Introduction to nanotechnology

Henrik Bruus

MIC – Department of Micro and Nanotechnology
Technical University of Denmark

Lyngby, spring 2004

Preface

In the spring 2002 MIC launched a new fourth semester course at the Technical University of Denmark (course no. 33320, 10 ECTS) to provide a general and broad introduction to the multi-disciplinary field of nanotechnology. The number of students attending the course has grown steadily from 24 in 2002, to 35 the following year and now more than 50 in 2004. Based on the feed-back from the students I have changed part of the course and expanded the lecture notes.

The aim of the course remains the same. It is intended for students who have completed three semesters in any engineering or science study programme at college level. During the course the students will be introduced to many fascinating phenomena on the nanometer scale, and they will hopefully acquire basic knowledge of the theoretical concepts and experimental techniques behind the recent vastly improved ability to observe, fabricate and manipulate individual structures on the nanometer scale.

The first part of the course, which is covered by these lecture notes, is an introduction to the top-down approach of microelectronics and micromechanics. Here selectred topics like the AFM and quantum transport are studied in some detail. The second part has a much broader focus. Here the aim is to give the students an overview of the on-going merge of the top-down approach with the bottom-up approach of chemistry/biochemistry; a development that is creating new and exciting cross-disciplinary research fields and technologies. Much of the material used in this part of the course is provided by guest lecturers

Henrik Bruus
MIC – Department of Micro and Nanotechnology
Technical University of Denmark
26 January 2004

Contents

Chapter 1

Top-down micro and nanotechnology

Nanotechnology deals with natural and artificial structures on the nanometer scale, i.e. in the range from 1 μm down to 10 Å. One nanometer, 1 nm = 10^{-9} m, is roughly the distance from one end to the other of a line of five neighboring atoms in an ordinary solid. The nanometer scale can also be illustrated as in Fig. 1.1: if the size of a soccer ball ($\sim$ 30 cm = 3×10^{-1} m) is reduced 10.000 times we reach the width of a thin human hair ($\sim$ 30 μm = 3×10^{-5} m). If we reduce the size of the hair with the same factor, we reach the width of a carbon nanotube ($\sim$ 3 nm = 3×10^{-9} m).

It is quite remarkable, and very exciting indeed, that we today have a technology that involves manipulation of the ultimate building blocks of ordinary matter: single atoms and molecules.

Nanotechnology owes it existence to the astonishing development within the field of micro electronics. Since the invention of the integrated circuit nearly half a century ago in 1958, there has been an exponential growth in the number of transistors per micro chip and an associated decrease in the smallest width of the wires in the electronic circuits. As

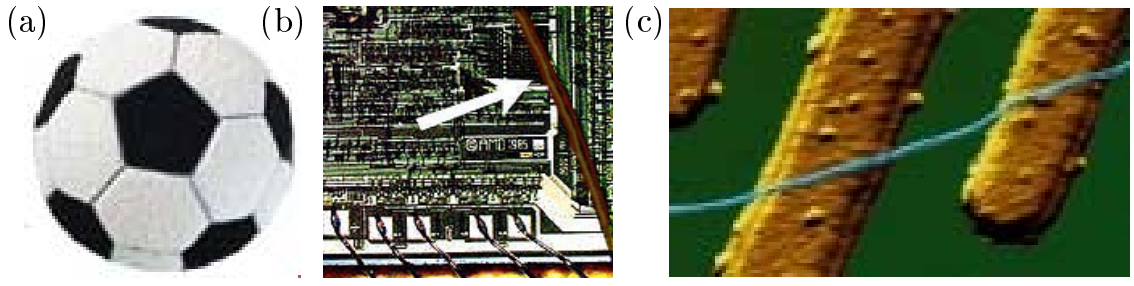


Figure 1.1: (a) A soccer ball with a diameter $\sim$ 30 cm = 3×10^{-1} m. (b) The width of a human hair (here placed on a microchip at the white arrow) is roughly 10^4 times, i.e. $\sim$ 30 μm = 3×10^{-5} m. (c) The diameter of a carbon nanotube (here placed on top of some metal electrodes) is yet another 10^4 times smaller, i.e. $\sim$ 3 nm = 3×10^{-9} m.

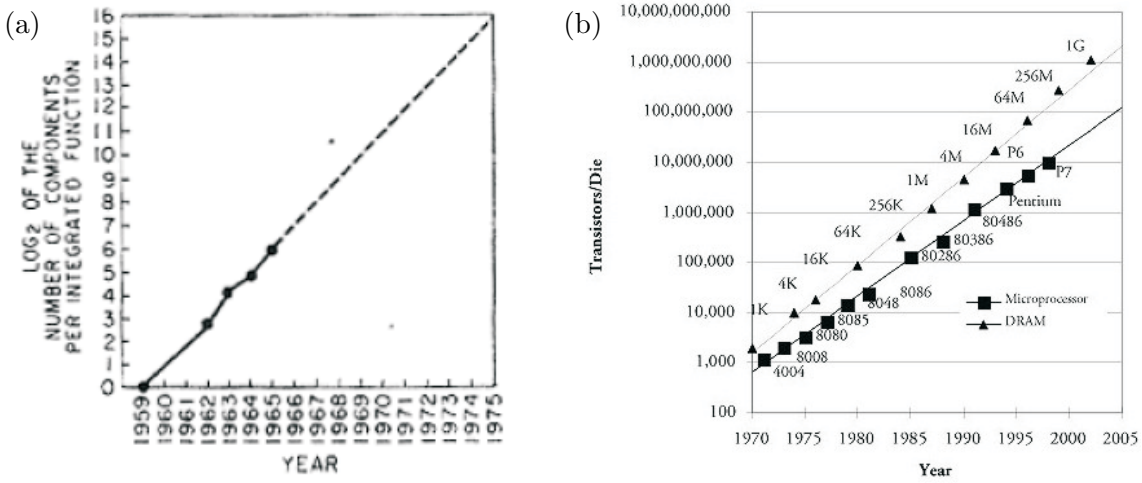


Figure 1.2: (a) Moore's law in the form of the original graph from 1965 suggesting a doubling of the number of components per microchip each year. (b) For the past 30 years Moore's law has been obeyed by the number of transistors in Intel processors and DRAM chips, however only with a doubling time of 18 months.

a result extremely powerful computers and efficient communication systems have emerged with a subsequent profound change in the daily lives of all of us.

A modern computer chip contains more than 10 million transistors, and the smallest wire width are incredibly small, now entering the sub 100 nm range. Just as the American microprocessor manufacturer, Intel, at the end of 2003 shipped its first high-volume 90 nm line width production to the market, the company announced that it expects to ramp its new 65 nm process in 2005 in the production of static RAM chips.[1] Nanotechnology with active components is now part of ordinary consumer products.

Conventional microtechnology is a top-down technology. This means that the microstructures are fabricated by manipulating a large piece of material, typically a silicon crystal, using processes like lithography, etching, and metallization. However, such an approach is not the only possibility. There is another remarkable consequence of the development of micro and nanotechnology.

Since the mid-1980'ies a number of very advanced instruments for observation and manipulation of individual atoms and molecules have been invented. Most notable are the atomic force microscope (AFM) and the scanning tunnel microscope (STM) that will be treated later in the lecture notes. These instruments have had en enormous impact on fundamental science as the key elements in numerous discoveries. The instruments have also boosted a new approach to technology denoted bottom-up, where instead of making small structure out over large structures, the small structures are made directly by assembling of molecules and atoms.

In the rest of this chapter we shall focus on the top-down approach, and describe some

[1]Learn more about the 65 nm SRAM at http://www.intel.com/labs/features/si11032.htm

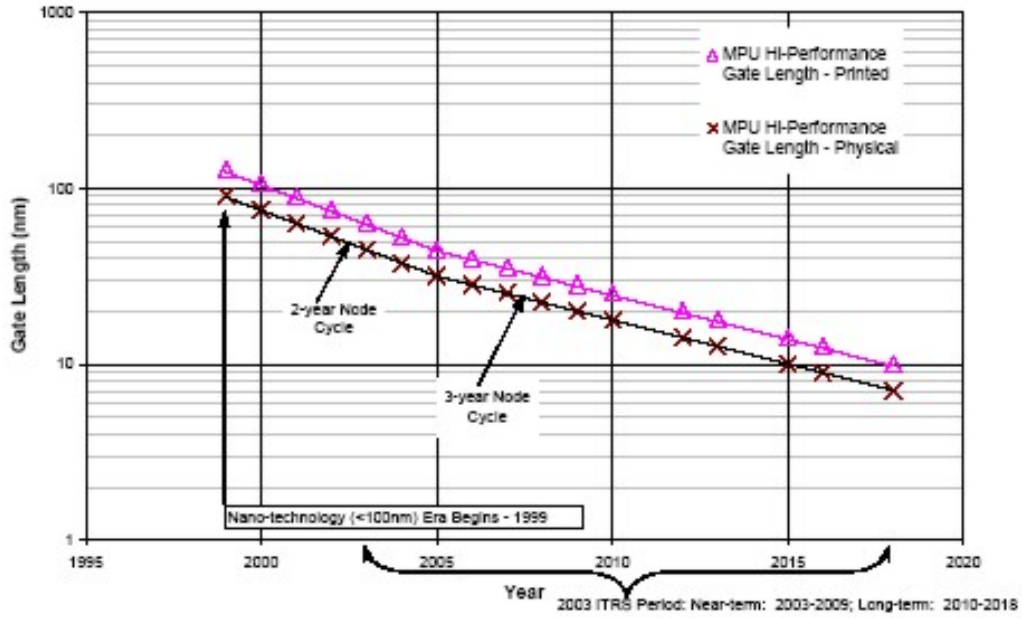


Figure 1.3: Moore's law applied to the shrinking of the length of the gate electrode in CMOS transistors. The length has deminished from about 100 nm in year 2000 to a projected length of 10 nm in 2015 [from the International Technology Roadmap for Semiconductors, 2003 Edition (http://public.itrs.net)].

of its main features.

1.1 Microfabrication and Moore's law

The top-down approach to microelectronics seems to be governed by an exponential time dependence. I 1965, when the most advanced integrated circuit contained only 64 transistors, Gordon E. Moore, Director of Fairchild Semiconductor Division, was the first to note this exponential behavior in his famous paper *Cramming more components onto integrated circuits* [Electronics, **38**, No. 8, April 19 (1965)]: "When unit cost is falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip". He observed a doubling of the number of transistors per circuit every year, a law that has become known as Moore's law. It is illustrated in Fig. 1.2.

Today there exist many other versions of Moore's law. One of them is shown in Fig. 1.3. It concerns the exponential decrease in the length of the gate electrode in standard CMOS transistors, and relates to the previous quoted values of 90 nm in 2003 and 65 nm in 2005. Naturally, there will be physical limitations to the exponential behavior expressed in Moore's law, see Exercise 1.1. However, also economic barriers play a major if not the decisive role in ending Moore's law developments. The price for constructing microprocessor fabrication units also rises exponentially for each generation of microchips. Soon the

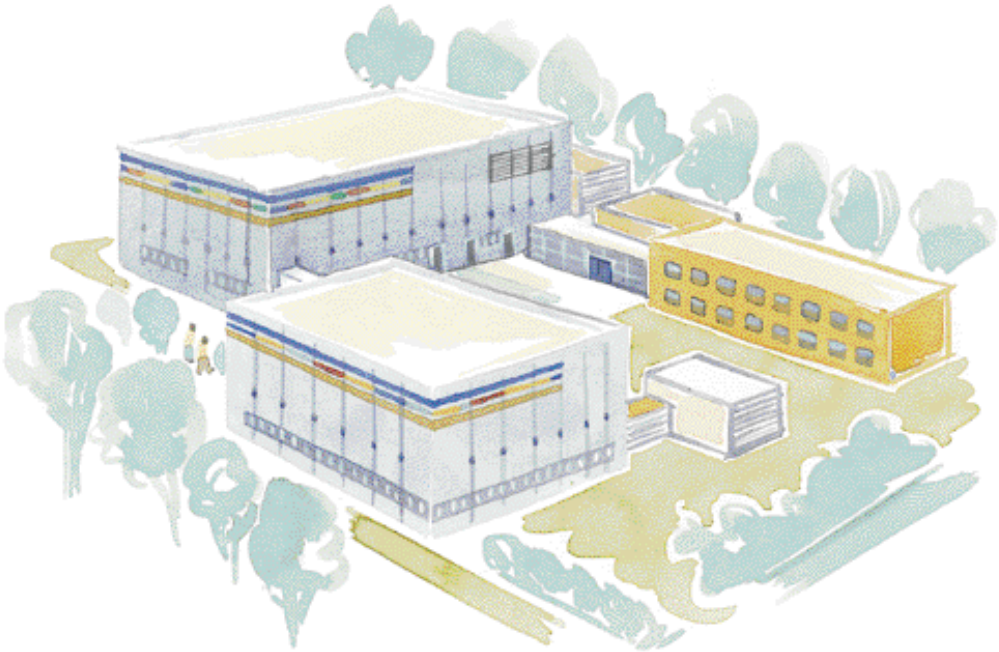


Figure 1.4: The clean room facilities DANCHIP, situated next to MIC at the Technical University of Denmark. The large building in background to the left is the original MIC clean room from 1992. The building in the front is under construction until the summer of 2004.

level is comparable to the gross national product of a mid-size country, and that might very well slow down the rate of progress.

1.2 Clean room facilities

The small geometrical features on a microchip necessitates the use of clean room facilities during the critical fabrication steps. Each cubic meter of air in ordinary laboratories may contain more than 10^7 particles with diameters larger than 500 nm. To avoid a huge flux of these "large" particles down on the chips containing micro and nanostructures, micro and nanofabrication laboratories are placed in so-called clean rooms equipped with high-efficiency particulate air (HEPA) filtering system. Such systems can retain nearly all particles with diameters down to 300 nm. Clean rooms are classified according to the maximum number of particles per cubic foot larger than 500 nm. Usually a class-1000 or class-100 clean room is sufficient for microfabrication.

The low particle concentration is ensured by keeping the air pressure inside the clean room slightly higher than the surroundings, and by combining the HEPA filter system with a laminar air flow system in the critical areas of the clean room. The latter system let the clean air enter from the perforated ceiling in a laminar flow and leave through the perforated floor. Moreover, all personnel in the clean room must be wearing a special suit covering the whole body to minimize the surprisingly huge emission of small particles from

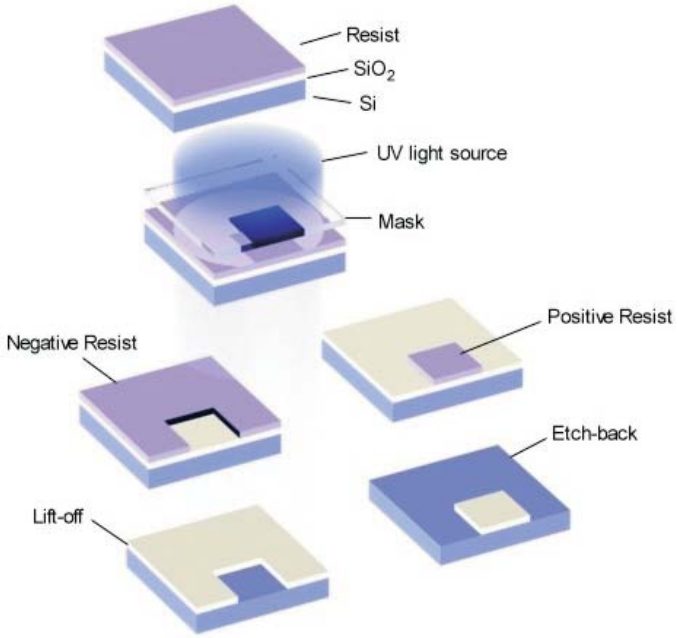


Figure 1.5: The basic principles of photolithography. The left-most figures illustrates the use of a negative resist to do lift-off. The right-most figures illustrates the use of a positive resist as an etch mask. See the text for more details.

each person.

The air flow inside the DANCHIP clean room is about 1.3×10^5 m^3 h^{-1}, most of which is recirculated particle-free air from the clean room itself. However, since the exhaust air from equipment and fume hoods is not recirculated, there is in intake of fresh air of 0.3×10^5 m^3 h^{-1}.

1.3 Photolithography

Almost all top-down manufacturing involves one or more photolithography fabrication steps, so we give a brief outline of this technique here. A generic photolithography process is sketched in Fig. 1.5. From a lightsource light is directed through a mask carrying the circuit design down onto the substrate wafer covered with a photo-sensitive film, denoted the photoresist. Depending on the local photo-exposure defined by the photolitographic mask the photoresist can be partly removed by a chemical developer leaving well-defined parts of the substrate wafer exposed to etching or metal deposition.

The substrate wafer is typically a very pure silicon disk with a thickness around 500 μm and diameter of 100 mm (for historic reasons denoted a 4 inch wafer). Wafers of different purities are purchased at various manufacturers.

The photolithographic mask contains (part of) the design of the microsystem that is to be fabricated. This design is created using computer-aided design (CAD) software. Once

completed the computer file containing the design is sent to a company producing the mask. At the company the design is transferred to a glass plate covered with a thin but non-transparent layer of chromium. The transfer process is normally based on either the relatively cheap and fast laser writing with a resolution of approximately 1.5 μm and a delivery time of around two weeks, or the expensive and rather slow electron beam writing with a resolution of 0.2 μm and a delivery time of several months.

The photo exposure is typically performed using the 356 nm UV line from a mercury lamp, but to achieve the line widths of sub 100 nm mentioned in Sec. 1.1 an extreme UV source or even an X-ray source is needed. To achieve the best resolution must minimize note only the wavelength λ of the exposure light, but also the distance d between the photolithographic mask and the photoresist-covered substrate wafer, and the thickness t of the photoresist layer. The minimum line width $w_{\min}$ is given by the approximate expression

$$w_{\min} = \frac{3}{2}\sqrt{\lambda(d+t)}. \tag{1.1}$$

If $d = 0$ nm the mask is touching the photo-resist. This situation, denoted contact printing, improves the resolution but wears down the mask. If $d > 0$ nm, a case denoted proximity printing, the resolution is pourer but the mask may last longer. It is difficult to obtain $w_{\min} < 2$ μm using standard UV photolithography.

