.4.11)$$

$$0 = \beta(1 + e^{ika}) u_k + (M_2 \omega^2 - 2\beta) v_k \quad (23.4.12)$$

Or in matrix form :

$$\begin{bmatrix} M_1 \omega^2 - 2\beta & \beta(1 + e^{-ika}) \\ \beta(1 + e^{ika}) & M_2 \omega^2 - 2\beta \end{bmatrix} \begin{bmatrix} u_k \\ v_k \end{bmatrix} = 0 \quad (23.4.13)$$

Because this is a homogenous pair of equations a solution only exists when the determinant of the matrix is 0. This gives a quadratic equation for ω^2 :

$$\begin{aligned} (\omega^2 M_1 - 2\beta)(\omega^2 M_2 - 2\beta) - \beta^2(1 + e^{-ika})(1 + e^{ika}) &= 0 \\ M_1 M_2 \omega^4 - 2\beta(M_1 + M_2)\omega^2 + 4\beta^2 - 2\beta^2(1 + \cos ka) &= 0; \end{aligned}$$

Using the trigonometric identities

$$\begin{aligned} 4\beta^2 - 2\beta^2(1 + \cos ka) &= 2\beta^2(1 - \cos ka) = 4\beta^2 \sin^2(ka/2) \\ &= 4\beta^2 \sin^2 kb \end{aligned}$$

We must solve this quadratic equation for ω_{\pm} .

$$\omega_{\pm}^2 = \frac{\beta}{M_1 M_2} \left(M_1 + M_2 \pm \sqrt{(M_1 + M_2)^2 - 4M_1 M_2 \sin^2 kb} \right) \quad (\text{where } a = 2b) \quad (23.4.14)$$

23.4.3 3D Linear model of lattice vibrations

Let us denote the displacement of the j -th atom of the basis at the Bravais vector \mathbf{R} with $\mathbf{u}_j(\mathbf{R})$. Then

$$U \equiv E_{total\ pot.\ energy} = U(\mathbf{u}_1(\mathbf{R}_1), \mathbf{u}_2(\mathbf{R}_1), \dots, \mathbf{u}_n(\mathbf{R}_1), \\ \mathbf{u}_1(\mathbf{R}_2), \mathbf{u}_2(\mathbf{R}_2), \dots, \mathbf{u}_n(\mathbf{R}_2), \\ \dots \\ \mathbf{u}_1(\mathbf{R}_N), \mathbf{u}_2(\mathbf{R}_N), \dots, \mathbf{u}_n(\mathbf{R}_N))$$

$$U = u_o + \sum_{\mathbf{R}, \mathbf{R}', j, j'} \frac{1}{2} \mathbf{u}_j(\mathbf{R}) \mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}') \mathbf{u}_{j'}(\mathbf{R}') \quad (23.4.15)$$

where u_o is the potential energy when all atoms are in their equilibrium position. Fortunately we know about the properties of tensor \mathbf{D} of the spring constants:

- It must have translational symmetry, therefore $\mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}')$ may only depend on the distance between sites \mathbf{R} and \mathbf{R}' :

$$\mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}') = \mathbf{D}_{j,j'}(\mathbf{R} - \mathbf{R}')$$

- $\mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}')$ must be real and symmetric:

$$\mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}') = \mathbf{D}_{j',j}(\mathbf{R}' - \mathbf{R})$$

- When the crystal is translated as a whole the total potential energy must remain the same:

$$\sum_{\mathbf{R}, \mathbf{R}', j, j'} \mathbf{D}_{j,j'}(\mathbf{R}, \mathbf{R}') = 0$$

Even with this knowledge it is easy to see that solving the equations in 3D is much more complicated than it was in the 1D case. We will not solve the equations in 3D, but will only discuss the results.

23.5 Mathematical note: From summation to integration

Summation over the allowed values of \mathbf{k} may be approximated with integrals if the physical quantity used does not vary appreciably over distances of order

$2\pi/L$ in k -space. Because the $d^3\mathbf{k}(\equiv V_k)$ volume of k -space per allowed \mathbf{k} in 3D is given by (14.3.6) $V_k = 8\pi^3/V$:

$$\sum_k F(\mathbf{k}) = \sum_k F(\mathbf{k}) \left(\frac{V}{8\pi^3} \right) d^3\mathbf{k} = \frac{V}{8\pi^3} \sum_k F(\mathbf{k}) d^3\mathbf{k}$$

in the limit $d^3\mathbf{k} \rightarrow 0$ (i.e. $V \rightarrow \infty$) for unit volume

$$\lim_{V \rightarrow \infty} \frac{1}{V} \sum_k F(\mathbf{k}) = \frac{1}{8\pi^3} \int_k F(\mathbf{k}) d^3\mathbf{k} \quad (23.5.1)$$

Usually when we apply (23.5.1) to finite, but macroscopic, systems we assume that the density at the left hand side differs only negligibly from its infinite volume limit (e.g it is the same for a crystal of 1 cm^3 and for another one of 2 cm^3)

23.6 Derivation of the Bloch function

Let us determine the form of the wave function of an electron in a (weak) periodic potential.

In 1D:

$$V(x + R) = V(x) \quad \text{and} \quad \psi(x + R) = \psi(x)$$

$$R = na, \quad n = 1, 2, \dots$$

For $n = 1$:

$$\begin{aligned} |\psi(x + a)|^2 &= |\psi(x)|^2 \\ \psi(x + a) &= C(a)\psi(x) \quad \text{therefore} \\ \psi(x + 2a) &= C(a)\psi(x + a) = C(a)C(a)\psi(x) = \\ &= C^2(a)\psi(x) \\ &\dots \end{aligned}$$

$$\psi(x + la) = C(a)\psi(x + (l - 1)a) = C(a)^l\psi(x) \quad (23.6.1)$$

...

Again, using periodic boundary conditions

$$\psi(x + Na) = \psi(x) \Rightarrow \psi(x + Na) = C^N(a)\psi(x) \quad (23.6.2)$$

i.e.

$$C^N(a) = 1 \quad (23.6.3)$$

$$C(a) = \sqrt[N]{1} = e^{i2\pi \frac{n}{N}} \quad n = 1, 2, \dots, N \quad (23.6.4)$$

$$C(a) = e^{ika} \quad \text{where} \quad k \equiv \frac{2\pi}{Na}n, \quad n = 1, 2, \dots, N \quad (23.6.5)$$

The number n has a maximum, because there are only N different unit roots. For $n > N$ the roots are the same as the ones already used, therefore k also have a maximum: $k_{max} = \frac{2\pi}{a}$

k is a vector of the reciprocal space, but it is a primitive vector of the reciprocal lattice only at its maximum.

$$k \in [0, 2\pi/a]$$

Instead of the range $[0, 2\pi/a]$ from (23.6.5) we generally use the equivalent $[-\pi/a, \pi/a]$ range (the first Brillouin zone).

From (23.6.3) and (23.6.1)

$$\psi(x) = \psi(x + na)e^{-ikna} \quad (23.6.6)$$

Multiply the right hand side of (23.6.5) with $e^{-ikx}e^{ikx} = 1$

$$\psi(x) = \psi(x + na)e^{-ik(x+na)}e^{ikx} \quad (23.6.7)$$

which may be written in the form:

$$\psi(x) = u(x)e^{ikx} \quad \text{where} \quad u(x + na) = u(x) \quad (23.6.8)$$

23.7 Kinetic energy of a Bloch electron

$$\begin{aligned} \mathcal{E}_{kin} &= \int \psi^* \frac{\hat{p}^2}{2m_e} \psi dx = \\ &= \int \psi^* \left(-\frac{\hbar^2}{2m_e} \frac{d^2}{dx^2} \right) \psi dx \\ &= -\frac{\hbar^2}{2m_e} \int \psi^* \frac{d^2}{dx^2} \psi dx \\ &= -\frac{\hbar^2}{2m_e} \int u^*(x) e^{-ikx} \frac{d^2}{dx^2} u(x) e^{ikx} dx \end{aligned}$$

$$\begin{aligned}
&= -\frac{\hbar^2}{2m_e} \int u^*(x) e^{-ikx} \frac{d}{dx} ((u'(x) + ik u(x)) e^{ikx}) dx \\
&= -\frac{\hbar^2}{2m_e} \int u^*(x) e^{-ikx} (u''(x) + ik u'(x) + ik u'(x) - k^2 u(x)) e^{ikx} dx \\
&= -\frac{\hbar^2}{2m_e} \int u^*(x) (u''(x) + 2ik u'(x)) dx - \frac{\hbar^2 k^2}{2m_e} \int u^*(x) u(x) dx \\
&= \frac{\hbar^2 k^2}{2m_e} - \frac{\hbar^2}{2m_e} \int u^*(x) (u''(x) + 2ik u'(x)) dx
\end{aligned}$$

$$\mathcal{E}_{kin}(k) = \frac{\hbar^2 k^2}{2m_e} + \mathcal{E}_{cryst}(k)$$

23.8 Tight-binding Bloch function

One dimensional calculation for a linear chain of atoms with a lattice constant a :

Let the atomic orbital (wave function) of a localized electron at the free atom be $\varphi(x)$ and create a Bloch function using a linear combination of such functions each localized around an atom²:

$$\psi(x) = \sum_{n=0}^{N-1} e^{ikna} \varphi(x - na) \quad (23.8.1)$$

This can be written as

$$\psi(x) = e^{ikx} \underbrace{\sum_{n=0}^{N-1} e^{-ik(x-na)} \varphi(x - na)}_{u(x)} \quad (23.8.2)$$

$$= u(x) e^{ikx} \quad (23.8.3)$$

²The wave functions $\varphi(x)$ may be the same as the actual atomic wave functions $\varphi_{n,l,m}(x)$, but any localized functions may be used instead.

This is a Bloch function, because $u(x)$ is a lattice periodic function:

$$\begin{aligned}
u(x) &= \sum_{n=0}^{N-1} e^{-ik(x-na)} \varphi(x-na) \\
u(x+a) &= \sum_{n=0}^{N-1} e^{-ik(x-(n-1)a)} \varphi(x-(n-1)a) \\
&= \sum_{l=-1}^{N-2} e^{-ik(x-la)} \varphi(x-la) \\
&= e^{-ik(x+a)} \varphi(x+a) \\
&\quad + \sum_{l=1}^{N-1} e^{-ik(x-la)} \varphi(x-la) \\
&\quad - e^{-ik(x-(N-1)a)} \varphi(x-(N-1)a)
\end{aligned}$$

After reordering:

$$u(x+a) = u(x) + e^{-ik(x+a)} \varphi(x+a) \quad (23.8.4)$$

$$- e^{-ik(x-(N-1)a)} \varphi(x-(N-1)a) \quad (23.8.5)$$

Even without using the periodic boundary condition we can argue that when N is very large ($\sim 10^{24}$) then the sum (that is $u(x)$) in (23.8.5) is much larger than the two terms with opposite signs outside it, therefore

$$u(x+a) = u(x)$$

To get the width of the band calculate the total energy:

$$\mathcal{E} = \frac{\int \psi^* H \psi dx}{\int \psi^* \psi dx} \quad (23.8.6)$$

where the Hamiltonian is written as the sum of the atomic Hamiltonians and a small perturbing periodic potential:

$$H = \sum_n H_{atomic}(x-na) + \Delta V_p(x)$$

Here

$$H_{atomic} = -\frac{\hbar^2}{2m_e} \frac{d^2}{dx^2} + V_{atomic}(x)$$

and $\Delta V_p(x+a) = \Delta V_p(x)$ is the periodic perturbing potential due to the overlap of wave functions.