The photoresist is a typically a melted and thus fluid polymer that is put on the substrate wafer, which then is rotated at more than 1000 rounds per minute to ensure an even and thin layer of resist spreading on the wafer. The photoresists carry exotic names like SU-8, PMMA, AZ4562 and Kodak 747. The solubility of the resists is proportional to the square of the molecular weight of the polymer. The photo-processes in a polymer photoresist will either cut the polymer chains in small pieces (chain scission) and thus lower the molecular weight, or they will induces cross-linking between the polymer chains and thus increase the molecular weight. The first type of resists is denoted the positive tone photoresists, they will be removed where they have been exposed to light. The second type is denoted the negative tone photoresists, they will remain where they have been exposed to light.

1.4 Electron beam lithography

To obtain resolutions better than the few μm of photolithography it is necessary to use either X-ray lithography or electron beam lithography. Here we give a brief overview of the latter technique.

After development of the resist one can choose to etch the exposed part of the wafer. Acid will typically not etch the polymer photoresist but only the substrate, so etching will carve out the design defined by the mask. The shape of the etching depends on the acid and the substrate. It can be isotropic and have the same etch rate in all spatial directions, or it can by anisotropic with a very large etch rate in some specific directions. One can choose the etching process that is most suitable for the design.

Metal deposition followed by lift-off is another core technique. Here a thin layer of metal (less than 500 nm) is deposited by evaporation technique on the substrate after de-

(a) (b)

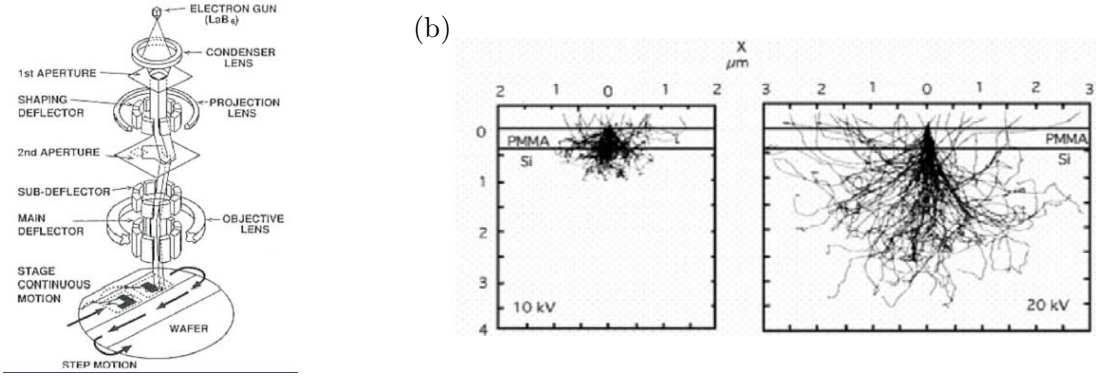


Figure 1.6: (a) The various components of an scanning electron microscope (SEM) from the electron gun in the top to the sample moved by a stepper motor in the bottom. (b) A computer simulation of the back scattering of electrons in the substrate. Note how the electrons enters the resist from the back in a much wider beam (500 - 1500 nm) than the incoming beam. The higher the energy (e.g., 20 kV) the higher is the spread of electrons (e.g., 1.5 μm), and the lower is the resolution.

veloping the resist. At the exposed places the metal is deposited directly on the substrate, and elsewhere the metal is residing on top of the remaining photoresist. After the metal deposition the substrate is rinsed in a chemical that dissolves the photoresist and thereby lift-off the metal residing on it. As a result a thin layer of metal is left on the surface of the wafer in the pattern defined by the photography mask.

The above mentioned process steps can be repeated many times with different masks and very complicated devices may be fabricated that way.

Electron beam lithography is based on a electron beam microscope, see Fig. 1.6, in which a focused beam of fast electrons are directed towards a resist-covered substrate. No mask is involved since the position of the electron beam can be controlled directly from a computer through electromagnetic lenses and deflectors.

The electrons are produced with an electron gun, either by thermal emission from hot tungsten filament or by cold field emission. The emitted electrons are then accelerated by electrodes with a potential $U \approx 10$ kV and the beam is focused by magnetic lenses and steered by electromagnetic deflectors.

As we shall discuss in great detail in Sec. 2.1 the electron is both a particle and a wave. The wavelength λ of an electron is given in terms by the momentum p of the electron and Planck's constant h by the de Broglie relation Eq. (2.3) $\lambda = h/p$. In the electron beam microscope the electron acquires a kinetic energy given by the acceleration voltage U as $\frac{1}{2}mv^2 = eU$, where m and $-e$ is the mass and charge of the electron, respectively. Since $p = mv$ the expression for the wavelength λ becomes

$$\lambda = \frac{h}{\sqrt{2meU}}, \tag{1.2}$$

which for a standard potential of 10 kV yields $\lambda = 0.012$ nm.

However, the resolution of an electron beam microscope is not given by λ. First of all, one can not focus the electron beam on such a small length scale. A typical beam spot size is around 0.1 nm. But more importantly are the scattering processes of the electrons inside the resist and the substrate. As illustrated by the computer simulation shown in Fig. 1.6(b) the backscattering of the electrons implies that an area much broader area is exposed to electrons than the area of the incoming electrons. This results in an increase of the resolution. It turns out that in practice it is difficult to get below a minimum linewidth of 10 nm.

Electron beam lithography is still the technique with the best resolution for lithography. A major drawback of the method is the long expose time required to cover an entire wafer with patterns. The exposure time t_{exp} is inversely proportional to the current I in the electron beam and proportional to the clearing dose D (required charge per area) and the exposed area A,

$$t_{\text{exp}} = \frac{DA}{I}. \tag{1.3}$$

This formula is discussed further in Exercise 1.3. In photolithography the entire wafer is exposed in one flash, like parallel processing, whereas in electron beam lithography it is necessary to write one pattern after the other in serial processing. For mass production electron beam lithography is therefore mainly used to fabricate masks for photolithography discussed in Sec. 1.3 and nanoimprint lithography discussed in Sec. 1.5.

1.5 Nanoimprint lithography

Nanoimprint lithography (NIL) is a relatively young technique compared to photolithography and electron beam lithography. The first results based on NIL was published in 1995.

The technique is in principle very simple. It consists of pressing machine that presses a stamp or mold containing the desired design down into a thin polymer film spun on top of a wafer and heated above its glass transition temperature. The basic principle of nanoimprint lithography is sketched in Fig. 1.7.

Naturally, a stamp is needed, and often its is produced by making a nanostructured surface in some wafer by use of electron beam lithography as described in Sec. 1.4 and subsequent etching techniques. Different materials have successfully been used as stamps among them silicon, silicon dioxide, metals, and polymers. Often it is necessary to coat the stamp with some anti-stiction coating to be able to release the stamp form the target material after pressing.

The target material is a polymer, which is useful for two reasons. First, above the glass transition temperature polymers are soft enough to make imprinting possible. Second, polymers can be functionalized to become sensitive various electric, magnetic, thermal, optical and biochemical input. Thus the resulting nanostructure can become very sophisticated indeed.

The pressing machine needs be able to deliver the necessary pressure. Moreover, it must contain an efficient temperature control in the form of heater plates and a thermostat,

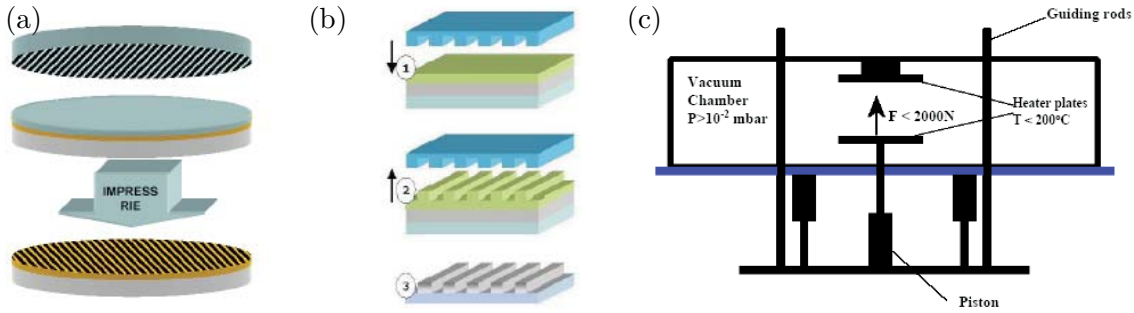


Figure 1.7: (a) The principle of nanoimprint lithography: after having made nanostructures in the surface of the stamp wafer using electron beam lithography, it is pressed into a thin layer of polymer film spun on top of a gold-plated silicon wafer. (b) A sketch of the setup before imprinting, after imprinting, and after etching. (c) A sketch of the pressing machine at MIC to used to perform nanoimprint lithography.

since it is crucial to operate at the correct temperature somewhat above the glass transition temperature of the polymer. To avoid impurities it must also operate under a sufficiently low vacuum, and finally it should allow for correct alignment of the sample before pressing.

Nanoimprint lithography is one of the few nanotechnologies that seems to be capable of mass production. Once the stamp is delivered, and the pressing machine is correctly set up, it should be possible to mass fabricate nanostructured wafer. The cycle time of a typical nanoimprint machine is of the order of minutes. This time scale is determined by the actual time it takes to press the stamp down and the various thermal time scales for heating and cooling of the sample.

Chapter 2

A brief intro to quantum physics

It is crucial to realize that the physics on the nanometer scale tends to become dominated by quantum physics. In the nanoworld one must always be prepared to take seemingly strange quantum phenomena into account and hence give up on an entirely classical description. Although this is not a course in quantum physics it is nevertheless imperative to get a grasp of the basic ideas and concepts of quantum physics. Without this it is not possible to reach a full understanding of the potentials of nanotechnology. Serious students of nanotechnology are hereby encouraged to study at least a minimum of quantum theory.

2.1 The particle-wave duality

The first laws of quantum physics dealt with energy quantization. They were discovered in studies of the electromagnetic radiation field by Planck and Einstein in 1900 and 1905, respectively. A new universal constant, Planck's constant h, was introduced in physics in addition to other constants like the speed of light, c, the gravitation constant, G, and the charge quantum, e. The 1998 CODATA values[1] for h and $\hbar = h/2\pi$ are

$$h = 6.62606876(52) \quad \times 10^{-34} \text{ Js}, \tag{2.1a}$$

$$\hbar = 1.054571596(82) \times 10^{-34} \text{ Js}. \tag{2.1b}$$

The energy E of light quanta (photons) in light of frequncy f or angular frequency $\omega = 2\pi f$ is given by

$$E = hf = \hbar\omega. \tag{2.2}$$

Already in 1906 Einstein applied this energy quantization on other objects than light, namely on oscillating atoms in his work on the heat capacity of solids. But it was first with Bohr's analysis of the stationary states in the hydrogen atom that quantum physics really proved to be essential for the understanding of not only the radiation field, but also of matter. With Heisenberg and Schrödinger's seminal papers from 1925 and 1926, respectively, modern quantum theory was born, a theory that eversince has been the foundation of our understanding of the physical world in which we live.

[1] See the NIST reference on constants, units, and uncertainty http://physics.nist.gov/cuu/

A central concept in quantum physics is the particle-wave duality, the fact that fundamental objects in the physical world, electrons, protons, neutrons, photons and other leptons, hadrons, and field quanta, all have the same dual nature: they are at the same time both particles and waves. In some situations the particle aspect may be the dominant feature, in other vice versa; but the behavior of any given object can never be understood fully by ascribing only one of these aspects to it.

Historically, as indicated, the particle-wave duality was first realized for the electromagnetic field. It is interesting to note that right from the beginning when the first theories of the nature of light was proposed in the 17th century, it was debated whether light were particles (corpuscles), as claimend by Newton, or waves, as claimed by Huygens. The debate appeared to end in the beginning of the 19th century with Young's famous double-slit interference experiments that demonstrated that light were waves. With Maxwell's theory (1873) and Hertz's experiments (1888) the light waves were shown to be of electromagnetic nature. But after nearly one hundred years of wave dominance Planck's formula for the energy distribution in black-body radiation demonstrated that light possesses some element of particle nature. This particle aspect became more evident with Einstein's explanation of the photoelectric effect: small energy parcels of light, the so-called light quanta, are able to knock out electrons from metals just like one billiard ball hitting another. For some years theorists tried to give alternative explanations of the photo-electric effect using maxwellian waves, but it proved impossible to account for the concentration of energy in a small point needed to explain the photoelectric effect without postulating the existence of light quanta – today called photons.

In 1913 Niels Bohr published his theory of the hydrogen atom, explaining its stability in terms of stationary states. Bohr did not explain why stationary states exist. He boldly postulated their existence and from that assumption he could explain the frequencies of the experimentally observed spectral lines, and in particular he could derive Balmer's empirical expression for the position of the spectral lines. He also derived a formula for the Rydberg constant appearing in that expression. In the following years it was postulated still without an explanation that a particle of momentum $\mathbf{p} = m\mathbf{v}$, where m and $\mathbf{v}$ is its mass and velocity, respectively, moving in a closed orbit must obey the Bohr-Sommerfeld quantization rule, $\oint \mathbf{p} \cdot d\mathbf{r} = nh$, where the integral is over one revolution and n is an integer.

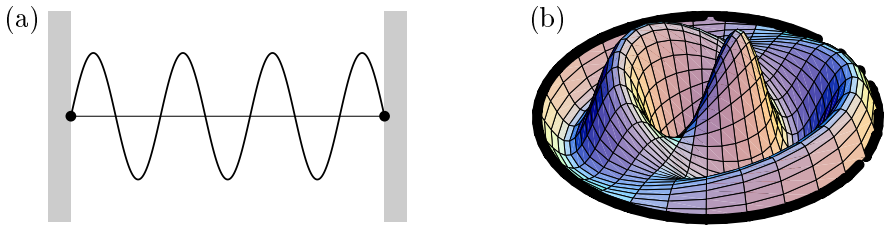


Figure 2.1: Resonance modes or eigenmodes in one and two dimensions: (a) a vibrating string described by $\sin(kx)$, and (b) a vibrating membrane described in polar coordinates in terms of a Bessel function by $J_1(kr)\cos(\phi)$.

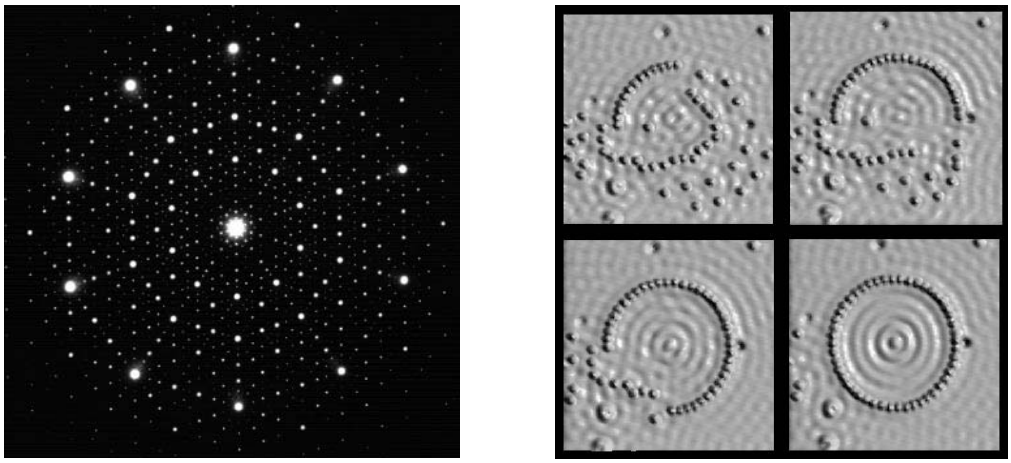


Figure 2.2: Two strong evidences for the existence of electron waves. (a) An electron diffraction from the quasicrystal $Al_{70}Co_{11}Ni_{19}$; by S. Ritsch et al., Phil. Mag. Lett. **80**, 107 (2000). (b) Electron waves on the surface of copper detected by scanning tunnel microscopy (STM). The waves are trapped inside a ring of iron atoms. The ring is created by pushing the iron atomes around on the copper surface using an atomic force microscope. By M.F. Crommie, C.P. Lutz, and D.M. Eigler, Science **262**, 218-220 (1993).

In 1923 de Broglie proposed that particles could be ascribed a pilot wave guiding their motion through space. This idea was based on the fundamental idea of duality: if light waves were also particles, then ought not also particles be waves? Moreover, if particles were waves it would somehow be possible to explain Bohr's stationary states as a kind of standing particle-waves. Think of a vibrating string fixed in both ends: only certain resonance frequencies are possible corresponding to matching an integral number of half-waves between the endpoints. Likewise for higher dimensional bodies as sketched in Fig. 2.1.

Below we shall go through a simple argument that leads to de Broglie's famous relation between the momentum p of a particle and the wave length λ (or wave number $k = 2\pi/\lambda$) of its associated wave,

$$p = \frac{h}{\lambda} = \hbar k. \tag{2.3}$$

Shortly after de Broglie's proposal Davisson and Germer verified his ideas by demonstrating the existence of diffraction patterns when electrons are shot through a thin metal film. The regularly spaced atoms in the film constituted a multi-slit analogue of Young's double-slit experiment for light. In the beginning of 1926 Schrödinger published his wave equation providing a firm mathematical foundation for de Broglie's ideas.

Today there can be no doubt about the reality of the wave nature of particles. Many experiments show this convincingly. In fact, all our understanding of matter on the microscopic level is based on the existence of these waves. Two particularly beautiful observations of electron waves are shown in Fig. 2.2.

2.2 de Broglie waves

Consider a given particle of mass m moving along the x axis with momentum p. If a wave $\psi(x,t)$ is to be associated with this motion, what is then the relation between the wavelength λ and p? In the following we shall show that the answer is the de Broglie relation Eq. (2.3), stating that $p = h/\lambda = \hbar k$.