For typographic reasons we will denote the numerator and denominator of (23.8.6) with \mathcal{N} and \mathcal{D} respectively.

The denominator of (23.8.6)

$$\begin{aligned}\mathcal{D} &:= \int \left(\sum_{n=0}^{N-1} e^{ikna} \varphi(x-na) \right)^* \cdot \left(\sum_{m=0}^{N-1} e^{ikma} \varphi(x-ma) \right) dx = \\ &= \sum_{n,m=0}^{N-1} \int (e^{-ikna} \varphi^*(x-na)) \cdot (e^{ikma} \varphi(x-ma)) dx = \\ &= \sum_{n,m=0}^{N-1} \int e^{ik(m-n)a} \varphi^*(x-na) \varphi(x-ma) dx\end{aligned}$$

The summation in \mathcal{D} can be split into sums containing integrals referring to the same atom ($\mathcal{D}_0\{n\}$), to nearest neighbors ($\mathcal{D}_1\{n, m\}$), to second nearest neighbors ($\mathcal{D}_2\{n, m\}$), etc.

$$\mathcal{D} = \sum_{\substack{n,m=0 \\ n=m}} I_0\{n, m\} + e^{\pm ika} \sum_{\substack{n,m=0 \\ n=m\pm 1}} I_1\{n, n \pm 1\} + e^{\pm i2ka} \sum_{\substack{n,m=0 \\ n=m\pm 2}} I_2\{n, n \pm 2\} + \dots \quad (23.8.7)$$

Similarly the numerator of (23.8.6)

$$\int \left(\sum_{n=0}^{N-1} e^{-ikna} \varphi^*(x-na) \right) \left(\sum_l H_{atomic}(x-la) + \Delta V_p(x) \right) \left(\sum_{m=0}^{N-1} e^{ikma} \varphi(x-ma) \right) dx$$

can also be split into similar sub-sums of n-th neighbor integrals. These will symbolically be denoted as $\mathcal{N}_0^H\{n\}$, $\mathcal{N}_1^H\{n, m\}$, $\mathcal{N}_2^H\{n, m\}$, etc and $\mathcal{N}_0^V\{n\}$, $\mathcal{N}_1^V\{n, m\}$, $\mathcal{N}_2^V\{n, m\}$, etc.

$$\begin{aligned}
\mathcal{N} &= \sum_{n,m,l=0}^{N-1} \int e^{-ikna} \varphi^*(x-na) H_{atomic}(x-la) e^{ikma} \varphi(x-ma) dx + \\
&+ \sum_{n,m=0}^{N-1} \int e^{-ikna} \varphi^*(x-na) \Delta V_p(x) e^{ikma} \varphi(x-ma) dx = \\
&= e^{ik(m-n)a} \sum_{n,m,l=0}^{N-1} \int \varphi^*(x-na) H_{atomic}(x-la) e^{ikma} \varphi(x-ma) dx + \\
&+ e^{ik(m-n)a} \sum_{n,m=0}^{N-1} \int \varphi^*(x-na) \Delta V_p(x) \varphi(x-ma) dx = \\
&= \sum_{\substack{n,m,l=0 \\ n=m=l}}^{N-1} \mathcal{N}_0^H\{n\} + e^{\pm ika} \sum_{\substack{n,m,l=0 \\ n=m\pm 1 \\ n=l\pm 1}}^{N-1} \mathcal{N}_1^H\{n,m\} + \dots + \\
&+ \sum_{\substack{n,m=0 \\ n=m}}^{N-1} \mathcal{N}_0^V\{n\} + e^{\pm ika} \sum_{\substack{n,m=0 \\ n=m\pm 1}}^{N-1} \mathcal{N}_1^V\{n,m\} + e^{\pm ika} \sum_{\substack{n,m=0 \\ n=m\pm 2}}^{N-1} \mathcal{N}_2^V\{n,m\}
\end{aligned}$$

The sums contain integrals of functions localized around the n -th and m -th atom and the integrals in the first sum additionally contain the atomic Hamiltonian of the l -th atom.

If we assume that only wave functions centered on the same or neighboring atoms can overlap then the only non zero terms will be $\mathcal{D}_0\{n\}$, $\mathcal{D}_1\{n, n \pm 1\}$, $\mathcal{N}_0^H\{n\}$, \dots , $\mathcal{N}_0^V\{n\}$ and $\mathcal{N}_1^V\{n, n \pm 1\}$.

From these if the φ functions are normalized then $\mathcal{D}_0\{n\} = 1$. $\mathcal{N}_0^H\{n\}$ is the atomic energy.

$$\begin{aligned}
\mathcal{N} &= \sum_n \mathcal{E}_{atomic,n} + \sum_n \mathcal{N}_0^V\{n\} + e^{ika} \sum_n \mathcal{N}_1^V\{n, n+1\} + e^{-ika} \sum_n \mathcal{N}_1^V\{n, n-1\} = \\
&= \mathcal{E}_{atomic} + \sum_n \mathcal{N}_0^V\{n\} + (e^{ika} + e^{-ika}) \sum_n \mathcal{N}_1^V\{n, n+1\} \\
\mathcal{D} &= 1 + e^{ika} \sum_n \mathcal{D}_1\{n, n+1\} + e^{-ika} \sum_n \mathcal{D}_1\{n, n-1\} = \\
&= 1 + (e^{ika} + e^{-ika}) \sum_n \mathcal{D}_1\{n, n+1\}
\end{aligned}$$

where we used that integrals $\mathcal{N}_1^V\{n, n+1\} = \mathcal{N}_1^V\{n, n-1\}$, and $\mathcal{D}_1\{n, n+1\} = \mathcal{D}_1\{n, n-1\}$.