For the wave description to be reasonable we demand that the wave has a high amplitude at the position x of the particle, i.e. $\psi(x,t)$ must be a wave packet that moves with the velocity $v = p/m$. Such a wave packet can be written in the form

$$\psi(x,t) = \int_{-\infty}^{\infty} dk\, g(k) e^{i(kx-\omega_k t)}, \tag{2.4}$$

where the weight function $g(k)$ is peaked around $k = k_0$. In these notes $i = \sqrt{-1}$. The wider $g(k)$ is in k space the narrower $\psi(x,t)$ is in x space. An example of a wave packet with $g(k) = \exp[-L^2(k-k_0)^2]$ is shown in Fig. 2.3. The k dependence of the frequency ω_k is called the dispersion relation for the particle. If $g(k)$ is sufficiently narrow, we can expand the dispersion relation around $k = k_0$,

$$\omega_k \approx \omega_0 + \frac{\partial \omega}{\partial k}\,(k-k_0), \tag{2.5}$$

where $\omega_0 = \omega_{k_0}$. We insert this expansion in Eq. (2.4) and use $q = k - k_0$ to obtain

$$\psi(x,t) \approx \int_{-\infty}^{\infty} dk\, g(k) e^{i[kx-(\omega_0+\frac{\partial\omega}{\partial k}(k-k_0))t]} \tag{2.6a}$$

$$= \left[\int_{-\infty}^{\infty} dq\, g(k_0+q) e^{iq[x-\frac{\partial\omega}{\partial k}t]}\right] e^{i(k_0x-\omega_0 t)} \tag{2.6b}$$

$$= f\Big(x - \frac{\partial\omega}{\partial k}t\Big)\, e^{i(k_0x-\omega_0 t)}. \tag{2.6c}$$

In this expression the argument of the function f is the only place containing information about the movement of the peak of the wave packet. If at $t = 0$ the peak is at $x = x_0$, then at time t its postion is given by $x(t) = x_0 + \frac{\partial\omega}{\partial k}t$, which clearly means, that its velocity v is given by

$$v = \frac{\partial\omega}{\partial k}. \tag{2.7}$$

This velocity is also known as the group velocity, since it refers to the movement of the shape of the group of waves forming the wave packet. The phase factor $\exp[i(k_0x - \omega_0 t)]$ contains another velocity, ω_0/k_0, the so-called phase velocity, but this is not related in any way to the shape of the wave packet.

The last step is to note that in classical mechanics the kinetic energy is given by $E = \frac{1}{2}mv^2 = \frac{p^2}{2m}$. From this and from Eqs. (2.2) and (2.7) we find

$$v = \frac{p}{m} = \frac{\partial E}{\partial p}, \tag{2.8a}$$

$$v = \frac{\hbar\partial\omega}{\hbar\partial k} = \frac{\partial E}{\partial(\hbar k)}. \tag{2.8b}$$

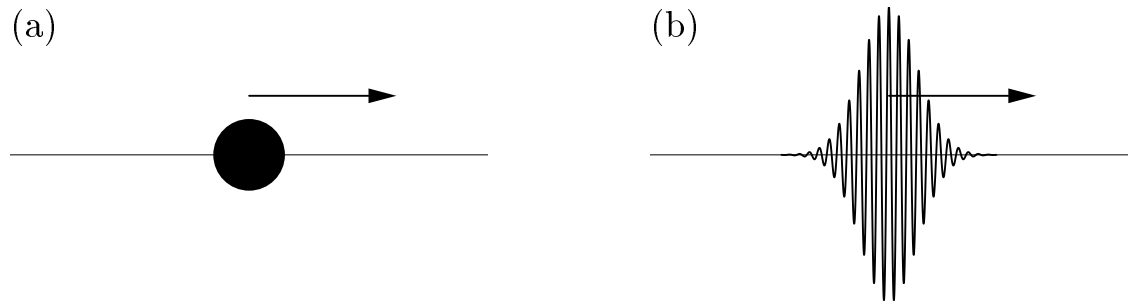


Figure 2.3: (a) A particle moving with speed v along the x-axis. (b) A wave packet with a center motion identical to that of the particle. This particular wave packet shown at $t = 0$ is obtiained from $\psi(x) = \int_{-\infty}^{\infty} dk \, \exp[-L^2(k-k_0)^2]e^{ikx} = (\sqrt{\pi}/L) \exp(-x^2/4L^2)e^{ik_0x}$ with $k = 16\pi$ and $L = 1$.

These equations become consistent if we make the following identification,

$$p = \hbar k = \frac{h}{\lambda}, \tag{2.9}$$

which in fact is the de Broglie relation.

The wave picture leads to a very fundamental relation between the uncertainty in position Δx and the uncertainty in wave number Δk. A clear example is shown in Fig. 2.3. Here the uncertainty Δk in k around k_0 is of the order $1/L$ as seen from the gaussian $g(k) = \exp[-L^2(k - k_0)^2]$, while the uncertainty in x is of the order $2L$ as seen from the gaussian $\exp(-x^2/4L^2)$. The more well defined the position is the less well defined is the wave vector and vice versa. This is summarized in the inequality $\Delta x \, \Delta k \geq 1$, which after multiplication with $\hbar$ becomes the famous Heisenberg uncertainty relation

$$\Delta x \, \Delta p \geq \hbar. \tag{2.10}$$

In the extreme quantum limit, e.g. for the ground state of a system the inequality becomes an equality. Furthermore, Δx becomes the characteristic length x_c, while Δp becomes the characteristic momentum p_c:

$$p_c \simeq \frac{\hbar}{x_c}. \tag{2.11}$$

2.3 The quantum pressure

Let us immediately focus on a very essential physical consequence of the wave nature of matter contained in the de Broglie relation, namely the appearance of a quantum pressure. Imagine we want to localize a particle by putting it in a little cubic box with side length L. Mathematically, localization is described by demanding $\psi = 0$ at the walls of the cube and everwhere outside it. The longest possible wave length $\lambda_{\max}$ of the particle in the box is given by the half wave length condition $\lambda_{\max}/2 = L$. The smallest possible kinetic

energy of the particle in the box is thus given by

$$E_{\min} = \frac{p_x^2 + p_y^2 + p_z^2}{2m} = \frac{3}{2m}\left(\frac{h}{\lambda_{\max}}\right)^2 = \frac{3h^2}{8mL^2} \quad \underset{L\to 0}{\longrightarrow} \quad \infty. \tag{2.12}$$

We see that it costs energy to localize a particle in a little region of space. The somewhat strange quantum nature of a particle reveals itself as a very concrete pressure acting against localization of the particle. This is the so-called quantum pressure, a very central concept in modern physics. Among other things it determines the size of atoms, the speed of sound in solids, the height of mountains on the Earth, and the size of neutron stars. Since $E_{\min} \propto L^{-2} = V^{-\frac{2}{3}}$ the quantum pressure $P_{\rm qu}$ becomes

$$P_{\rm qu} = -\frac{\partial E_{\min}}{\partial V} = \frac{2}{3}\,\frac{E_{\min}}{V}\,. \tag{2.13}$$

As studied in Exercise 2.3 the combination of quantum pressure and electrostatics defines the size of atoms. If the characteristic radius of the hydrogen atom is denoted a we can use of the Heisenberg equality of Eq. (2.11) with $x_c = a$ to rewrite the expression for the classical energy as follows:

$$E(a) \;=\; \frac{p^2}{2m} - \frac{e^2}{4\pi\epsilon_0}\,\frac{1}{a} \;\simeq\; \frac{\hbar^2}{2m}\,\frac{1}{a^2} - \frac{e^2}{4\pi\epsilon_0}\,\frac{1}{a}. \tag{2.14}$$

Finding the size a_0 that minimizes $E(a)$ we arrive (a bit by chance given the approximation) at the famous expression for the Bohr radius of the hydrogen atom,

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} = 0.053 \text{ nm}, \tag{2.15}$$

and for the ground state energy $E(a_0)$,

$$E_0 = E(a_0) = -\frac{1}{2}\,\frac{1}{4\pi\epsilon_0}\,\frac{e^2}{a_0} = -\frac{\hbar^2}{2ma_0^2} = -\frac{me^4}{2(4\pi\epsilon_0)^2\hbar^2} = -13.6 \text{ eV}. \tag{2.16}$$

2.4 The Schrödinger equation in one dimension

In 1926 Schrödinger wrote down the first proper wave equation for de Broglie's matter waves. The time-dependent Schrödinger equation for the wave function $\psi(x,t)$ describing a single non-relativistic particle of mass m moving in one dimension in a potential $V(x)$ is

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2}\psi(x,t) + V(x)\psi(x,t). \tag{2.17}$$

The Schrödinger equation forms the foundation of quantum physics, and as such it cannot logically be derived from classical physics. However, we will nevertheless give a heuristic derivation of it in the following.

A wave equation for a given field is a partial differential equation that describes the local evolution of the field in any given space-time point (x,t) in terms of the partial

derivatives $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial x}$. Since quantum behavior through Eqs. (2.2) and (2.3) has been linked to energy, $E = \hbar\omega$ and momentum $p = \hbar k$, it is natural to seek a wave equation corresponding to the classical energy expression containing both these quantities,

$$E = \frac{p^2}{2m} + V(x). \tag{2.18}$$

But how do we get the partial derivatives into play? And how do we take the spatially varying potential into account? Well, the clue comes from considering at first a pure plane wave, $\psi_0(x,t) = e^{i(kx-\omega t)}$, together with the quantum conditions Eqs. (2.2) and (2.3):

$$i\hbar\frac{\partial}{\partial t}\, e^{i(kx-\omega t)} = \hbar\omega\, e^{i(kx-\omega t)} = E\, e^{i(kx-\omega t)} \tag{2.19a}$$

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\, e^{i(kx-\omega t)} = \hbar k\, e^{i(kx-\omega t)} = p\, e^{i(kx-\omega t)}. \tag{2.19b}$$

This result is then generalized to the case of any shape of the wave function $\psi(x,t)$ by making the fundamental assumption of quantum theory, namely that the classical energy E and momentum p are replaced with the following differential operators, the so-called Hamiltonian operator $\hat{H}$ and the momentum operator $\hat{p}$:

$$E \quad \longrightarrow \quad \hat{H} \quad = \quad i\hbar\frac{\partial}{\partial t}, \tag{2.20a}$$

$$p \quad \longrightarrow \quad \hat{p} \quad = \quad \frac{\hbar}{i}\frac{\partial}{\partial x}. \tag{2.20b}$$

The Schrödinger equation Eq. (2.17) now follows by direct substitution,

$$E = \frac{p^2}{2m} + V(x) \quad \longrightarrow \quad \hat{H} = \frac{\hat{p}^2}{2m} + V(x) \quad \longrightarrow \quad i\hbar\frac{\partial}{\partial t} = -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2} + V(x). \tag{2.21}$$

2.5 The Schrödinger equation in three dimensions

It is a simple matter to extend the results of the previous section to three spatial dimensions and write down the corresponding time-dependent Scrödinger equation. Using cartesian coordinates, positions are given by $\mathbf{r} = (x,y,z)$ and momenta by $\mathbf{p} = (p_x, p_y, p_z)$. We derive the Schrödinger equation for a particle of mass m moving in the potential $V(x,y,z)$ by applying the quantum rule Eq. (2.20b) for each spatial direction. The energy

$$E = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(x,y,z) \tag{2.22}$$

becomes the time-dependent Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}\,\psi(x,y,z,t) = \left[-\frac{\hbar^2}{2m}\,\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V(x,y,z)\right]\psi(x,y,z,t). \tag{2.23}$$

This is often written in the abbreviated form,

$$i\hbar\frac{\partial}{\partial t}\,\psi(\mathbf{r},t) = \left[-\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2 \;+\; V(\mathbf{r})\right]\psi(\mathbf{r},t), \tag{2.24}$$

and the momentum operator $\hat{p}$ itself as

$$\hat{\mathbf{p}} \quad = \quad \frac{\hbar}{i}\left(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right) \quad = \quad \frac{\hbar}{i}\,\boldsymbol{\nabla}. \tag{2.25}$$

Often the symmetry of a given problem makes it obvious which particular coordinate system to choose. For example, the Coulomb potential around a single proton is spherical symmetric hence making spherical coordinates (r,θ,ϕ) the natural choise in studies of the electron wave functions in the hydrogen atom. The form of the differential operators nabla, $\boldsymbol{\nabla}$, and the Laplacian, $\boldsymbol{\nabla}^2$, in Cartesian, cylindrical, and spherical coordinates are the following:

a) Cartesian coordinates with basis vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$:

$$\boldsymbol{\nabla} \;= \mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z}, \tag{2.26a}$$

$$\boldsymbol{\nabla}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{2.26b}$$

b) Cylindrical coordinates with basis vectors $\mathbf{e}_r$, $\mathbf{e}_\phi$, and $\mathbf{e}_z$:

$$\boldsymbol{\nabla} \;= \mathbf{e}_r\frac{\partial}{\partial r} + \mathbf{e}_\phi\frac{1}{r}\frac{\partial}{\partial \phi} + \mathbf{e}_z\frac{\partial}{\partial z}, \tag{2.27a}$$

$$\boldsymbol{\nabla}^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}. \tag{2.27b}$$

c) Spherical coordinates with basis vectors $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_\phi$:

$$\boldsymbol{\nabla} \;= \mathbf{e}_r\frac{\partial}{\partial r} + \mathbf{e}_\theta\frac{1}{r}\frac{\partial}{\partial \theta} + \mathbf{e}_\phi\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}, \tag{2.28a}$$

$$\boldsymbol{\nabla}^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial \phi^2}. \tag{2.28b}$$

2.6 Superposition and interference of quantum waves

The Schrödinger equation is a linear differential equation. This means that if ψ_1 and ψ_2 are two solutions of it, then so is the sum $\psi = \psi_1 + \psi_2$. In other words, quantum waves obey the superposition principles in analogy with electromagnetic waves. It follows immediatly that interference is possible,

$$|\psi|^2 = |\psi_1 + \psi_2|^2 = |\psi_1|^2 + |\psi_2|^2 + 2\mathrm{Re}\,(\psi_1^*\psi_2). \tag{2.29}$$

We see that the total probability density $|\psi|^2$ is not merely the sum of the two partial probability densities $|\psi_1|^2$ and $|\psi_2|^2$. The interference term $2\mathrm{Re}(\psi_1^*\psi_2)$ has to be taken into account, and the measurement of interference effects like the electron diffraction pattern shown in Fig. 2.2a is crucial in demonstrating the existence of quantum waves.

Another consequence of the linearity of the Schrödinger equation is the abstract formulation of quantum physics in terms of a particular vector space, the so-called Hilbert space. Each wavefunction can be thought of as a vector in this abstract vector space. Addition of these vectors are possible due to the superposition principle. We refer the reader to any standard text on quantum mechanics for further studies of the Hilbert space formulation of quantum theory.

2.7 Energy eigenstates

A particularly important class of wave functions are the so-called energy eigenfunctions. They are characterized by having a well defined energy at all times, i.e. they correspond to Bohr's idea of stationary states. We try to find solutions of the form

$$\psi(\mathbf{r},t) = \psi_E(\mathbf{r})\, e^{-i\frac{E}{\hbar}t}. \tag{2.30}$$

Almost by definition this wave function can be ascribed the energy E since

$$i\hbar\frac{\partial}{\partial t}\psi_E(\mathbf{r})\, e^{-i\frac{E}{\hbar}t} = E\,\psi_E(\mathbf{r})\, e^{-i\frac{E}{\hbar}t}. \tag{2.31}$$

If we insert this particular form of $\psi(\mathbf{r},t)$ into the time-dependent Schrödinger equation Eq. (2.17) and divide out the common factor $\exp[-i\frac{E}{\hbar}t]$, we end up with the so-called time-independent Schrödinger equation,

$$\left[-\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2 + V(\mathbf{r})\right]\psi_E(\mathbf{r}) = E\psi_E(\mathbf{r}). \tag{2.32}$$

Eigenvalue problems like the time-independent Schrödinger equation Eq. (2.31) are solved by finding both the wavefunction $\psi_E(\mathbf{r})$ and the corresponding eigenenergy E. Often, as we shall see later, the set of possible values E is restricted to a discrete set. The state $\psi_E(\mathbf{r})$ is seen to be the stationary states originally proposed by Bohr.

2.8 The interpretation of the wavefunction ψ

Let us now address the question of how to interpret the wavefunction that now has been associated with a given particle. It would certainly be desirable if the wavefunction ψ has a large amplitude where the particle is present, and a vanishing amplitude where the particle is absent. We can be guided from our experience with light and photons.

2.8.1 The intensity argument

The wavefunction for light is the electric field $\mathbf{E}(\mathbf{r},t)$, while the intensity $I(\mathbf{r},t)$ is proportional to the square of the electric field. On the other hand, using the photon representation of light, we find that the intensity is proportional with the local number of photons per volume, the photon density $n(\mathbf{r},t)$,

$$n(\mathbf{r},t) \propto I(\mathbf{r},t) \propto |\mathbf{E}(\mathbf{r},t)|^2. \tag{2.33}$$

Naturally, the total number of photons $N(t)$ at time t is found by integrating over the entire volume $\mathcal{V}$,

$$N(t) = \int_{\mathcal{V}} d\mathbf{r}\, n(\mathbf{r},t). \tag{2.34}$$

Returning to the quantum wavefunction $\psi(\mathbf{r},t)$ of a single particle we postulate in analogy with Eq. (2.33) that the intensity $|\psi(\mathbf{r},t)|^2$ of the wavefunction is related to the particle density n and write

$$n(\mathbf{r},t) \propto I(\mathbf{r},t) \propto |\psi(\mathbf{r},t)|^2. \tag{2.35}$$

But since we have only one particle the analogy of Eq. (2.34) reduces to

$$1 = \int_{\mathcal{V}} d\mathbf{r}\, n(\mathbf{r},t). \tag{2.36}$$

This looks like the condition for a probability distribution. The total probability for finding the particle somewhere in space at time t is indeed unity. In fact, the particle and wave description can be reconciled by the fundamental postulate of quantum physics:

$\lvert\psi(\mathbf{r},t)\rvert^2\, d\mathbf{r}$	is the probability that the particle is in the volume $d\mathbf{r}$ around the point $\mathbf{r}$ at time t.
$\psi(\mathbf{r},t)$	itself is denoted the probability amplitude.