Introducing the notations $\alpha \equiv \sum_n \mathcal{N}_0^V \{n\}$ and $\beta \equiv \sum_n \mathcal{N}_0^V \{n, n+1\}$ and observing that the second term in \mathcal{D} is very small, therefore

$$1/\mathcal{D} \approx (1 - (\text{small number}))$$

we arrive to the energy formula:

$$\mathcal{E} = \frac{\mathcal{N}}{\mathcal{D}} \approx \mathcal{E}_{atomic} - \alpha - 2\beta \cos ka \quad (23.8.8)$$

23.9 The explanation of the mass action law for semi-conductors

Let us again examine an n-type semiconductor. At temperature T the conduction electron concentration is

$$n_c(T) = n_c^d(T) + n_i(T)$$

But what is the hole concentration $p_v(T)$ in the valence band?

In thermal equilibrium the generation and the recombination of electron-hole pairs are in equilibrium. When donors are added to an intrinsic semiconductor the additional electrons from ionized donors may not only be excited to the conduction band but they can also recombine with holes in the valence band so the recombination rate for holes will increase, therefore the number of movable holes will decrease. At the same time the number of conduction electrons will also increase, because although some of the additional electrons may recombine with holes the remaining donor electrons will still go the conduction band. And as we saw the number of electrons due to the dopants are much higher than the intrinsic electron or hole concentration therefore the decrement in the concentration of dopant supplied electrons due to the electron-hole recombination is negligible while the effect the recombination has on the hole concentration is not. As a consequence in n-type semiconductors the electron concentration is higher while hole concentration is lower compared to those in an intrinsic semiconductor. When the donor concentration is high enough to be useful (but still much smaller than the concentration of the constituents of the semiconductor) the product $n_c \cdot p_v$ will approximately be the same as in an intrinsic semiconductor. The higher the dopant concentration (up to a point) the better is this approximation.

For p-type semiconductors the argument would be similar.

23.10 Fabrication of Si based integrated circuits

It starts with the creation of semiconductor *wafers* from extremely pure (only a few parts per million of impurities) Si. (The following section is mainly from Wikipedia.)

A typical wafer is made out of extremely pure silicon that is grown into mono-crystalline cylindrical ingots (boules) up to 300 mm (slightly less than 12 inches) in diameter using the *Czochralski process*. These ingots are then sliced into wafers about 0.75 mm thick and polished to obtain a very regular and flat surface.

Czochralski process: High-purity, semiconductor-grade silicon (only a few parts per million of impurities) is melted in a crucible, usually made of quartz. Dopant impurity atoms such as boron or phosphorus can be added to the molten silicon in precise amounts to dope the silicon, thus changing it into p-type or n-type silicon. This influences the electronic properties of the silicon. A precisely oriented rod-mounted seed crystal is dipped into the molten silicon. The seed crystal's rod is slowly pulled upwards and rotated simultaneously. By precisely controlling the temperature gradients, rate of pulling and speed of rotation, it is possible to extract a large, single-crystal, cylindrical ingot from the melt. Occurrence of unwanted instabilities in the melt can be avoided by investigating and visualizing the temperature and velocity fields during the crystal growth process. This process is normally performed in an inert atmosphere, such as argon, in an inert chamber, such as quartz.

Once the wafers are prepared, many process steps are necessary to produce the desired semiconductor integrated circuit. In general, the steps can be grouped into two major parts:

- Front-end-of-line (FEOL) processing : refers to the formation of the transistors directly in the silicon. This includes:
 - *Deposition* is any process that grows, coats, or otherwise transfers a material onto the wafer. Available technologies consist of physical vapor deposition (PVD), chemical vapor deposition (CVD), electrochemical deposition (ECD), molecular beam epitaxy (MBE) and more recently, atomic layer deposition (ALD) among others.
 - *Removal processes* are those that remove material from the wafer either in bulk or selectively and consist primarily of etch processes, either wet etching or dry etching. Chemical-mechanical planarization (CMP) is also a removal process used between levels.

- *Patterning* covers the series of processes that shape or alter the existing shape of the deposited materials and is generally referred to as lithography. For example, in conventional lithography, the wafer is coated with a chemical called a photoresist. The photoresist is exposed by a stepper, a machine that focuses, aligns, and moves the mask, exposing select portions of the wafer to short wavelength light. The unexposed regions are washed away by a developer solution. After etching or other processing, the remaining photoresist is removed by plasma ashing.
- Modification of electrical properties has historically consisted of doping transistor sources and drains originally by diffusion furnaces and later by ion implantation. These doping processes are followed by furnace anneal or in advanced devices, by rapid thermal anneal (RTA) which serve to activate the implanted dopants. Modification of electrical properties now also extends to reduction of dielectric constant in low-k insulating materials via exposure to ultraviolet light in UV processing (UVP).

Modern chips (2011) have up to eleven metal levels produced in over 300 sequenced processing steps.

- Back-end-of-line (BEOL) processing : refers to the formation of interconnections with external circuitry.

23.11 Determination of $n_c(x)$ and $p_v(x)$ in a p-n structure

Quantitatively for non-degenerate semiconductors:

$$\begin{aligned} n_c(x) &= N_c(T) e^{-(\mathcal{E}_c - e\varphi(x) - E_F)/k_B T} \\ p_v(x) &= P_v(T) e^{-(\mathcal{E}_F + e\varphi(x) - E_v)/k_B T} \end{aligned} \quad (23.11.1)$$

Far from the space charge region (for simplicity at $\pm\infty$ ³) and supposing the donor and acceptor atoms are all completely ionized in the whole crystal:

$$\begin{aligned} n_c(\infty) &= N_c(T) e^{-(\mathcal{E}_c - e\varphi(\infty) - E_F)/k_B T} = N_d \\ p_v(-\infty) &= P_v(T) e^{-(\mathcal{E}_F + e\varphi(-\infty) - E_v)/k_B T} = N_a \end{aligned} \quad (23.11.2)$$

³Because there are no free charge carriers there almost the whole of this potential difference is realized over the depletion region therefore this assumption will be valid for even small semiconductor samples.