(2.37)

The probability interpretation of the wave function means that the wave function ψ must alway be normalized to unity. Therefore all wavefunctions $C\psi$ proportional to ψ represent the same physical state.

2.8.2 The continuity equation argument

The probability interpretation can be supported by the following analysis. If $n(\mathbf{r},t) = |\psi(\mathbf{r},t)|^2$ really is a probablity density, then, since a non-relativistic particle as the one studied here cannot disappear, it must obey a continuity equation

$$\frac{\partial n}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{J}. \tag{2.38}$$

Here $\mathbf{J}$ is some probability current density to be determined. Such an equation is well-known in electromagnetism, where n is the electric charge density and $\mathbf{J}$ is the electric

current density. The probability current density $\mathbf{J}$ is found using $n = |\psi|^2 = \psi^*\psi$ and the Schrödinger equation as follows. Begin by calculating the time derivative:

$$\frac{\partial n}{\partial t} = \frac{\partial \psi^*\psi}{\partial t} = \psi^* \frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi^*}{\partial t}. \tag{2.39}$$

Then write down the Schrödinger equation for ψ and its complex conjugate ψ^*,

$$\frac{\partial \psi}{\partial t} = +\frac{1}{i\hbar}\left[-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi + V\,\psi\right], \tag{2.40a}$$

$$\frac{\partial \psi^*}{\partial t} = -\frac{1}{i\hbar}\left[-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi^* + V\,\psi^*\right]. \tag{2.40b}$$

The equation for ψ is multiplied by ψ^*, and the equation for ψ^* is multiplied by ψ. The resulting equations are added. The V-terms cancel and using Eq. (2.39), leads to

$$\frac{\partial \psi^*\psi}{\partial t} = -\frac{\hbar}{2mi}\left[\psi^*\boldsymbol{\nabla}^2\psi - \psi\boldsymbol{\nabla}^2\psi^*\right]. \tag{2.41}$$

Now it is a simple matter to rewrite this as a divergence,

$$\frac{\partial \psi^*\psi}{\partial t} = -\frac{\hbar}{2mi}\boldsymbol{\nabla}\cdot\left[\psi^*\boldsymbol{\nabla}\psi - \psi\boldsymbol{\nabla}\psi^*\right]. \tag{2.42}$$

Furthermore, it is tempting to bring in the momentum operator $\hat{\mathbf{p}} = (\hbar/i)\boldsymbol{\nabla}$, which yields

$$\frac{\partial \psi^*\psi}{\partial t} = -\boldsymbol{\nabla}\cdot\left[\frac{1}{2m}\left\{\psi^*\left(\frac{\hbar}{i}\boldsymbol{\nabla}\psi\right) + \psi\left(\frac{\hbar}{i}\boldsymbol{\nabla}\psi\right)^*\right\}\right] = -\boldsymbol{\nabla}\cdot\left[\frac{1}{2m}\Big\{\psi^*(\hat{\mathbf{p}}\psi) + \psi(\hat{\mathbf{p}}\psi)^*\Big\}\right]. \tag{2.43}$$

We finally arrive at

$$\frac{\partial \psi^*\psi}{\partial t} = -\boldsymbol{\nabla}\cdot\frac{1}{m}\mathrm{Re}\Big[\psi^*(\hat{\mathbf{p}}\psi)\Big]. \tag{2.44}$$

The continuity equation Eq. (2.38) for the probability density is established if we simply identify $\mathbf{J}$ as

$$\mathbf{J}(\mathbf{r},t) = \frac{1}{m}\,\mathrm{Re}\Big[\psi^*(\hat{\mathbf{p}}\psi)\Big]. \tag{2.45}$$

Fortunately for the probability interpretation this is a very reasonable result. Recall that the classical expression for current density is $\mathbf{J} = n\,\mathbf{v} = n\,\mathbf{p}/m$. Given that $n = \psi^*\psi$, a naïve guess for the quantum expression would be $\mathbf{J} = \psi^*\psi\,\mathbf{p}/m$, which in fact is not far from the exact result Eq. (2.45). The latter is just slightly more complicated due to $\hat{\mathbf{p}}$ being a differential operator and not a number.

2.8.3 Quantum operators and their expectation values

We have already met some quantum operators, e.g. the energy operator $\hat{H} = i\hbar\frac{\partial}{\partial t}$, the momentum operator $\hat{\mathbf{p}} = \frac{\hbar}{i}\boldsymbol{\nabla}$, and the potential operator $\hat{V} = V(\mathbf{r})$. When an operator $\mathcal{O}$ acts on a given wavefunction ψ the result is in general not proportional to the wavefunction,

$\mathcal{O}\psi \neq \text{const}\ \psi$. How should one find the value of the operator? The answer comes from probability theory. Since the wavefunction has been identified with a probability amplitude, one is reminded of how mean values are calculated in classical probability theory. If X is some stochastic variable and $P(X)$ is the probability distribution of X, then the mean value or expectation value $\langle f(X)\rangle$ of any function f of the stochastic variable is given by

$$\langle f(X)\rangle = \int dX\ f(X)\, P(X). \tag{2.46}$$

Since ψ loosely speaking is the squareroot of P we arrive at the following definition of the expectation value $\langle\hat{\mathcal{O}}\rangle_\psi$ of quantum operator $\mathcal{O}$ acting on the state ψ:

$$\langle\hat{\mathcal{O}}\rangle_\psi = \int d\mathbf{r}\ \psi^*(\mathbf{r})\,\hat{\mathcal{O}}\psi(\mathbf{r}). \tag{2.47}$$

We shall use this equation many times.

2.9 Many-particle quantum states

In general a physical system contains more than one particle, and in that case we need to extend the wavefunction formalism.

2.9.1 The N-particle wavefunction

Consider a system containing N identical particles, say, electrons. Let the positions of the particles be given by the the N vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N$. The wavefunction of the system is given by $\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, which is a complex function in the $3N$-dimensional configuration space. The first fundamental postulate for many-particle systems is to interpret the N-particle wavefunction as a probability amplitude such that its absolute square is related to a probability,

$\lvert\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)\rvert^2 \prod_{j=1}^N d\mathbf{r}_j$	is the probability that the N particles are in the $3N$-dimensional volume $\prod_{j=1}^N d\mathbf{r}_j$ surrounding the point $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ in the $3N$-dimensional configuration space.

(2.48)

2.9.2 Permutation symmetry and indistinguishability

A fundamental difference between classical and quantum mechanics concerns the concept of indistinguishability of identical particles. In classical mechanics each particle in an many-particle system can be equipped with an identifying marker (e.g. a colored spot on a billiard ball) without influencing its behavior, and moreover it follows its own continuous path in phase space. Thus in principle each particle in a group of identical particles can be identified. This is not so in quantum mechanics. Not even in principle is it possible to mark a particle without influencing its physical state; and worse, if a number of identical

particles are brought to the same region in space, their wavefunctions will rapidly spread out and overlap with one another, thereby soon render it impossible to say which particle is where. It has for example no strict physical meaning to say that particle 1 is in state ψ_{ν_1} and particle 2 is in state ψ_{ν_2}, instead we can only say that two particles are occupying the two states ψ_{ν_1} and ψ_{ν_2}. The second fundamental assumption for N-particle systems is therefore that

> **Identical particles**, i.e. particles characterized by the same quantum numbers such as mass, charge and spin, **are in principle indistinguishable**. (2.49)

From the indistinguishability of particles follows that if two coordinates in an N-particle state function are interchanged the same physical state results, and the corresponding state function can at most differ from the original one by a simple prefactor C. If the same two coordinates then are interchanged a second time, we end with the exact same state function,

$$\psi(\mathbf{r}_1,..,\mathbf{r}_j,..,\mathbf{r}_k,..,\mathbf{r}_N) = C\psi(\mathbf{r}_1,..,\mathbf{r}_k,..,\mathbf{r}_j,..,\mathbf{r}_N) = C^2\psi(\mathbf{r}_1,..,\mathbf{r}_j,..,\mathbf{r}_k,..,\mathbf{r}_N), \quad (2.50)$$

and we conclude that $C^2 = 1$ or $C = \pm 1$. Only two species of particles are thus possible in quantum physics, the so-called bosons and fermions:

$$\psi(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_k,\ldots,\mathbf{r}_N) = +\psi(\mathbf{r}_1,\ldots,\mathbf{r}_k,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_N) \quad \text{(bosons)}, \quad (2.51a)$$

$$\psi(\mathbf{r}_1,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_k,\ldots,\mathbf{r}_N) = -\psi(\mathbf{r}_1,\ldots,\mathbf{r}_k,\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_N) \quad \text{(fermions)}. \quad (2.51b)$$

The importance of the assumption of indistinguishability of particles in quantum physics cannot be exaggerated, and it has been introduced due to overwhelming experimental evidence. For fermions it immediately leads to the Pauli exclusion principle stating that two fermions cannot occupy the same state, because if in Eq. (2.51b) we let $\mathbf{r}_j = \mathbf{r}_k$ then $\psi = 0$ follows. It thus explains the periodic table of the elements, and consequently forms the starting point in our understanding of atomic physics, condensed matter physics and chemistry. It furthermore plays a fundamental role in the studies of the nature of stars and of the scattering processes in high energy physics. For bosons the assumption is necessary to understand Planck's radiation law for the electromagnetic field, and spectacular phenomena like Bose–Einstein condensation, superfluidity and laser light.

2.9.3 Fermions: wavefunctions and occupation number

Consider two identical fermions, e.g. electrons, that are occupying two single-particle states ψ_{ν_1} and ψ_{ν_2}. It is useful to realize how to write down the two-particle wavefunction $\psi_{\nu_1\nu_2}(\mathbf{r}_1,\mathbf{r}_2)$ that explicitly fulfill the fermionic antisymmetry condition Eq. (2.51b). It is easy to verify that the following function will do the job,

$$\psi_{\nu_1\nu_2}(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}\Big[\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2) - \psi_{\nu_1}(\mathbf{r}_2)\psi_{\nu_2}(\mathbf{r}_1)\Big], \quad (2.52)$$

where the prefactor is inserted to ensure the correct normalization:

$$\int d\mathbf{r}_1 \int d\mathbf{r}_2 \, |\psi_{\nu_1\nu_2}(\mathbf{r}_1, \mathbf{r}_2)|^2 = 1. \tag{2.53}$$

In the case of N fermions occupying N orbitals $\psi_{\nu_1}, \psi_{\nu_2} \dots \psi_{\nu_N}$ the operator $\hat{S}_-$ that anti-symmetrizes the simple product of the N single-particle wavefunctions is the so-called Slater determinant

$$\psi_{\nu_1 \dots \nu_N}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \hat{S}_-[\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2) \dots \psi_{\nu_N}(\mathbf{r}_N)] \tag{2.54a}$$

$$= \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{\nu_1}(\mathbf{r}_1) & \psi_{\nu_1}(\mathbf{r}_2) & \dots & \psi_{\nu_1}(\mathbf{r}_N) \\ \psi_{\nu_2}(\mathbf{r}_1) & \psi_{\nu_2}(\mathbf{r}_2) & \dots & \psi_{\nu_2}(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\nu_N}(\mathbf{r}_1) & \psi_{\nu_N}(\mathbf{r}_2) & \dots & \psi_{\nu_N}(\mathbf{r}_N) \end{vmatrix}. \tag{2.54b}$$

Now that we know a little about what the fermion N-particle states look like, we turn to the question of what is the probability for having a given orbital occupied at non-zero temperature. Let n_ν be the occupation number for the state ψ_ν. At zero temperature n_ν is either constantly 0 or constantly 1. As the temperature T is raised from zero n_ν begins to fluctuate between 0 and 1. The thermal mean value $\langle n_\nu \rangle$ is denoted $f_F(\varepsilon_\nu)$, where ε_ν is the eigenenergy of ψ_ν. From statistical mechanics we know that the for a system with a fluctuating particle number thermal averages are calculated using the weight factor $\exp\big[-\beta(n_\nu\varepsilon_\nu - \mu n_\nu)\big]$, where $\beta = 1/k_B T$ and μ is the chemical potential.

$$f_F(\varepsilon_\nu) = \langle n_\nu \rangle = \frac{\sum\limits_{n_\nu=0,1} n_\nu e^{-\beta(n_\nu\varepsilon_\nu - \mu n_\nu)}}{\sum\limits_{n_\nu=0,1} e^{-\beta(n_\nu\varepsilon_\nu - \mu n_\nu)}} = \frac{0 + e^{-\beta(\varepsilon_\nu - \mu)}}{1 + e^{-\beta(\varepsilon_\nu - \mu)}} = \frac{1}{e^{\beta(\varepsilon_\nu - \mu)} + 1}. \tag{2.55}$$

This is the famous Fermi–Dirac distribution, which we shall use later in the course.

2.9.4 Bosons: wavefunctions and occupation number

Consider now two identical bosons, e.g. helium atoms, that are occupying two single-particle states ψ_{ν_1} and ψ_{ν_2}. The two-particle wavefunction $\psi_{\nu_1\nu_2}(\mathbf{r}_1, \mathbf{r}_2)$ that explicitly fulfill the bosonic symmetry condition Eq. (2.51a) is

$$\psi_{\nu_1\nu_2}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\Big[\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2) + \psi_{\nu_1}(\mathbf{r}_2)\psi_{\nu_2}(\mathbf{r}_1)\Big], \tag{2.56}$$

where the prefactor is to ensure the correct normalization:

$$\int d\mathbf{r}_1 \int d\mathbf{r}_2 \, |\psi_{\nu_1\nu_2}(\mathbf{r}_1, \mathbf{r}_2)|^2 = 1. \tag{2.57}$$

In contrast to the fermionic case there can be any number of bosons in a given orbital. In the case of N bosons occupying N orbitals (of which some may be the same) the operator

$\hat{S}_+$ that symmetrizes the N single-particle wavefunction $\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2)\dots\psi_{\nu_N}(\mathbf{r}_N)$ can be written as "sign-less" determinant

$$\psi_{\nu_1\dots\nu_N}(\mathbf{r}_1,\mathbf{r}_2,\dots,\mathbf{r}_N) = \hat{S}_+[\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2)\dots\psi_{\nu_N}(\mathbf{r}_N)] \tag{2.58a}$$

$$= \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{\nu_1}(\mathbf{r}_1) & \psi_{\nu_1}(\mathbf{r}_2) & \dots & \psi_{\nu_1}(\mathbf{r}_N) \\ \psi_{\nu_2}(\mathbf{r}_1) & \psi_{\nu_2}(\mathbf{r}_2) & \dots & \psi_{\nu_2}(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\nu_N}(\mathbf{r}_1) & \psi_{\nu_N}(\mathbf{r}_2) & \dots & \psi_{\nu_N}(\mathbf{r}_N) \end{vmatrix}_+ . \tag{2.58b}$$

The subsript "+" indicates that upon expanding this "determinant" all signs are taken to be plus.

Next we find the distribution function for bosons. Again using the grand canonical ensemble we derive the equally famous Bose–Einstein distribution $f_B(\varepsilon)$. It is derived like its fermionic counterpart, the Fermi–Dirac distribution $f_F(\varepsilon)$.

Consider a bosonic state characterized by its single-particle energy ε_ν. The occupation number of the state can be any non-negative integer $n_\nu = 0, 1, 2, \dots$. In the grand canonical ensemble the average occupation number $\langle f_B(\varepsilon_\nu)\rangle$ is found by writing $\lambda_\nu = e^{-\beta(\varepsilon_\nu-\mu)}$ and using the formulas $\sum_{n=0}^{\infty} n\lambda^n = \lambda\frac{d}{d\lambda}\sum_{n=0}^{\infty}\lambda^n$ and $\sum_{n=0}^{\infty}\lambda^n = \frac{1}{1-\lambda}$:

$$f_B(\varepsilon_\nu) = \frac{\sum_{n_\nu=0}^{\infty} n_\nu\, e^{-\beta(n_\nu\varepsilon_\nu-\mu n_\nu)}}{\sum_{n_\nu=0}^{\infty} e^{-\beta(n_\nu\varepsilon_\nu-\mu n_\nu)}} = \frac{\lambda_\nu\frac{d}{d\lambda_\nu}\sum_{n_\nu=0}^{\infty}\lambda_\nu^{n_\nu}}{\sum_{n_\nu=0}^{\infty}\lambda_\nu^{n_\nu}} = \frac{\frac{\lambda_\nu}{(1-\lambda_\nu)^2}}{\frac{1}{1-\lambda_\nu}} = \frac{1}{e^{\beta(\varepsilon_\nu-\mu)}-1}. \tag{2.59}$$

The Bose–Einstein distribution differs from the Fermi–Dirac distribution by having -1 in the denominator instead of $+1$. Both distributions converge towards the classical Maxwell–Boltzmann distribution, $f_M(n_\nu) = e^{-\beta(\varepsilon_\nu-\mu)}$, for very small occupation numbers, where the particular particle statistics is not felt very strongly.