From this the total potential difference between the two sides is

$$e\Delta\varphi = \underbrace{(\mathcal{E}_c - \mathcal{E}_v)}_{\mathcal{E}_g} + k_B T \ln \left(\frac{N_d N_a}{N_c P_v} \right) \quad (23.11.3)$$

But from (16.1.17)

$$\begin{aligned} \ln(n_i^2) &= \ln(N_c P_v) - \mathcal{E}_g / k_B T \quad \Rightarrow \quad -k_B T \ln(N_c P_v) \\ &= -E_g - k_B T \ln(n_i^2) \\ e\Delta\varphi &= k_B T \ln \left(\frac{N_d N_a}{n_i^2(T)} \right) \end{aligned} \quad (23.11.4)$$

The value of $\varphi(x)$ can be determined from the Poisson equation. In Si units:

$$\begin{aligned} -\frac{d^2\varphi(x)}{dx^2} &= \frac{\rho(x)}{\epsilon} \quad \text{where} \\ \rho(x) &= e [N_d(x) - N_a(x) + p_v(x) - n_c(x)] \end{aligned} \quad (23.11.5)$$

Substituting (23.11.2) into (23.11.1)

$$\begin{aligned} n_c(x) &= N_d e^{-e(\varphi(\infty) - \varphi(x)) / k_B T} \\ p_v(x) &= N_a e^{-e(\varphi(x) - \varphi(-\infty)) / k_B T} \end{aligned} \quad (23.11.6)$$

This equation can only be solved numerically, because the unknown function appears at the right hand side in the exponent. However a quite reasonable assumption is that the total change in $\varphi(x)$ is of order $\mathcal{E}_g \gg k_B T$. If the $\Delta\varphi$ change occurs in the interval $-d_p < x < d_n$ ($d_n + d_p$ is the width of the space charge or depletion region) then $n_c = N_d$ and $p_v = N_a$ outside this interval, which means $\rho = 0$. Within this region, except quite near to the boundaries, $e\varphi$ differs by many $k_B T$ from its asymptotic value, so $n_c \ll N_d$ on the n side and $p_v \ll N_a$ at the p side:

$$\rho = \begin{cases} 0 & x < d_p \\ e(N_d(x) - N_a(x)) & -d_p < x < d_n \\ 0 & x > d_n \end{cases}$$

23.12 Temperature dependent resistivity of materials

Scattering on crystal defects

Crystal defects destroy the periodicity of the crystal potential locally, so the Bloch model will break down there. This effect will be independent of the

temperature, because the scattering cross section of crystal defects:

$$\sigma_{s,def} = const.$$

is independent of the temperature⁴ Because the amplitude of the lattice vibrations will decrease with decreasing temperatures this will be the dominant scattering mechanism at very low temperatures.

Scattering on small amplitude lattice vibrations

Far from the melting point lattice vibrations increase the probability of a scattering, which may be taken into account as the increase of the scattering cross section of ions, which depend on the amplitude of the lattice vibrations:

$$\sigma_{s,vib} \propto A^2$$

A itself is proportional to the energy \mathcal{E}_{vib} of the lattice vibrations (See section 13.3):

$$\mathcal{E}_{vib} = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_B T} - 1}$$

Not too near to $T = 0K$ where $e^{h\nu/k_B T}$ is near to 1

$$\mathcal{E}_{vib} \approx \frac{8\pi h\nu^3}{c^3} \frac{k_B T}{h\nu}$$

therefore

$$\sigma_{s,vib} = const T$$

For the resultant τ then

$$\frac{1}{\tau} = \frac{1}{\tau_{def}} + \frac{1}{\tau_{vib}} = n_{s,def} v_F \sigma_{s,def} + n_{s,vib} v_F \sigma_{s,vib}$$

According to Eq.(16.1.1), the conductivity:

$$\sigma = \frac{ne^2\tau}{m_{eff}(E_F)}$$

The resistivity

$$\rho = \frac{1}{\sigma} = \frac{m_{eff}(E_F)}{ne^2} \frac{1}{\tau} = \frac{m_{eff}(E_F)}{ne^2} (n_{s,def} v_F \sigma_{s,def} + n_{s,vib} v_F \sigma_{s,vib})$$

Because of the temperature dependence of the cross sections

$$\rho = A + B T$$

⁴Strictly speaking this formula is not exactly true as the equilibrium vacancy concentration does depend somewhat on the temperature.

23.13 The explanation of the color of gold

The color of metals such as silver and gold is mainly due to absorption of light when a d electron jumps to an s orbital. For silver, the $4d \rightarrow 5s$ transition has an energy corresponding to ultraviolet light, so frequencies in the visible band are not absorbed. With all visible frequencies reflected equally, silver has no color of its own; it's silvery.

Silver and aluminum powders appear black because the white light that has been re-emitted is absorbed by nearby grains of powder and no light reaches the eye.

But to explain the yellow color of gold we need to turn to an unexpected direction: toward special relativity!

With an atomic number of 79, gold is in the last row of the periodic table containing stable elements, and only four stable elements (mercury, thallium, lead, and bismuth) have greater atomic number. With 79 protons in its nucleus, the electrons of the gold atom are subjected to an intense electrostatic attraction. Using the naïve classical Bohr model of the atom for the moment, electrons in the $1s$ orbital, closest to the nucleus, would have to orbit with a velocity v of $1.6 \cdot 10^8$ metres per second to have sufficient kinetic energy to avoid falling into the nucleus. This is more than half the speed of light, which, according to Einstein's equation increases the electron's momentum (older terminology its mass) by about 20%.

Quantum mechanics replaces the Bohr orbits with a probability distribution of the electron's position, with the Bohr orbit radius interpreted as the distance from the nucleus where the peak probability occurs. The relativistic increase in mass of the electron causes a relativistic contraction of its orbit because, as the electron's mass increases, the radius of an orbit with constant angular momentum shrinks proportionately.

So in gold, relativistic contraction of the s orbitals causes their energy levels to shift closer to those of the d orbitals (which are less affected by relativity). This, in turn, shifts the light absorption (primarily due to the $5d \rightarrow 6s$ transition) from the ultraviolet down into the lower energy and frequency blue visual range. A substance which absorbs blue light will reflect the rest of the spectrum: the reds and greens which, combined, result in the yellowish hue we call golden⁵.

⁵Special relativity is also responsible for gold's resistance to tarnishing and other chemical reactions. Chemistry is mostly concerned with the electrons in the outermost orbitals. With a single $6s$ electron, you might expect gold to be highly reactive; after all, caesium has the same $6s^1$ outer shell, and it is the most alkaline of natural elements: it explodes if dropped in water, and even reacts with ice. Gold's $6s$ orbital, however, is relativistically contracted toward the nucleus, and its electron has a high probability to be among the electrons of the filled inner shells. This, along with the stronger electrostatic attraction

23.14 Derivation of the Larmor formula

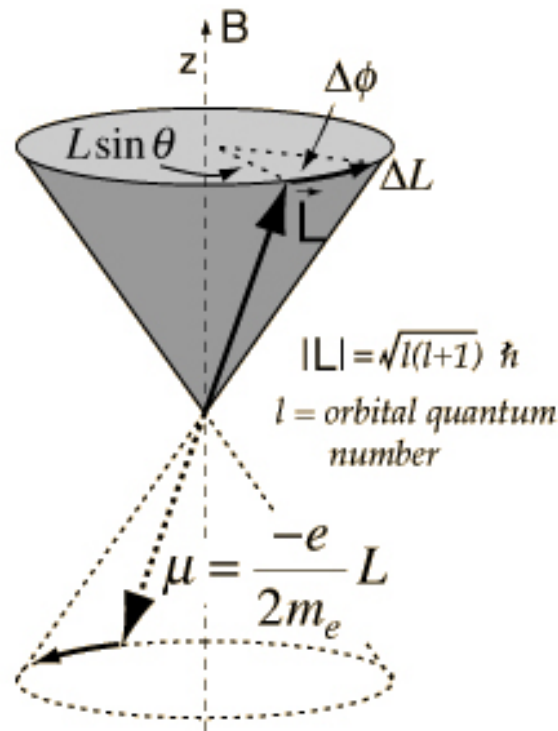


Figure 23.10: Principle of the Larmor precession

From classical mechanics:

$$\frac{d\mathbf{L}}{dt} = \mathbf{\Gamma}$$

where $\mathbf{\Gamma}$ denotes the torque.

$$\frac{d\mathbf{L}}{dt} = \mathbf{p}_m \times \mathbf{B} = \gamma \cdot \mathbf{L} \times \mathbf{B}$$

i.e. \mathbf{L} changes in a direction perpendicular to both itself and \mathbf{B} , so its endpoint will rotate around \mathbf{B} . Let us select the z -axis parallel with the direction of \mathbf{B} .

of the 79 protons in the nucleus, reduce the atomic radius of gold to 135 picometres compared to 260 picometres for caesium with its 55 protons and electrons – the gold atom is almost 50% heavier, yet only a little over half the size of caesium. Only the most reactive substances can tug gold's $6s^1$ electron out from where it's hiding among the others, and hence not only the color of gold, but its immunity from tarnishing and corrosion are consequences of special relativity. See http://www.fourmilab.ch/documents/golden_glow/

The trajectory of the endpoint of \mathbf{L} is a circle whose plane is perpendicular to \mathbf{B} (Fig. 23.10) and if we denote the angle between \mathbf{L} and \mathbf{B} with θ , then because

$$\left| \frac{d\mathbf{L}}{dt} \right| = |\gamma \cdot \mathbf{L} \times \mathbf{B}|$$

both the derivative and the vector product contain the sine of the θ angle:

$$\begin{aligned} \left| \frac{d\mathbf{L}}{dt} \right| &= \frac{d(L \cdot \sin \theta \cdot \varphi)}{dt} = L \cdot \sin \theta \cdot \frac{d\varphi}{dt} \\ |\gamma \cdot \mathbf{L} \times \mathbf{B}| &= \gamma \cdot L \cdot B \sin \theta \end{aligned}$$

therefore

$$L \cdot \sin \theta \cdot \frac{d\varphi}{dt} = |\gamma| \cdot L \cdot B \sin \theta$$

from here:

$$\omega \equiv \frac{d\varphi}{dt} = |\gamma| B = g \frac{e B}{2 m_e}$$

23.15 Calculating the Pauli paramagnetic moment of metals

Orient the z-axis in the direction of \mathbf{B} then at $0K$

$$n_{\uparrow\uparrow} = \frac{1}{2} \int_{-\mu_S B}^{\mathcal{E}_F} \frac{g(\mathcal{E} + \mu_S B)}{e^{-(\mathcal{E} - \mathcal{E}_F)/k_B T} + 1} d\mathcal{E} \quad (23.15.1)$$

$$n_{\uparrow\downarrow} = \frac{1}{2} \int_{\mu_S B}^{\mathcal{E}_F} \frac{g(\mathcal{E} - \mu_S B)}{e^{-(\mathcal{E} - \mathcal{E}_F)/k_B T} + 1} d\mathcal{E} \quad (23.15.2)$$

Here $g(\mathcal{E})$ is the (14.3.10) free electron density of state function:

$$g(\mathcal{E}) = \frac{8\pi \sqrt{2m_e^3}}{h^3} \sqrt{\mathcal{E}}$$

Substituting a new variable $E \equiv \mathcal{E} \pm \mu_S B$ into the integrals the lower limit of both integral become 0, while the upper limits change to $\mathcal{E}_F \pm \mu_S B$.