2.9.5 Operators acting on many-particle states

The third fundamental postulate for many-particle systems is that the total action of few-particle operators are given by summation of their action on the individual coordinates. For a one-particle operator $\hat{\mathcal{O}}_1(\mathbf{r}_i)$ and a two-particle operator $\hat{\mathcal{O}}_2(\mathbf{r}_i,\mathbf{r}_j)$ we have

$$\boxed{\begin{aligned} \hat{\mathcal{O}}_1\psi &= \sum_{i=1}^{N}\hat{\mathcal{O}}_1(\mathbf{r}_i)\psi(\mathbf{r}_1,\mathbf{r}_2,\dots,\mathbf{r}_N), \\ \hat{\mathcal{O}}_2\psi &= \sum_{i>j}^{N}\hat{\mathcal{O}}_2(\mathbf{r}_i,\mathbf{r}_j)\psi(\mathbf{r}_1,\mathbf{r}_2,\dots,\mathbf{r}_N). \end{aligned}} \tag{2.60}$$

For fermions, when $\psi_{\nu_1\dots\nu_N}(\mathbf{r}_1,\dots,\mathbf{r}_N) = \hat{S}_-[\psi_{\nu_1}(\mathbf{r}_1)\psi_{\nu_2}(\mathbf{r}_2)\dots\psi_{\nu_N}(\mathbf{r}_N)]$ is a simple anti-symmetrized product of single-particle orbitals, the single-particle operator acts evenly on

all these single-particle orbitals, and we end up with the simple result

$$\langle \mathcal{O}_1 \rangle_\psi = \sum_{\nu_j} \langle \mathcal{O}_1 \rangle_{\psi_{\nu_j}} \tag{2.61a}$$

$$= \sum_{\nu_j} \int d\mathbf{r}\, \psi^*_{\nu_j}(\mathbf{r})\, \mathcal{O}_1 \psi_{\nu_j}(\mathbf{r}). \tag{2.61b}$$

This is simply the sum of all the single-particle contributions.

Similarly for the two-particle operator $\mathcal{O}_2$, which also for the same ψ acts evenly on all two-particle states $\psi_{\nu_j\nu_k}$ (the prefactor $\frac{1}{2}$ takes care of double counting)

$$\langle \mathcal{O}_2 \rangle_\psi = + \frac{1}{2} \sum_{\nu_j\nu_k} \langle \mathcal{O}_2 \rangle_{\psi_{\nu_j\nu_k}} \tag{2.62a}$$

$$= + \frac{1}{2} \sum_{\nu_j\nu_k} \int d\mathbf{r}_1 \int d\mathbf{r}_2\, \psi^*_{\nu_j}(\mathbf{r}_1)\psi^*_{\nu_k}(\mathbf{r}_2)\, \mathcal{O}_2 \psi_{\nu_j}(\mathbf{r}_1)\psi_{\nu_k}(\mathbf{r}_2) \tag{2.62b}$$

$$- \frac{1}{2} \sum_{\nu_j\nu_k} \int d\mathbf{r}_1 \int d\mathbf{r}_2\, \psi^*_{\nu_j}(\mathbf{r}_1)\psi^*_{\nu_k}(\mathbf{r}_2)\, \mathcal{O}_2 \psi_{\nu_j}(\mathbf{r}_2)\psi_{\nu_k}(\mathbf{r}_1). \tag{2.62c}$$

Here we have used the explicit form Eq. (2.52) for the two-fermion state $\psi_{\nu_j\nu_k}(\mathbf{r}_1, \mathbf{r}_2)$. Note that the term where the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ has been exchanged appears with a minus sign. This particular quantum expectation value has no classical analoge. It is intimately tied to the indistinguishability of quantum particles. This contribution to the expectation value is denoted the exchange term.

Chapter 3

Metals and conduction electrons

The study of electrons moving in a charge compensating background of positively charged ions is central in the understanding of solids. We shall restrict ourselves to the study of simple metals, but the results obtained are important for understanding of the scanning tunnel microscope (STM) and the atomic force microscope (AFM).

Any atom in a metal consists of three parts: the positively charged heavy nucleus at the center, the light cloud of the many negatively charged core electrons tightly bound to the nucleus, and finally, the outermost few valence electrons. The nucleus with its core electrons is called the ion. The ion mass is denoted M, and if the atom has Z valence electrons the charge of the ion is $+Ze$. To a large extend the inner degrees of freedom of the ions do not play a significant role leaving the center of mass coordinates R_j and total spin S_j of the ions as the only dynamical variables. In contrast to the core electrons the Z valence electrons, with mass m and charge $-e$, are often free to move away from their respective host atoms. Thus a gas of electrons swirling around among the ions is formed. In metals the electrons constituting this gas are able to conduct electric currents, and consequently they are denoted conduction electrons. The formation of a metal from N independent atoms is sketched in Fig. 3.1.

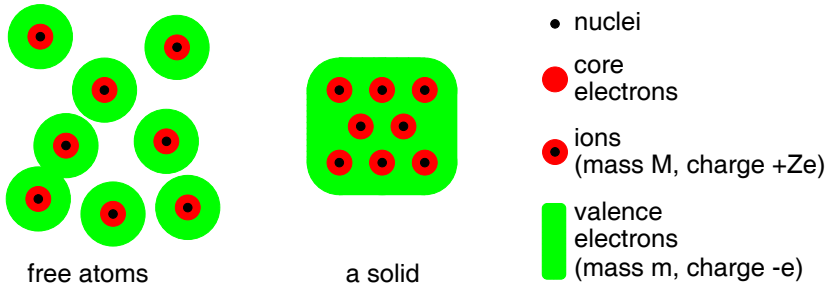


Figure 3.1: A sketch showing N free atoms merging into a metal. The ions are unchanged during the process where they end up by forming a periodic lattice. The valence electrons are freed from their host atoms and form an electron gas holding the ionic lattice together.

3.1 The single-electron states: travelling waves

In our study of conduction electrons in metals we employ the particularly simple Sommerfeld model. In this model the ion charges are imagined to be smeared out to form a homogeneous and static positive charge density, $+Z\rho_{\text{jel}}$, called the ion jellium. Let the metal be shaped as a box with side lengths L_x, L_y, and L_z and volume $\mathcal{V} = L_x L_y L_z$. The periodic potential, $V_{\text{el}-\text{latt}}$, present in a real lattice becomes the constant potential $V_{\text{el}-\text{jel}}$ confined by steep walls to the region $0 < x < L_x$, $0 < y < L_y$, and $0 < z < L_z$ as sketched in Fig. 3.2. We are free to choose the zero point of the energy, so it is natural to choose it such that the potential is zero, $V_{\text{el}-\text{jel}} = 0$, inside the box. In this case the Schrödinger equation Eq. (3.1) simply becomes

$$-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi(\mathbf{r}) = E\psi(\mathbf{r}). \tag{3.1}$$

It is straightforward to check that any plane wave with wavevector $\mathbf{k}$,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{V}}}\, e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{\sqrt{\mathcal{V}}}\, e^{i(k_x x + k_y y + k_z z)}, \tag{3.2}$$

solves the Schrödinger equation with a $\mathbf{k}$-dependent eigenenergy $\varepsilon_{\mathbf{k}}$:

$$-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2\psi_{\mathbf{k}}(\mathbf{r}) = \varepsilon_k\psi_{\mathbf{k}}(\mathbf{r}) \tag{3.3a}$$

$$\varepsilon_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_z^2\right). \tag{3.3b}$$

We note that the wavefunctions are properly normalized since $\int d\mathbf{r}\, |\psi(\mathbf{r})|^2 = \int d\mathbf{r}\, \frac{1}{\mathcal{V}} = 1$.

But what about boundary conditions? And is any $\mathbf{k}$ an admissable solution? At first, it seems natural to demand the vanishing of the wave function at the walls, $\psi(0, y, z) = 0$ and $\psi(L_x, y, z) = 0$ (likewise for the y and z directions). This leads to the standing wave solutions

$$\tilde{\psi}(\mathbf{r}) = C \sin\left(n_x\pi\frac{x}{L_x}\right)\sin\left(n_y\pi\frac{y}{L_y}\right)\sin\left(n_z\pi\frac{z}{L_z}\right), \quad n_x, n_y, n_z = 1, 2, 3, 4, \ldots \tag{3.4}$$

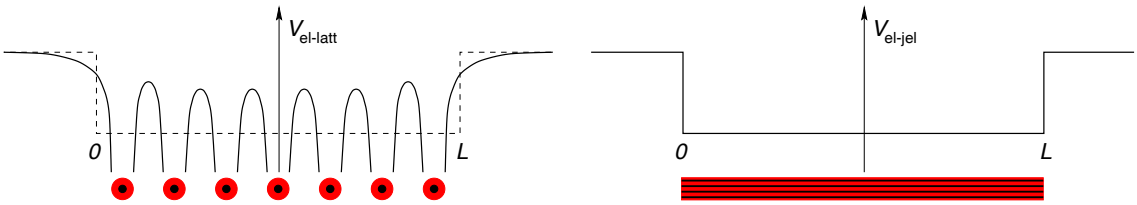


Figure 3.2: A sketch showing the periodic potential, $V_{\text{el}-\text{latt}}$, present in a real lattice, and the imagined smeared out potential $V_{\text{el}-\text{jel}}$ of the jellium model.

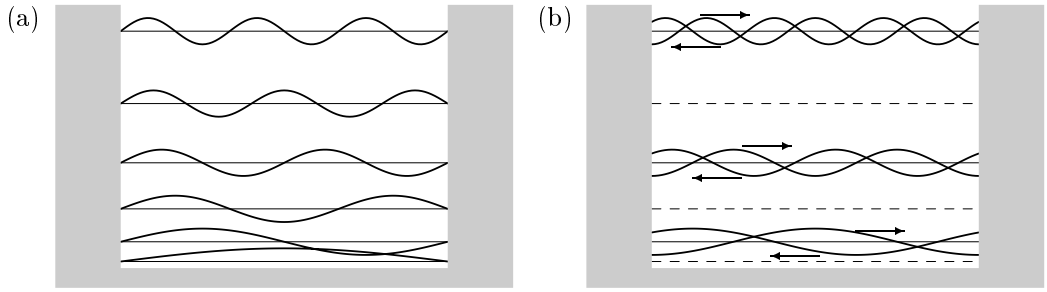


Figure 3.3: (a) Standing waves in a box obeying the boundary conditions $\psi(0) = 0$ and $\psi(L) = 0$. There is one standing wave at each allowed energy. (b) Travelling waves obeying periodic boundary conditions $\psi(0) = \psi(L)$ and $\psi'(0) = \psi'(L)$. There are only half as many allowed energy levels in (b) as compared to (a), but there are two waves at each allowed energy, one moving to the left the other to the right.

However, although these states are energy eigestates, they are not eigenstates for the momentum operator $\hat{\mathbf{p}}$,

$$\frac{\hbar}{i}\boldsymbol{\nabla}\tilde{\psi}(\mathbf{r}) \neq \hbar\mathbf{k}\,\tilde{\psi}(\mathbf{r}). \tag{3.5}$$

Instead we therefore choose the travelling wave solutions that arises from imposing periodic boundary conditions $\psi(L, y, z) = \psi(0, y, z)$ and $\psi'(L, y, z) = \psi'(0, y, z)$ (likewise for the y and z directions).

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{V}}}\, e^{i\mathbf{k}\cdot\mathbf{r}}, \qquad \begin{cases} k_x = \frac{2\pi}{L_x} n_x \text{ (same for } y \text{ and } z) \\ n_x = 0, \pm 1, \pm 2, \ldots \text{ (same for } y \text{ and } z) \\ \mathcal{V} \;= L_x L_y L_z, \end{cases} \tag{3.6}$$

The travelling waves are eigenstates both to the energy operator $\hat{H}$ and to the momentum operator $\hat{\mathbf{p}}$,

$$\hat{H}\psi_{\mathbf{k}}(\mathbf{r}) = \frac{\hbar^2 k^2}{2m}\,\psi_{\mathbf{k}}(\mathbf{r}) = \frac{h^2}{2m}\,\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right)\psi_{\mathbf{k}}(\mathbf{r}) \tag{3.7a}$$

$$\hat{\mathbf{p}}\psi_{\mathbf{k}}(\mathbf{r}) = \frac{\hbar}{i}\boldsymbol{\nabla}\psi_{\mathbf{k}}(\mathbf{r}) = \hbar\mathbf{k}\,\psi_{\mathbf{k}}(\mathbf{r}). \tag{3.7b}$$

3.2 The ground state for non-interacting electrons

Having identified the single-electron states we are now in a position to find the ground state for N non-interacting electrons. In this first analysis we neglect the Coulomb interaction between the electrons. We shall see later why this is justifiable.

For each allowed wavevector $\mathbf{k}$, see Eq. (3.6), there exists an electron state $\psi_{\mathbf{k}}$, that according to the Pauli exclusion principle can hold two electrons, one with spin up and

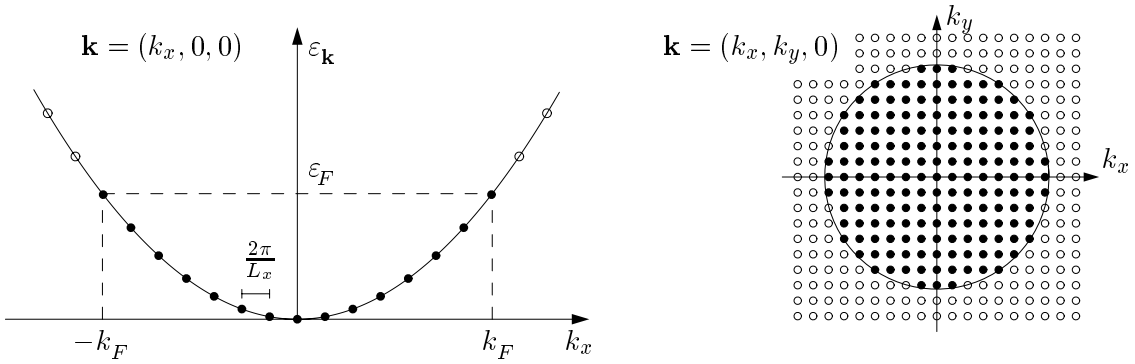


Figure 3.4: Two aspects of the Fermi sphere in **k**-space. To the left the dispersion relation $\varepsilon_{\mathbf{k}}$ is plotted along the line $\mathbf{k} = (k_x, 0, 0)$, and ε_{F} and k_{F} are indicated. To the right the occupation of the states is shown in the plane $\mathbf{k} = (k_x, k_y, 0)$. The Fermi sphere is shown as a circle with radius k_{F}. Filled and empty circles represent occupied and unoccupied states, respectively.

one with spin down. The energy of an electron in this state is $\varepsilon_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m}$. It is natural to order the single-particle states according to their energies in ascending order,

$$\psi_{\mathbf{k}_1\uparrow}, \psi_{\mathbf{k}_1\downarrow}, \psi_{\mathbf{k}_2\uparrow}, \psi_{\mathbf{k}_2\downarrow}, \ldots, \quad \text{where } \varepsilon_{\mathbf{k}_1} \leq \varepsilon_{\mathbf{k}_2} \leq \varepsilon_{\mathbf{k}_3} \leq \ldots \tag{3.8}$$

where the arrows indicate the electron spin in a given state. In the following we suppress this spin index and simply take spin into account by allowing two electrons (one with spin up and the other with spin down) in each state $\psi_{\mathbf{k}}(\mathbf{r})$.

The ground state, i.e. the state with lowest energy, of a system containing N electrons is therefore constructed by filling up states with the lowest possible energy, i.e. the smallest values of k, until all N electrons are placed. If we represent the allowed **k**-vectors as points in a (k_x, k_y, k_z)-coordinate system, we easily see that the ground state is obtained by occupying with two electrons (spin up and spin down) all states inside the smallest sphere centered around the zero-energy state $\mathbf{k} = 0$ containing $N/2$ allowed **k**-points, see Fig. 3.4.

This sphere in **k**-space representing the ground state of the non-interacting N-electron system is called the Fermi sphere. Due to the Pauli principle it is not possible to move electrons around deep inside the Fermi sphere, so when an N-electron system is disturbed by external perturbations most of the action involves occupied and un-occupied states near the surface. It is therefore relevant to focus on the surface. The radius of the Fermi sphere is denoted the Fermi wavenumber k_{F}. The energy of the topmost occupied state is denoted the Fermi energy, ε_{F}. Associated with ε_{F} and the Fermi wavenumber k_{F} are the Fermi wave length λ_{F} and the Fermi velocity v_{F}:

$$\varepsilon_{\mathrm{F}} = \frac{\hbar^2 k_{\mathrm{F}}^2}{2m}, \qquad k_{\mathrm{F}} = \frac{1}{\hbar}\sqrt{2m\varepsilon_{\mathrm{F}}}, \qquad \lambda_{\mathrm{F}} = \frac{2\pi}{k_{\mathrm{F}}}, \qquad v_{\mathrm{F}} = \frac{\hbar k_{\mathrm{F}}}{m}. \tag{3.9}$$

To facilitate calculations involving the Fermi sphere we note how the quantization of **k** in Eq. (3.6) means that one state fills a volume $\frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{(2\pi)^3}{\mathcal{V}}$ in **k**-space. From this

follows the very useful rule of how to evaluate **k**-sums by **k**-integrals:

$$\sum_{\mathbf{k}} \quad \rightarrow \quad \frac{\mathcal{V}}{(2\pi)^3}\int d\mathbf{k}\,. \tag{3.10}$$

We shall immediatly use this rule to calculate the relation between the macroscopic quantity $n = N/\mathcal{V}$, the density, and the microscopic quantity $k_{\rm F}$, the Fermi wavenumber.

$$N = \sum_{\text{allowed } \mathbf{k}}^{k_{\rm F}} 2 \; = 2\,\frac{\mathcal{V}}{(2\pi)^3}\int_{|\mathbf{k}|<k_{\rm F}} d\mathbf{k} = 2\,\frac{\mathcal{V}}{(2\pi)^3}\int_0^{k_{\rm F}} 4\pi k^2\,dk = \frac{\mathcal{V}}{3\pi^2}\,k_{\rm F}^3, \tag{3.11}$$

where we have inserted a factor of 2 for spin. We arrive at

$$k_{\rm F}^3 = 3\pi^2 n. \tag{3.12}$$

This extremely important formula allows us to obtain the values of the microscopic parameters $k_{\rm F}$, $\varepsilon_{\rm F}$, and $v_{\rm F}$. Hall measurements in macroscopic samples yield the electron density of copper[1], $n = 8.47\times10^{28}$ m^{-3}, and from Eqs. (3.9) and (3.12) it thus follows that for copper

$$\begin{aligned} k_{\rm F} &= 13.6 \text{ nm}^{-1} \qquad & \varepsilon_{\rm F} &= 7.03 \text{ eV} = 81600 \text{ K} \\ \lambda_{\rm F} &= 0.46 \text{ nm} & v_{\rm F} &= 1.57\times10^6 \text{ m/s} = 0.005\,c. \end{aligned} \tag{3.13}$$

Note that the Fermi energy corresponds to an extremely high temperature, which we shall return to shortly, and even though the Fermi velocity is large it is still less than a percent of the velocity of light, and we need not invoke relativistic considerations.

3.3 The energy of the non-interacting electron gas

In this section we calculate the ground state energy of the electron gas in the limit of high electron densities. The reason for choosing this limit is, as we shall see, that the Fermi sphere is an excellent approximation to the ground state even if the Coulomb interaction between the electrons is taken into account. The question is why this is so. The key to the answer lies in how the kinetic energy $E_{\rm kin}$ and the potential Coulomb energy $E_{\rm pot}$ of a typical electron depend on the electron density n. From Eqs. (3.9) and (3.12) we have $E_{\rm kin} \approx \varepsilon_{\rm F} \propto k_{\rm F}^2 \propto n^{\frac{2}{3}}$. At the density n the typical distance $\bar{d}$ between the electrons is proportional to $n^{-\frac{1}{3}}$, and consequently the Coulomb energy behaves as $E_{\rm pot} \propto \bar{d}^{-1} \propto n^{\frac{1}{3}}$. So we find that

$$\frac{E_{\rm pot}}{E_{\rm kin}} \propto \frac{n^{\frac{1}{3}}}{n^{\frac{2}{3}}} = n^{-\frac{1}{3}} \quad \underset{n\to\infty}{\longrightarrow} \quad 0, \tag{3.14}$$

revealing the following perhaps somewhat counter intuitive fact: the importance of the electron-electron interaction diminishes as the density of the electron gas increases. Due

[1]The density can also be estimated as follows. The inter-atomic distances are typically $\simeq 2$ Å. In monovalent Cu one electron thus occupies a volume $\simeq (2\times10^{-10}\text{ m})^3$, and $n \approx 10^{29}$ m^{-3} follows.

to the Pauli exclusion principle the kinetic energy simply becomes the dominant energy scale in the interacting electron gas at high densities.

We move on to calculate the ground state energy for the non-interacting electron gas. Due to the absense of interaction this is purely kinetic energy:

$$E_{\text{kin}} = \sum_{\text{allowed } \mathbf{k}}^{k_{\text{F}}} 2\,\varepsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}\,\frac{2\mathcal{V}}{(2\pi)^3}\int_{|\mathbf{k}|<k_{\text{F}}} d\mathbf{k}\,k^2 = \frac{\mathcal{V}\hbar^2}{m(2\pi)^3}\int_0^{k_{\text{F}}} 4\pi k^4\,dk = \frac{3}{5}N\varepsilon_{\text{F}}. \tag{3.15}$$

The kinetic energy per particle as a function of the density n is therefore

$$\frac{E_{\text{kin}}}{N} = \frac{3}{5}\frac{\hbar^2}{2m}{k_{\text{F}}}^2 = \frac{3}{5}\frac{\hbar^2}{2m}(3\pi^2)^{\frac{2}{3}}\,n^{\frac{2}{3}}. \tag{3.16}$$

3.4 The energy of the interacting electron gas

The Coulomb interaction is now taken into account. First we note that the sum of classical electrostatic energies from ion-ion, ion-electron, and electron-electron interactions is zero. This is because the both the electrons and the ions are smeared out homogeneously leaving a system that is completely charge neutral everywhere in space. Any non-zero contribution to the Coulomb energy must therefore be of quantum nature.

In quantum theory the Coulomb interaction is given by a two-particle operator $\hat{V}_{\text{Coul}}$, since two particles are involved each with its own coordinates,

$$\hat{V}_{\text{Coul}}(\mathbf{r}_1,\mathbf{r}_2) = \frac{e^2}{4\pi\epsilon_0}\,\frac{1}{\mathbf{r}_2-\mathbf{r}_1}. \tag{3.17}$$

We disregard the spin index here. The interesting quantum effects in the following comes only from electron pairs with the same spin. The potential energy E_{pot} is just the expectation value of of the Coulomb interaction operator $\hat{V}_{\text{Coul}}$ acting on the N-particle state ψ, the Fermi sphere. Using the results of Sec. 2.8.3 in particular the equation Eq. (2.62a) valid for two-particle operators we get

$$\begin{aligned}
E_{\text{pot}} = &\quad \langle\hat{V}_{\text{Coul}}\rangle_{\psi} && \text{(3.18a)}\\
= &+ \frac{1}{2}\sum_{\mathbf{k}_1\mathbf{k}_2}\langle\hat{V}_{\text{Coul}}\rangle_{\psi_{\mathbf{k}_1\mathbf{k}_2}} && \text{(3.18b)}\\
= &+ \frac{e^2}{8\pi\epsilon_0}\sum_{\mathbf{k}_1\mathbf{k}_2}\int d\mathbf{r}_1\int d\mathbf{r}_2\,\frac{|\psi_{\mathbf{k}_1}(\mathbf{r}_1)|^2|\psi_{\mathbf{k}_2}(\mathbf{r}_2)|^2}{|\mathbf{r}_2-\mathbf{r}_1|} && \text{(3.18c)}\\
&- \frac{e^2}{8\pi\epsilon_0}\sum_{\mathbf{k}_1\mathbf{k}_2}\int d\mathbf{r}_1\int d\mathbf{r}_2\,\frac{\psi^*_{\mathbf{k}_1}(\mathbf{r}_1)\psi^*_{\mathbf{k}_2}(\mathbf{r}_2)\psi_{\mathbf{k}_1}(\mathbf{r}_2)\psi_{\mathbf{k}_2}(\mathbf{r}_1)}{|\mathbf{r}_2-\mathbf{r}_1|}. && \text{(3.18d)}
\end{aligned}$$

The term Eq. (3.18c) is easy to interpret. It is simply the classical electrostatic energy from two electron charge distributions $|\psi_{\mathbf{k}_1}(\mathbf{r}_1)|^2$ and $|\psi_{\mathbf{k}_2}(\mathbf{r}_2)|^2$. We have already discussed how this is cancelled by the interaction with the postive ions. However, the term Eq. (3.18d)

is not a classical electrostatic energy, and it will therefore yield a non-zero contribution. We note that it involves exchange of coordinates between the two single-particle orbitals $\psi_{\mathbf{k}_1}$ and $\psi_{\mathbf{k}_2}$ typical of quantum states with two (or more) indistingushiable particles. The term is therefore called the exchange energy, denoted E_{exch}. It is not difficult to calculate the integral when $\psi_{\mathbf{k}}$ are travelling waves, however, it is somewhat cumbersome. We therefore just state the result:

$$E_{\text{exch}} = \frac{-e^2}{4\pi\epsilon_0 \mathcal{V}^2} \sum_{\mathbf{k}_1\mathbf{k}_2}^{k_{\text{F}}} \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, \frac{e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot(\mathbf{r}_2-\mathbf{r}_1)}}{|\mathbf{r}_2-\mathbf{r}_1|} = -\frac{3N}{4\pi} \, \frac{e^2}{4\pi\epsilon_0} k_{\text{F}}. \tag{3.19}$$

Or expressing k_{F} in terms of density:

$$\frac{E_{\text{exch}}}{N} = -\frac{3(3\pi^2)^{\frac{1}{3}}}{4\pi} \, \frac{e^2}{4\pi\epsilon_0} n^{\frac{1}{3}}. \tag{3.20}$$

We note that the exchange energy E_{exch} is negative. This can be explained by noting that the Pauli exclusion principle prohibits electrons to be close one another. This quantum effect therefore forces the electrons to stay further apart on average than would have been the case in a purely classical world. But larger distance means smaller Coulomb energy, and the quantum correction to the classical Coulomb energy must therefore be negative.

The total energy per electron E/N in the metal is now obtained by combining the quantum kinetic energy Eq. (3.16) with the quantum exchange energy Eq. (3.20):

$$\frac{E}{N} = \frac{E_{\text{kin}} + E_{\text{exch}}}{N} = \frac{3}{5}\frac{\hbar^2}{2m}(3\pi^2)^{\frac{2}{3}} \, n^{\frac{2}{3}} - \frac{3(3\pi^2)^{\frac{1}{3}}}{4\pi} \, \frac{e^2}{4\pi\epsilon_0} n^{\frac{1}{3}}. \tag{3.21}$$

It is common practice to express the result in terms of a dimensionless measure r_s of the average distance between the electrons and the Bohr radius $a_0 = 0.053$ nm. By definition $r_s a_0$ is the radius of a sphere containing exactly one electron,

$$\frac{4\pi}{3}(r_s a_0)^3 = \frac{1}{n} \quad \Rightarrow \quad r_s = \left(\frac{3}{4\pi}\right)^{\frac{1}{3}} \frac{1}{n^{\frac{1}{3}}} \frac{1}{a_0}. \tag{3.22}$$

Putting in numbers we obtain for the total energy per particle:

$$\frac{E}{N} = \left(\frac{2.211}{r_s^2} - \frac{0.916}{r_s}\right) 13.6 \text{ eV}. \tag{3.23}$$

This result shows that the electron gas is stable when the repulsive Coulomb interaction is turned on. No external confinement potential is needed to hold the electron gas in the ion jellium together. There exists an optimal density n^*, or inter-particle distance r_s^*, which minimizes the energy and furthermore yields an energy $E^* < 0$. The negative exchange energy overcomes the positive kinetic energy. The equilibrium situation is obtained from

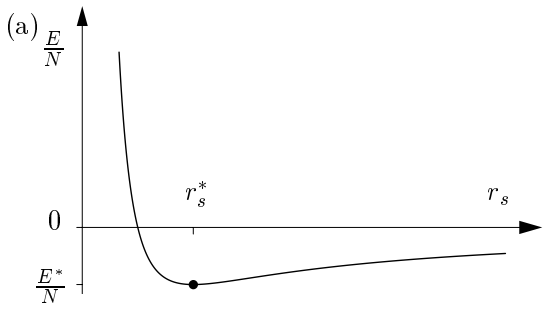


Figure 3.5: (a) The energy per particle E/N of the 3D electron gas in the jellium model Eq. (3.23) as a function of the dimensionless inter-particle distance r_s. Due to the quantum kinetic energy and the quantum exchange energy the electron gas is stable at $r_s = r_s^* = 4.83$ with an ionization energy $E/N = E^*/N = -1.29$ eV.

$\frac{\partial}{\partial r_s}(E^{(0)} + E^{(1)}) = 0$, and we can compare the result with experiment:

$$r_s^* = 4.83, \quad \frac{E^*}{N} = -1.29 \text{ eV} \quad \text{(the jellium model)} \tag{3.24a}$$

$$r_s = 3.96, \quad \frac{E}{N} = -1.13 \text{ eV} \quad \text{(experiment on Na)} \tag{3.24b}$$

It is gratifying that our crude jellium model is able to grasp some central features in the physics of metals. It shows how kinetic and potential energies of quantum nature keep the metal together, and it predicts an almost correct electron density and ionization energy.

3.5 The density of states

The concept of density of states, $D(\varepsilon) = \frac{dN}{d\varepsilon}$, is widely used. The function $D(\varepsilon)$ simply gives the number ΔN of states in the energy interval $\Delta\varepsilon$ around the energy ε, $\Delta N = D(\varepsilon)\Delta\varepsilon$, and the density of states per volume $d(\varepsilon) = D(\varepsilon)/\mathcal{V} = \frac{dn}{d\varepsilon}$. Again using Eq. (3.12) we find

$$\varepsilon_{\rm F} = \frac{\hbar^2}{2m}{k_{\rm F}}^2 = \frac{\hbar^2}{2m}(3\pi^2)^{\frac{2}{3}}\, n^{\frac{2}{3}} \quad \Rightarrow \quad n(\varepsilon) = \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \varepsilon^{\frac{3}{2}}, \text{ for } \varepsilon > 0, \tag{3.25}$$

and from this

$$d(\varepsilon) = \frac{dn}{d\varepsilon} = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \varepsilon^{\frac{1}{2}}\,\theta(\varepsilon), \qquad D(\varepsilon) = \frac{dN}{d\varepsilon} = \frac{\mathcal{V}}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \varepsilon^{\frac{1}{2}}\,\theta(\varepsilon). \tag{3.26}$$

Here we have introduced the much used Heavyside step function $\theta(\varepsilon)$

$$\theta(\varepsilon) = \begin{cases} 1, & \text{for } \varepsilon > 0, \\ 0, & \text{for } \varepsilon < 0. \end{cases} \tag{3.27}$$

The density of states $D(\varepsilon)$ is a very useful function. It can forexample be used to calculate the particle number, $N = \int d\varepsilon\, D(\varepsilon)$, and the total energy, $E^{(0)} = \int d\varepsilon\, \varepsilon\, D(\varepsilon)$.

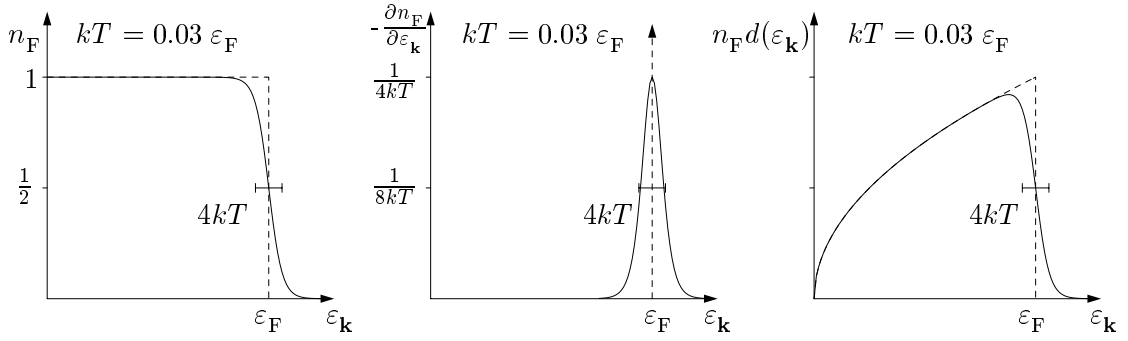


Figure 3.6: The Fermi-Dirac distribution $n_{\rm F}(\varepsilon_{\mathbf{k}})$, its derivative $-\frac{\partial n_{\rm F}}{\partial \varepsilon_{\mathbf{k}}}$, and its product with the density of states, $n_{\rm F}(\varepsilon_{\mathbf{k}})d(\varepsilon_{\mathbf{k}})$, shown at the temperature $kT = 0.03\,\varepsilon_{\rm F}$, corresponding to $T = 2400$ K in metals. This rather high value is chosen to have a clearly observable deviation from the $T = 0$ case, which is indicated by the dashed lines.

3.6 The electron gas at finite temperature

In the previous section temperature was tacitly taken to be zero. As temperature is raised from zero the occupation number is given by the Fermi-Dirac distribution $n_{\rm F}(\varepsilon_{\mathbf{k}})$, see Eq. (2.55). The main characteristics of this function is shown in Fig. 3.6. Note that to be able to see any effects of the temperature in Fig. 3.6, kT is set to $0.03\,\varepsilon_{\rm F}$ corresponding to $T \approx 2400$ K. Room temperature yields $kT/\varepsilon_{\rm F} \approx 0.003$, thus the low temperature limit of $n_{\rm F}(\varepsilon_{\mathbf{k}})$ is of importance:

$$n_{\rm F}(\varepsilon_{\mathbf{k}}) = \frac{1}{e^{\beta(\varepsilon_{\mathbf{k}}-\mu)}+1} \quad \underset{T\to 0}{\longrightarrow} \quad \theta(\mu - \varepsilon_{\mathbf{k}}), \tag{3.28a}$$

$$-\frac{\partial n_{\rm F}}{\partial \varepsilon_{\mathbf{k}}} = \frac{\beta}{4}\frac{1}{\cosh^2[\frac{\beta}{2}(\varepsilon_{\mathbf{k}}-\mu)]} \quad \underset{T\to 0}{\longrightarrow} \quad \delta(\mu - \varepsilon_{\mathbf{k}}). \tag{3.28b}$$

At $T = 0$ the chemical potential μ is identical to $\varepsilon_{\rm F}$. But in fact μ varies slightly with temperature. A careful analysis based on the so-called Sommerfeld expansion combined with the fact that the number of electrons does not change with temperature yields

$$n(T=0) = n(T) = \int_0^\infty d\varepsilon\, d(\varepsilon) f(\varepsilon) \quad \Rightarrow \quad \mu(T) = \varepsilon_{\rm F}\left[1 - \frac{\pi^2}{12}\Big(\frac{kT}{\varepsilon_{\rm F}}\Big)^2 + \ldots\right] \tag{3.29}$$

Because $\varepsilon_{\rm F}$ according to Eq. (3.13) is around 80000 K for metals, we find that even at the melting temperature of metals only a very limited number ΔN of electrons are affected by thermal fluctuations. Indeed, only the states within $2kT$ of $\varepsilon_{\rm F}$ are actually affected, and more precisely we have $\Delta N/N = 6kT/\varepsilon_{\rm F}$ ($\approx 10^{-3}$ at room temperature). The Fermi sphere is not destroyed by heating, it is only slightly smeared. Now we have at hand an explanation of the old paradox in thermodynamics, as to why only the ionic vibrational degrees of freedom contribute significantly to the specific heat of solids. The electronic

degrees of freedom are simply 'frozen' in. Only at temperatures comparable to ε_{F} they begin to play a major role.

Chapter 4

Atomic orbitals and carbon nanotubes

In this chapter we shall study atomic orbitals. These are the basis for the formation of chemical bonds, i.e., for the formation of molecules and solids. Thanks to the modern nanotechnological tools the atomic orbitals are no longer merely a mathematical abstraction, but something very real that can be observed directly. We begin with a short introduction to hydrogen-like atoms, and then move on to carbon atoms, graphene sheets, and finally the carbon nanotube molecules.