$$\begin{aligned} n_{\uparrow\uparrow} &= \frac{1}{2} \int_0^{\mathcal{E}_F + \mu_S B} \frac{g(E)}{e^{-(\mathcal{E} - \mathcal{E}_F + \mu_S B)/k_B T} + 1} d\mathcal{E} \\ n_{\uparrow\downarrow} &= \frac{1}{2} \int_0^{\mathcal{E}_F - \mu_S B} \frac{g(E)}{e^{-(\mathcal{E} - \mathcal{E}_F - \mu_S B)/k_B T} + 1} d\mathcal{E} \end{aligned}$$

If $\mu_S B \ll \mathcal{E}_F$ then the $\mu_S B$ is negligible in the exponent. If we denote the integral

$$\frac{1}{2} \int_0^{\mathcal{E}_F} \frac{g(E)}{e^{-(\mathcal{E}-\mathcal{E}_F-\mu_S B)/k_B T} + 1} d\mathcal{E}$$

with \mathcal{I} then

$$\begin{aligned} n_{\uparrow\uparrow} &= \mathcal{I} + \frac{1}{2} \int_{\mathcal{E}_F}^{\mathcal{E}_F + \mu_S B} \frac{g(E)}{e^{-(\mathcal{E}-\mathcal{E}_F + \mu_S B)/k_B T} + 1} d\mathcal{E} \\ n_{\uparrow\downarrow} &= \mathcal{I} + \frac{1}{2} \int_{\mathcal{E}_F}^{\mathcal{E}_F - \mu_S B} \frac{g(E)}{e^{-(\mathcal{E}-\mathcal{E}_F - \mu_S B)/k_B T} + 1} d\mathcal{E} \end{aligned}$$

In the limit of $k_B T \ll \mathcal{E}_F$

$$\begin{aligned} n_{\uparrow\uparrow} &= \mathcal{I} + \frac{1}{2} \int_{\mathcal{E}_F}^{\mathcal{E}_F + \mu_S B} g(E) d\mathcal{E} \approx \mathcal{I} + \frac{g(\mathcal{E}_F) \mu_S B}{2} \\ n_{\uparrow\downarrow} &= \mathcal{I} + \frac{1}{2} \int_{\mathcal{E}_F}^{\mathcal{E}_F - \mu_S B} g(E) d\mathcal{E} \approx \mathcal{I} - \frac{g(\mathcal{E}_F) \mu_S B}{2} \end{aligned}$$

The resulting magnetic moment (the magnetic polarization) is

$$\mathcal{M} = \mu_S \cdot (n_{\uparrow\uparrow} - n_{\uparrow\downarrow}) = g(\mathcal{E}_F) \mu_S^2 B \quad (23.15.3)$$

Substituting $g(\mathcal{E}_F)$ from (14.3.11)

$$\mathcal{M} = \mu_S \cdot (n_{\uparrow\uparrow} - n_{\uparrow\downarrow}) = \mu_S^2 g(\mathcal{E}_F) B = \frac{3 n_{tot} \mu_S^2}{2 e \mathcal{E}_F} B \quad (23.15.4)$$

where n_{tot} is the electron density in the metal and the paramagnetic susceptibility of the electron gas is

23.16 Derivation of the orientation polarization

From statistical physics the probability P_E that a molecule gains

$$\mathcal{E}_{pol} = -\mathbf{p}_e \cdot \mathbf{E} = -p_e E \cos \theta$$

energy is

$$P_E(\theta) = \frac{1}{Z} e^{-\mathcal{E}_{pol}/k_B T} = \frac{1}{Z} e^{p_e E \cos \theta / k_B T}$$

where Z is the sum of states:

$$\begin{aligned} Z &= \sum_{\substack{\text{unit volume} \\ \text{all polar angles}}} e^{-\mathcal{E}_{pol}/k_B T} = \\ &= \int_0^\pi \int_0^{2\pi} e^{p_e E \cos \theta / k_B T} \sin \theta d\varphi d\theta \end{aligned}$$

Because no term depends on φ the integral by φ gives just a 2π factor:

$$Z = 2\pi \int_0^\pi e^{p_e E \cos \theta / k_B T} \sin \theta d\theta$$

Substituting $\alpha \equiv p_e E / k_B T$ and $x \equiv \alpha \cos \theta$ ⁶

$$dx = \alpha(-\sin \theta) d\theta \quad \Rightarrow \quad \sin \theta d\theta = -\frac{dx}{\alpha}$$

$$\begin{aligned} Z &= -\frac{2\pi}{\alpha} \int_\alpha^{-\alpha} e^x dx = \frac{2\pi}{\alpha} [e^x]_{-\alpha}^\alpha \\ &= \frac{4\pi}{\alpha} \left[\frac{e^\alpha - e^{-\alpha}}{2} \right] = \frac{4\pi}{\alpha} \sinh \alpha \end{aligned}$$

For small values of alpha (i.e. when $p_e E \ll k_B T$) $\sinh \alpha \approx \alpha$ from which $Z = 4\pi$.
The number of those dipoles whose angle to the z-axis is θ is

$$n(\theta) = N P_E(\theta) = \frac{N}{Z} e^{p_e E \cos \theta / k_B T}$$

Summing up $n(\theta)$ naturally gives N .