4.1 The Schrödinger equation for hydrogen-like atoms

In the following we analyze one-electron hydrogen-like atoms like hydrogen itself, H, the helium-I ion, He^+, and the lithium-II ion, Li^{++}, and our starting point is therefore the Schrödinger equation for an electron moving in the Coulomb potential from a heavy nucleus at rest with the positive charge $+Ze$.

We consider an infinitely heavy nucleus[1] with charge $+Ze$ placed in the center of our coordinate system. The electrostatic potential energy $V(\mathbf{r})$ for an electron (with charge $-e$ and mass m_e) moving around such a nuclus is given by

$$V(\mathbf{r}) = -\frac{\alpha}{r}, \qquad \text{with } \alpha = \frac{Ze^2}{4\pi\epsilon_0}. \tag{4.1}$$

This potential energy depends only on the distance r between the electron and the nucleus and not on any angles. It is therefore natural to work with spherical coordinates (r, θ, ϕ) instead of the usual cartesian coordinates (x, y, z). These two sets of coordinates are

[1]The proton and neutron are both roughly 1836 times heavier than the electron, so in the infinite mass approximation we make an error of the order 0.2 %. Correct results are obtained by substituing the electron mass m_e with its so-called reduced mass $\mu = \frac{m_e M}{m_e+M} = \frac{m_e}{1+m_e/M}$, where M is the mass of the nucleus.

related by

$$\begin{aligned} x &= r\ \sin\theta\ \cos\phi, \\ y &= r\ \sin\theta\ \sin\phi, \\ z &= r\ \cos\theta. \end{aligned} \tag{4.2}$$

Substituting as usual the classical momentum $\mathbf{p}$ with the differential operator $\frac{\hbar}{i}\boldsymbol{\nabla}$, the classical energy equation for the electron $E = \frac{p^2}{2m_e} - \frac{\alpha}{r}$ becomes the time-independent Schrödinger equation

$$\left[-\frac{\hbar^2}{2m_e}\,\boldsymbol{\nabla}^2 - \frac{\alpha}{r}\right]\psi_{_E}(\mathbf{r}) = E\psi_{_E}(\mathbf{r}). \tag{4.3}$$

In spherical coordinates $\boldsymbol{\nabla}^2$ is given by Eq. (2.28b), and the Schrödinger equation takes the following form

$$\left[-\frac{\hbar^2}{2m_e}\left\{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\} - \frac{\alpha}{r}\right]\psi_{_E}(\mathbf{r}) = E\psi_{_E}(\mathbf{r}). \tag{4.4}$$

It seems to be a very difficult affair to find the eigenstates $\psi_{_E}(\mathbf{r})$ that solve this equation. But the spherical coordinates make things easier. Fortunately, the complicated differential operator acting on $\psi_{_E}(\mathbf{r})$ on the left-hand side of the equation has very distinct parts acting on each of the coordinates r, θ, and ϕ, respectively. We therefore separate the wavefunction as a product of functions of each of the three coordinates,

$$\psi_{_E}(\mathbf{r}) = R(r)\,\Theta(\theta)\,\Phi(\phi), \tag{4.5}$$

and find simple solutions for each of these functions.

4.1.1 The azimuthal functions $\Phi_{_m}(\phi)$

We begin with the azimuthal angle ϕ. The only place where the differential oprator in Eq. (4.4) acts on ϕ is the term containing $\frac{\partial^2}{\partial\phi^2}$. Hence we seek a function $\Phi_{_m}(\phi)$ that gives a simple result when acted upon by this differential operator:

$$\frac{\partial^2\Phi_{_m}(\phi)}{\partial\phi^2} = -m^2\Phi_{_m}(\phi), \tag{4.6}$$

where m is a constant (maybe complex) to be determined. We see immediatly that $\Phi_{_m}(\phi) = \exp(im\phi)$ solves Eq. (4.6), but what about the boundary condition? Naturally, we would like the wavefunction to be single-valued, so upon changing the azimuthal angle by 2π when should obtain the same value, i.e. $\Phi_{_m}(\phi + 2\pi) = \Phi_{_m}(\phi)$. But with $\Phi_{_m}(\phi) = \exp(im\phi)$ this is only possible for integer vaules of m. We therefore obtain the solution

$$\Phi_{_m}(\phi) = \frac{1}{\sqrt{2\pi}}e^{im\phi}, \quad m = 0, \pm1, \pm2, \ldots \tag{4.7}$$

The prefactor ensures proper normalization: $\int_0^{2\pi} d\phi\,|\Phi_{_m}(\phi)|^2 = 1$.

4.1.2 The polar functions $\Theta_{lm}(\theta)$

We now choose one particular integer value of m and keep it fixed. When this is done $\frac{\partial^2}{\partial\phi^2}$ simply gets replaced by the number $-m^2$. We then focus on the θ-dependent part of the Schrödinger equation and seek functions $\Theta_{lm}(\theta)$ that gives a simple result when the θ-part of the differential operator acts on it:

$$\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{m^2}{\sin^2\theta}\right]\Theta_{lm}(\theta)=-l(l+1)\,\Theta_{lm}(\theta). \tag{4.8}$$

Apriori, the constant $l(l+1)$ can be any complex number, but anticipating the solution we have written it as seen. The solution of Eq. (4.8) is found by changing variables from θ to $\mu\equiv\cos\theta$, whereby the equation for $\Theta_{lm}(\theta)=P_l^m(\mu)$ becomes

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{dP_l^m(\mu)}{d\mu}\right]-\frac{m^2}{1-\mu^2}P_l^m(\mu)=-l(l+1)P_l^m(\mu),\qquad \mu=\cos\theta. \tag{4.9}$$

This is one of the classical differential equation of mathematical analysis. Given the boundary condition $-1\le\mu\le 1$ the solutions are known to be the so-called associated Legendre polynomials $P_l^m(\mu)$, where l is a non-negative integer, and where $l\ge|m|\ge 0$:

$$P_l^m(\mu)=\frac{(-1)^m}{2^l l!}\,(1-\mu^2)^{\frac{m}{2}}\,\frac{d^{l+m}}{d\mu^{l+m}}(\mu^2-1)^l,\quad\left\{\begin{array}{ll} l & =0,1,2,\ldots\\ m & =-l,\ldots,l.\end{array}\right. \tag{4.10}$$

With proper normalization, $\int_{-1}^{1}d(\cos\theta)\,|\Theta_{lm}(\theta)|^2=1$, we arrive at

$$\Theta_{lm}(\theta)=\sqrt{\frac{(l+\frac{1}{2})(l-m)!}{(l+m)!}}\,P_l^m(\cos\theta),\quad\left\{\begin{array}{ll} l & =0,1,2,\ldots\\ m & =-l,-(l-1),\ldots,(l-1),l.\end{array}\right. \tag{4.11}$$

We give some explicit expressions for this wavefunction in the next subsection.

4.1.3 The spherical harmonics $Y_l^m(\theta,\phi)=\Theta_{lm}(\theta)\,\Phi_m(\phi)$

We have now solved the angular part of the problem. For reasons to be discussed in Sec. 4.4 it is costumary to combine the two angular functions $\Theta_{lm}(\theta)$ and $\Phi_m(\phi)$ into one, the so-called spherical harmonic $Y_l^m(\theta,\phi)$:

$$Y_l^m(\theta,\phi)=\Theta_{lm}(\theta)\,\Phi_m(\phi),\quad\left\{\begin{array}{ll} l & =0,1,2,\ldots\\ m & =-l,-(l-1),\ldots,(l-1),l.\end{array}\right. \tag{4.12}$$

The spherical harmonics are eigenfunctions to the angular part of the Laplace operator $\boldsymbol{\nabla}^2$,

$$\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]Y_l^m(\theta,\phi)=-l(l+1)\,Y_l^m(\theta,\phi),\quad\left\{\begin{array}{ll} l & =0,1,2,\ldots\\ m & =-l,\ldots,l.\end{array}\right. \tag{4.13}$$

Note that eventhough m is an index in $Y_l^m(\theta,\phi)$ it does not appear in the eigenvalue $-l(l+1)$. Below we list all spherical harmonics with $l\le 3$.

Y_0^0	=	$\sqrt{\frac{1}{4\pi}}$		Y_1^1	=	$-\sqrt{\frac{3}{8\pi}}$	$\sin\theta\, e^{i\phi}$
Y_3^3	=	$-\sqrt{\frac{35}{64\pi}}$	$\sin^3\theta\, e^{i3\phi}$	Y_1^0	=	$\sqrt{\frac{3}{4\pi}}$	$\cos\theta$
Y_3^2	=	$\sqrt{\frac{105}{32\pi}}$	$\sin^2\theta\cos\theta\, e^{i2\phi}$	Y_1^{-1}	=	$\sqrt{\frac{3}{8\pi}}$	$\sin\theta\, e^{-i\phi}$
Y_3^1	=	$-\sqrt{\frac{21}{64\pi}}$	$\sin\theta(5\cos^2-1)\, e^{i\phi}$	Y_2^2	=	$\sqrt{\frac{15}{32\pi}}$	$\sin^2\theta\, e^{i2\phi}$
Y_3^0	=	$\sqrt{\frac{7}{16\pi}}$	$(5\cos^3\theta - 3\cos\theta)$	Y_2^1	=	$-\sqrt{\frac{15}{8\pi}}$	$\sin\theta\cos\theta\, e^{i\phi}$
Y_3^{-1}	=	$\sqrt{\frac{21}{64\pi}}$	$\sin\theta(5\cos^2-1)\, e^{-i\phi}$	Y_2^0	=	$\sqrt{\frac{5}{16\pi}}$	$(3\cos^2\theta - 1)$
Y_3^{-2}	=	$\sqrt{\frac{105}{32\pi}}$	$\sin^2\theta\cos\theta\, e^{-i2\phi}$	Y_2^{-1}	=	$\sqrt{\frac{15}{8\pi}}$	$\sin\theta\cos\theta\, e^{-i\phi}$
Y_3^{-3}	=	$\sqrt{\frac{35}{64\pi}}$	$\sin^3\theta\, e^{-i3\phi}$	Y_2^{-2}	=	$\sqrt{\frac{15}{32\pi}}$	$\sin^2\theta\, e^{-i2\phi}$

(4.14)

4.1.4 The radial functions $R_{nl}(r)$

The last part of the Schrödinger equation Eq. (4.4) to be solved concerns the radial coordinate r. We plug in $\psi_E(\mathbf{r}) = R(r)\, Y_l^m(\theta,\phi)$, use the eigenvalue equation Eq. (4.13), and divide out the spherical harmonic $Y_l^m(\theta,\phi)$. The result is the following equation for $R(r)$,

$$\frac{d^2R(r)}{dr^2} + \frac{2}{r}\frac{dR(r)}{dr} - \frac{l(l+1)}{r^2}R(r) + \frac{2m_e}{\hbar^2}\left[E + \frac{\alpha}{r}\right]R(r) = 0. \tag{4.15}$$

This differential equation is then made dimensionless. Since the equation contains the parameters m_e, $\hbar$, and $\alpha = Ze^2/4\pi\epsilon_0$ we form the characteristic length a_Z and the characteristic energy E_Z (the index Z refers to the charge $+Ze$ of the nucleus),

$$a_Z = \frac{\hbar^2}{m_e\alpha} \tag{4.16a}$$

$$E_Z = \frac{m_e\alpha^2}{\hbar^2} \tag{4.16b}$$

In analogy with the parameter $l(l+1)$ for Θ_{lm} we now introduce the dimensionless (and perhaps complex) parameter n, and use that together with a_Z and E_Z to define dimen-

sionless measures of energy, n, and of length, ρ:

$$n = \sqrt{\frac{E_Z}{2|E|}}, \qquad \Rightarrow \qquad E = \frac{-1}{2n^2}\, E_Z, \tag{4.17a}$$

$$\rho = \frac{2}{n}\,\frac{r}{a_Z}, \qquad \Rightarrow \qquad r = \frac{n}{2}\,\rho\, a_Z. \tag{4.17b}$$

For negative energies in terms of ρ and n the radial equation Eq. (4.15) becomes

$$\frac{d^2R_{nl}(\rho)}{d\rho^2} + \frac{2}{\rho}\frac{dR_{nl}(\rho)}{d\rho} + \left[-\frac{1}{4} + \frac{n}{\rho} - \frac{l(l+1)}{\rho^2}\right] R_{nl}(\rho) = 0. \tag{4.18}$$

This equation is solved by studying the assymptotic behavior for $\rho \to \infty$ and $\rho \to 0$. In the limit $\rho \to \infty$ the terms in Eq. (4.18) containing $1/\rho$ and $1/\rho^2$ drop out, and thus

$$\frac{d^2R_{nl}(\rho)}{d\rho^2} - \frac{1}{4}R_{nl}(\rho) \approx 0 \qquad \Rightarrow \qquad R_{nl}(\rho) \approx e^{-\frac{1}{2}\rho}. \tag{4.19}$$

We have chosen the exponentially decaying solution since the probability of the electron to be infinitely far away from the nucleus must vanish. In the other limit, $\rho \to 0$ we keep the derivatives of R as well as the dominant term $l(l+1)/\rho^2$:

$$\frac{d^2R_{nl}(\rho)}{d\rho^2} + \frac{2}{\rho}\frac{dR_{nl}(\rho)}{d\rho} - \frac{l(l+1)}{\rho^2}R_{nl}(\rho) \approx 0 \qquad \Rightarrow \qquad R_{nl}(\rho) \approx \rho^l, \tag{4.20}$$

where we have chosen the solution that remains finite at $\rho = 0$. We thus end with the following form of $R_{nl}(\rho)$:

$$R_{nl}(\rho) = w_{nl}(\rho)\, \rho^l\, e^{-\frac{1}{2}\rho}. \tag{4.21}$$

Upon insertion of Eq. (4.21) into Eq. (4.18) we arrive at

$$\rho\, w''_{nl} + (2l + 2 - \rho)\, w'_{nl} + (n - l - 1)\, w_{nl} = 0. \tag{4.22}$$

To ensure the exponential decay for $\rho \to \infty$ given in Eq. (4.21) $w(\rho)$ must be a polynomium of finite degree, and from this follows that n must be an integer subject to the condition $n \geq l + 1$. Such polynomial solutions to the differential equation Eq. (4.22) are known to be the so-called generalized Laguerre polynomials L^{2l+1}_{n+l},

$$L^p_n(\rho) = \frac{(-1)^p(n!)^2}{p!(n-p)!}\, \rho^{-p} e^{\rho}\, \frac{d^{n-p}}{d\rho^{n-p}}(\rho^n e^{-\rho}), \quad 0 \leq p \leq n. \tag{4.23}$$

To ensure the normalization condition $\int_0^\infty |R_{nl}(\rho)|^2 \rho^2 d\rho = 1$ we find the radial wavefunction $R_{nl}(\rho)$ to be

$$R_{nl}(\rho) = -\frac{n}{4}\sqrt{\frac{(n-l-1)!}{[(n+l)!]^3}}\, \rho^l\, L^{2l+1}_{n+l}(\rho)\, e^{-\frac{1}{2}\rho}, \quad \begin{cases} n = 1, 2, 3, \ldots \\ n \geq l + 1. \end{cases} \tag{4.24}$$

Below we list the radial wavefunctions $R_{nl}(r)$ for $n = 1, 2, 3$ and $l = 0, 1, \ldots, n-1$. We have reinstated the actual radial coordinate r.

$$
\begin{array}{lcll}
R_{10}(r) & = & \dfrac{2}{\sqrt{a_Z^3}} & \exp\Big(-\dfrac{r}{a_Z}\Big) \\
\hline
R_{20}(r) & = & \dfrac{1}{\sqrt{2a_Z^3}} & \Big(1-\dfrac{1}{2}\dfrac{r}{a_Z}\Big)\exp\Big(-\dfrac{r}{2a_Z}\Big) \\
R_{21}(r) & = & \dfrac{1}{2\sqrt{6a_Z^3}} & \Big(\dfrac{r}{a_Z}\Big)\exp\Big(-\dfrac{r}{2a_Z}\Big) \\
\hline
R_{30}(r) & = & \dfrac{2}{3\sqrt{3a_Z^3}} & \Big(1-\dfrac{2}{3}\dfrac{r}{a_Z}-\dfrac{2}{27}\dfrac{r^2}{a_Z^2}\Big)\exp\Big(-\dfrac{r}{3a_Z}\Big) \\
R_{31}(r) & = & \dfrac{8}{27\sqrt{6a_Z^3}} & \Big(\dfrac{r}{a_Z}-\dfrac{1}{6}\dfrac{r^2}{a_Z^2}\Big)\exp\Big(-\dfrac{r}{3a_Z}\Big) \\
R_{32}(r) & = & \dfrac{4}{81\sqrt{30a_Z^3}} & \Big(\dfrac{r^2}{a_Z^2}\Big)\exp\Big(-\dfrac{r}{3a_Z}\Big)
\end{array}
\tag{4.25}
$$

4.2 The energies and sizes of the atomic orbitals

Returning to the normal units, we find the following energies E_n of the atomic orbitals,

$$E_n = -\frac{1}{2n^2}\,E_Z = -\frac{m_e\alpha^2}{2\hbar^2}\,\frac{1}{n^2} \qquad \Rightarrow \qquad E_n = -\frac{m_e Z^2 e^4}{32\pi^2\epsilon_0^2\hbar^2}\,\frac{1}{n^2}, \quad n = 1, 2, 3, \ldots \tag{4.26}$$

Note that n is the only quantum number entering the expression for the eigenenergies. Each energy level therefore has a degeneracy given by the number allowed values of l and m for a given n. Since $l = 0, 1, \ldots (n-1)$ and $m = -l, -(l-1), \ldots, (l-1), l$ we have

$$\text{degeneracy of } E_n = \sum_{l=0}^{n-1}(2l+1) = n^2. \tag{4.27}$$

The size of the atomic orbitals is not well defined due to the statistical nature of the wavefunction. It is only characterized fully by calculating all the expectation values $\langle r^k\rangle$ for any power k. Since the potential behaves as $1/r$ it is natural to estimate the size of the orbitals by calculating $\langle 1/r\rangle$,

$$\Big\langle\frac{1}{r}\Big\rangle_{nlm} = \int_0^\infty dr\, r^2\,\frac{1}{r}\,|R_{nl}(r)|^2 = \frac{1}{n^2}\,\frac{1}{a_Z}\,. \tag{4.28}$$

For the ground state ψ_{100} of the hydrogen atom we find (with $Z = 1$) the ground state

energy E_0 and the size a_0 to be

$$a_0 = a_{Z=1} = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} \quad = 0.053 \text{ nm}, \tag{4.29a}$$

$$E_0 = E_{Z=1} = -\frac{m_e e^4}{32\pi^2\epsilon_0^2\hbar^2} = -13.6 \text{ eV}. \tag{4.29b}$$

4.3 Atomic orbitals: shape and nomenclature

In Fig. 4.1 are shown the shape of some of the atomic orbitals $\psi_{nlm}(\mathbf{r})$ in hydrogen-like atoms. The figure contains all orbitals with $n \leq 4$ with the omission of negative values of m since the orbitals (n, l, m) and $(n, l, -m)$ have the same shape.