The polarization density vector is the average polarization dipole moment of the unit volume. Because adding $p_e(\theta)$ vectors with opposite φ polar angles the components perpendicular to the z axis cancel each other out the resulting vector will point into direction z . The z component of the moment is $p_e \cos \theta$

$$\begin{aligned} \mathcal{P} &= \sum_{\substack{\text{unit volume} \\ \text{all } \theta \text{ and } \varphi}} p_e \cos \theta \cdot P_E(\theta) = \\ &= \frac{N}{Z} \int_0^\pi \int_0^{2\pi} n(\theta) (p_e \cos \theta) \sin \theta d\varphi d\theta = \\ &= \frac{2\pi N}{Z} \int_0^\pi n(\theta) p_e \cos \theta \sin \theta d\theta = \\ &= \frac{2\pi N}{Z} \int_0^\pi e^{-p_e E \cos \theta / k_B T} \cos \theta \sin \theta d\theta \end{aligned}$$

⁶In the following we use the hyperbolic *sine* and *cosine* and *cot* functions, which are defined by

$$\sinh \alpha = \frac{e^\alpha - e^{-\alpha}}{2}, \quad \cosh \alpha = \frac{e^\alpha + e^{-\alpha}}{2}, \quad , \quad \coth \alpha = \frac{\cosh \alpha}{\sinh \alpha} = \frac{e^\alpha - e^{-\alpha}}{e^\alpha + e^{-\alpha}}$$

Using the same substitutions as before and noting that p_e is independent of θ this becomes:

$$\mathcal{P} = \frac{2\pi N p_e}{\alpha^2 Z} \int_{-\alpha}^{\alpha} x e^x dx$$

Integrating by parts

$$\begin{aligned} \mathcal{P} &= \frac{N 2\pi p_e}{Z \alpha^2} [(x-1)e^x]_{-\alpha}^{\alpha} = \\ &= \frac{N 2\pi p_e}{Z \alpha^2} ((\alpha-1)e^{\alpha} - (-\alpha-1)e^{-\alpha}) = \\ &= \frac{N 2\pi p_e}{Z \alpha^2} (2\alpha \cosh \alpha - 2 \sinh \alpha) = \end{aligned}$$

Using the value of Z calculated above yields:

$$\begin{aligned} \mathcal{P} &= \frac{\frac{N 2\pi p_e}{\alpha^2} 2 (\alpha \cosh \alpha - \sinh \alpha)}{\frac{N 2\pi}{\alpha} 2 \sinh \alpha} = \\ &= N p_e \left(\coth \alpha - \frac{1}{\alpha} \right) \end{aligned}$$

We expect that in the high E field limit ($\alpha \rightarrow \infty$) all dipoles are turned into the direction of the field and in the small field limit ($\alpha \rightarrow 0$) we expect a polarization density proportional to the field⁷

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \mathcal{P} &= N p_e \cdot (1 - 0) = N p_e && \text{high field limit} \\ \lim_{\alpha \rightarrow 0} \mathcal{P} &= \frac{N p_e \alpha}{3} = \frac{N p_e^2}{3 k_B T} E && \text{low field limit} \end{aligned}$$

The latter one corresponds to (20.2.1).

23.17 Determination of the local electric field E_{loc}

The E field of a uniformly polarized sphere can be calculated using the definition of \mathbf{P} , the superposition principle and Gauss's law. The polarization

⁷The first case is trivial and the second one requires the use of the Taylor series of \coth , which – as it is easy to check using e.g. the Taylor series of the exponential function

$$\coth \alpha = \frac{1}{\alpha} + \frac{\alpha}{3} - \frac{\alpha^3}{45} + \frac{2\alpha^5}{945} + \dots$$

occurs, because the charge center of the positive and negative charge densities ($\pm\rho$) do not coincide, there is a distance l between them. According to the superposition principle

$$\mathbf{P} = \rho l$$

Of course neither ρ , nor l is known. We know however that l very small. So we may consider both the positive and negative charge density distributions are spherical with the same radius as that of the sphere's R . The field of a uniformly charged sphere is easy to calculate using Gauss's law, which gives:

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \frac{\rho r}{3\epsilon_o} & \text{if } r \geq R \\ \frac{\rho R^3}{3\epsilon_o} \frac{1}{r^2} & \text{if } r \leq R \end{cases}$$

Instead of adding up the field strengths directly it will be simpler to first calculate the $\varphi(r)$ potentials by integrating the field strengths, then adding the potentials of the two charged spheres together and finally calculate the electric field by derivation. The connection between \mathbf{E} and φ is:

$$\mathbf{E} = -\text{grad } \varphi \text{ and } \varphi = \int \mathbf{E} d^3r$$

From the second formula:

$$\begin{aligned} \varphi(r) &= -\int_{\infty}^R \mathbf{E}(\mathbf{r}') dr' - \int_R^r \mathbf{E}(\mathbf{r}') dr' = \\ &= \left[\frac{\rho R^3}{3\epsilon_o} \frac{1}{r^2} \right]_{\infty}^R - \left[\frac{\rho r^2}{3\epsilon_o} \right]_R^r = \\ &= \frac{\rho}{2\epsilon_o} \left(R^2 - \frac{r^2}{3} \right) \end{aligned}$$

The difference between the positive and negative charge distributions appears in the sign of ρ and in the fact that two distinct r^{\pm} must be used in place of r in the formula above. The resulting potential:

$$\varphi = \varphi_+(r) + \varphi_-(r) = \frac{\rho}{2 \cdot 3 \cdot \epsilon_o} (-r_+^2 + r_-^2) =$$

Observing that the difference of the square of the length of the two r_{\pm} vectors may be written as a product containing the vectors:

$$\varphi = \frac{\rho}{2 \cdot 3 \cdot \epsilon_o} (\mathbf{r}_- - \mathbf{r}_+) (\mathbf{r}_- + \mathbf{r}_+)$$

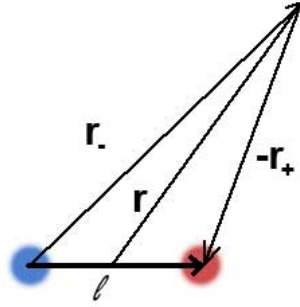


Figure 23.11:

$r_- - r_+ = l$. Using that $l \ll r$ it follows that $r_- \approx r_+ \approx r$, therefore $r_- + r_+ \approx 2r$:

$$\varphi = \frac{\rho l \mathbf{r}}{3 \epsilon_o} = \frac{\mathbf{P} \mathbf{r}}{3 \epsilon_o}$$

The resulting E_{plug} field then

$$E_{plug} = -grad \varphi = -\frac{\mathbf{P}}{3 \epsilon_o}$$