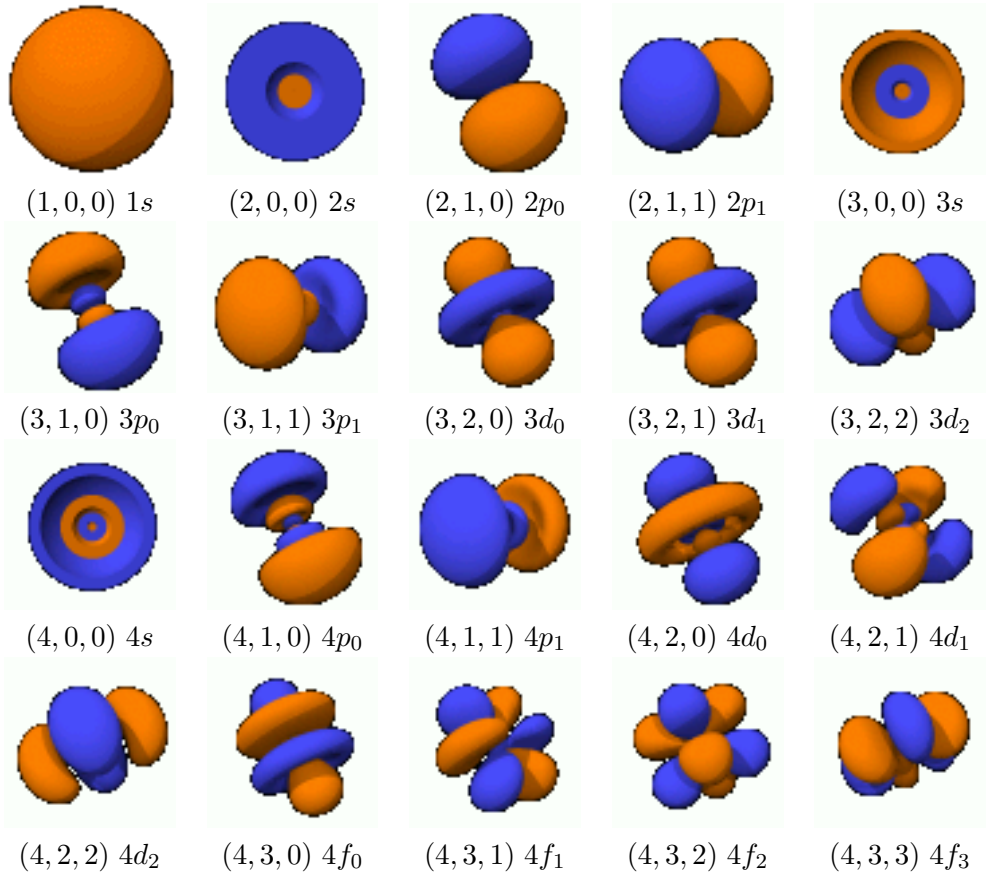


Figure 4.1: The shape of the electron orbitals $\psi_{nlm}(\mathbf{r})$ in a hydrogen-like atom for $n = 1, 2, 3, 4$, while $l = 0, 1, \ldots, n-1$, and $m = 0, 1, \ldots, l$. Each orbital is labelled by (n, l, m) and the standard code nX_m, where $X = s, p, d, f$ for $l = 0, 1, 2, 3$, respectively. Change of m to $-m$ does not change the shape of the orbital.

When specifying an orbital it is standard to write the value for n followed by a code letter for the value of l according to the following table

$$\begin{array}{lccccc} \text{value of } l: & 0 & 1 & 2 & 3 & 4 \\ \text{code letter:} & s & p & d & f & g \end{array} \tag{4.30}$$

The value of m can be given as a subscript to the code letter for l, but often it is omitted, and instead one writes the number of electrons with the same value of n and l as a superscript to the code letter for l,

$$\text{orbital term} = (n \text{ value})(l \text{ code letter})^{(\text{number of electrons})}. \tag{4.31}$$

As an example of the nomenclature we take the carbon atom. In its ground state it has two electrons (one spin up and one spin down) with $(n,l) = (1,0)$, two with $(n,l) = (2,0)$, and two with $(n,l) = (2,1)$. Thus the electron configuration for carbon is written as

$$[\text{C}] = 1s^2 2s^2 2p^2 = [\text{He}] 2s^2 2p^2, \tag{4.32}$$

where we also have use the electron configuration of He, the noble gas in the periodic table before C. This latter notation is helpful, since it emphasizes the orbitals of the valence electrons and just summarizes the inner core electrons as the filled electron orbitals of a nobel gas.

4.4 Angular momentum: interpretation of l and m

The angular part of the Laplacian $\boldsymbol{\nabla}^2$ in spherical coordinates is closely related to the angular momentum operator $\hat{\mathbf{L}}$. In quantum mechanics $\hat{\mathbf{L}}$ is defined by

$$\hat{\mathbf{L}} = \mathbf{r} \times \hat{\mathbf{p}} = \frac{\hbar}{i} \begin{pmatrix} y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \\ z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \\ x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \end{pmatrix}. \tag{4.33}$$

Transforming to spherical coordinates yields

$$\hat{\mathbf{L}} = \begin{pmatrix} \hat{L}_x \\ \hat{L}_y \\ \hat{L}_z \end{pmatrix} = \frac{\hbar}{i} \begin{pmatrix} -\sin\phi\frac{\partial}{\partial\theta} - \cot\theta\, \cos\phi\, \frac{\partial}{\partial\phi} \\ +\cos\phi\frac{\partial}{\partial\theta} - \cot\theta\, \sin\phi\, \frac{\partial}{\partial\phi} \\ \frac{\partial}{\partial\phi} \end{pmatrix}. \tag{4.34}$$

From this follows by straightforward algebra that

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{4.35}$$

But this is noting but the angular part of the Laplacian $\boldsymbol{\nabla}^2$, and combining Eqs. (4.12) and (4.13) with Eqs. (4.34) and (4.35) leads to the following important results

$$\hat{L}_z\, Y_l^m(\theta,\phi) = \hbar\, m\, Y_l^m(\theta,\phi) \tag{4.36}$$

$$\hat{\mathbf{L}}^2\, Y_l^m(\theta,\phi) = \hbar^2\, l(l+1)\, Y_l^m(\theta,\phi). \tag{4.37}$$

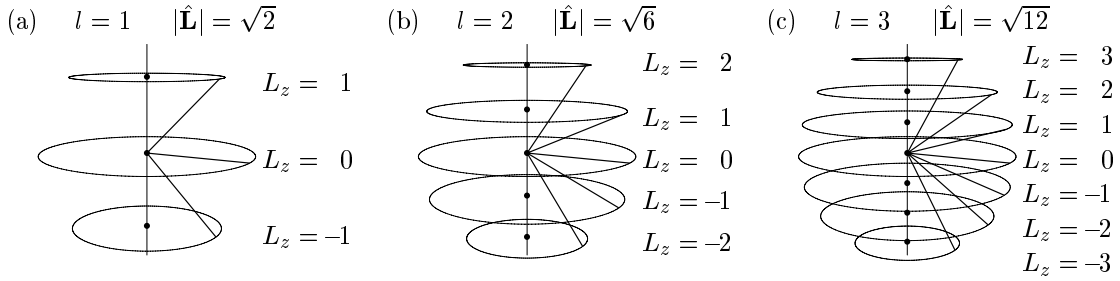


Figure 4.2: Illustration of the precessing angular momentum for $l = 1, 2, 3$ and $m = -l, -(l-1), \ldots, (l-1), l$. All angular momenta are given in units of $\hbar$. For a given set (l, m) both $\hat{\mathbf{L}}$ and $\hat{L}_z$ are fixed at the values $\sqrt{l(l+1)} \approx l + \frac{1}{2}$ and m, respectively. In contrast both $\hat{L}_x$ and $\hat{L}_y$ are completely indetermined. This can be depicted by letting $\hat{\mathbf{L}}$ precess around the z-axis keeping the z-component $\hat{L}_z$ fixed.

Consequently, for the wavefunction $Y_l^m(\theta, \phi)$ we can identify $\sqrt{l(l+1)}\,\hbar$ with the absolute size of the angular momentum, and $m\,\hbar$ with its z-component L_z.

Since one cannot determine the position $\mathbf{r}$ and the momentum $\hat{\mathbf{p}}$ simultaneously with absolute certainty, it is expected that one cannot dermine all three components of the angular momentum $\hat{\mathbf{L}} = \mathbf{r} \times \hat{\mathbf{p}}$ simultaneously with absolute certainty. This is in fact true. As seen in Eqs. (4.36) and (4.37) it is possible to determine the values of $\hat{L}_z$ and $\hat{\mathbf{L}}$ simultaneously. But we also note that that even if $\hat{L}_z$ takes its maximal value $l\hbar$ it is still smaller than the total size $\sqrt{l(l+1)}$ of $|\hat{\mathbf{L}}|$,

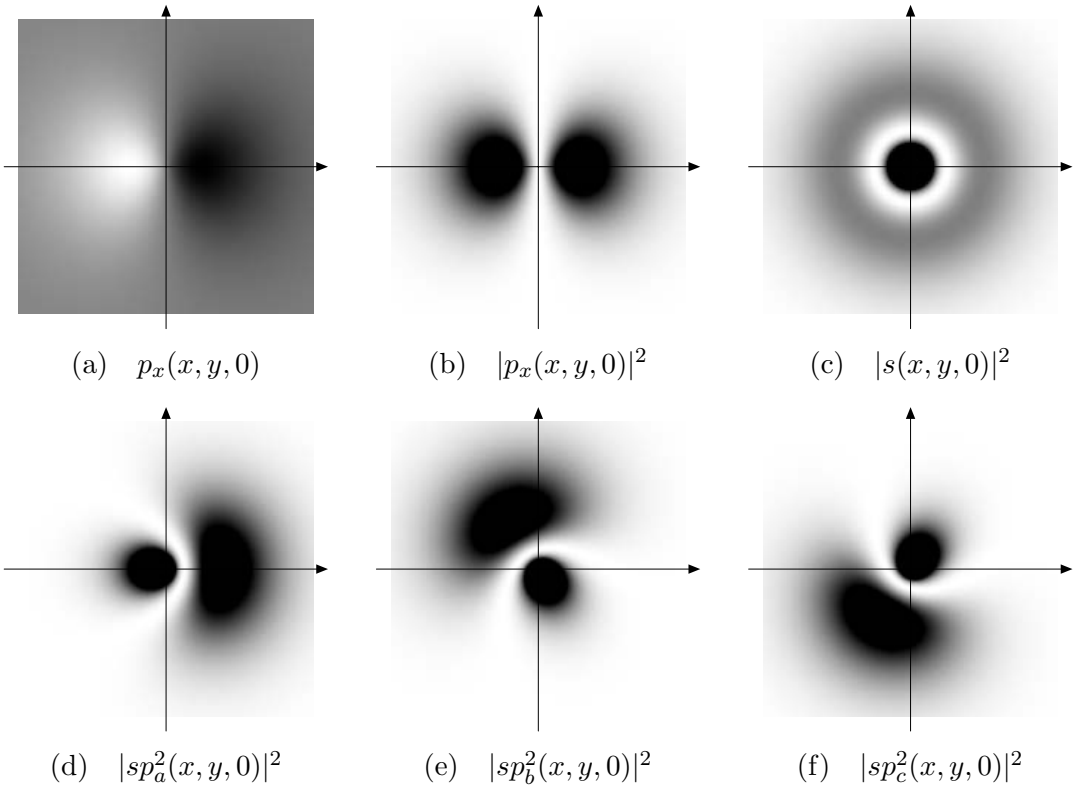
$$\max\{\hat{L}_z\} = l\hbar < \sqrt{l(l+1)}\,\hbar = |\hat{\mathbf{L}}|. \tag{4.38}$$

We can conclude that there is always some angular momentum left in the x and y direction. This can be illustrated as shown in Fig. 4.2 by imagining that the $\hat{\mathbf{L}}$-vector precesses around the z-axis with a fixed z-component $\hat{L}_z$.

4.5 The carbon atom and sp^2 hybridization

The ability of a single carbon atom to form up to four strong covalent bonds makes it a key player in organic chemistry and hence a corner stone in the formation of biological tissue. Moreover, the discoveries of the C_{60} "bucky ball" and the carbon nanotube molecules in 1986 and 1992, respectively, have given carbon a status in nanoscience and nanotechnology comparable to that of silicon in microtechnology. We therefore give a brief overview of carbon and carbon nanotubes in the following sections.

Carbon is the sixth element in the periodic table. The nucleus of the most abundant carbon isotope contains six protons and six neutrons, while the electron configuration is given by $1s^2 2s^2 2p^2$ as stated in Eq. (4.32). Of these six electrons the four in the $2s$ and $2p$ orbitals are valence electrons able to participate in the formation of chemical bonds.



(a) $p_x(x,y,0)$ (b) $|p_x(x,y,0)|^2$ (c) $|s(x,y,0)|^2$

(d) $|sp_a^2(x,y,0)|^2$ (e) $|sp_b^2(x,y,0)|^2$ (f) $|sp_c^2(x,y,0)|^2$

Figure 4.3: A cross-section in the xy plane of wavefunctions related to the sp^2 hybridization. (a) Density plot of $p_x(x,y,0)$, black is positive and white is negative. Panels (b)-(f) are density plots of the absolute square of the indicated wavefunctions, white is zero and black is maximal density. Note the specific directions, $0°$, $120°$, and $-120°$ of the three hybridized sp^2 orbitals $sp_a^2(\mathbf{r})$, $sp_b^2(\mathbf{r})$, and $sp_c^2(\mathbf{r})$ in panels (d), (e), and (f), respectively.

It is a well known fact that atoms are most stable when the s and p orbitals of the outer shell[2] are occupied. An s orbital can contain two electrons (spin up and spin down), while a p orbital, where per definition $l = 1$, can contain $2 \times 3 = 6$ (2 for spin and 3 for $m = -1, 0, 1$), so stability is obtained when eight electrons are present in the valence shell. This is known as the octet rule,

$$[\text{stable configuration for shell } n] = ns^2np^6. \tag{4.39}$$

We thus see that a carbon atom is short of four electron to have a maximally stable configuration. The octet rule can be fulfilled by the formation of chemical bonds. This is done most efficiently by making superpositions of the symmetric $2s$, $2p_1$, $2p_0$, and $2p_{-1}$ orbitals of Fig. 4.1, which all have the same energy, to form asymmetric orbitals that points in specific directions. This is known as hybridization.

[2] the outer shell contains all orbitals having the highest value of the radial quantum number n.

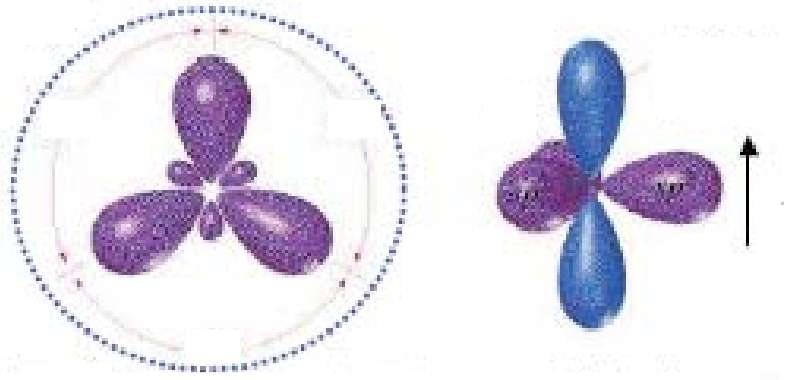


Figure 4.4: (a) A top-view down on the xy plane showing the three hybridized orbitals $sp_a^2(\mathbf{r})$, $sp_b^2(\mathbf{r})$, and $sp_c^2(\mathbf{r})$ with the characteristic 120 $\circ$ angle between them. (b) A 3D view of the non-hybridized p_z orbital pointing in the z direction (indicated by the verticla arrow) perpendicular to the three sp^2 orbitals in the xy plane.

First we transform $2s$, $2p_1$, $2p_0$, and $2p_{-1}$ into s, p_x, p_y, and p_z (where the index $n = 2$ is suppressed) as follows,

$$s(\mathbf{r}) \equiv 2s \qquad = R_{20}Y_0^0 \qquad = A\sqrt{2}(1-\rho)e^{-\rho}, \tag{4.40a}$$

$$p_x(\mathbf{r}) \equiv 2p_1 - 2p_{-1} = R_{21}Y_1^1 - R_{21}Y_{-1}^1 = A\cos\phi\,\sin\theta\,\rho e^{-\rho}, \tag{4.40b}$$

$$p_y(\mathbf{r}) \equiv 2p_1 + 2p_{-1} = R_{21}Y_1^1 + R_{21}Y_{-1}^1 = A\sin\phi\,\sin\theta\,\rho e^{-\rho}, \tag{4.40c}$$

$$p_z(\mathbf{r}) \equiv 2p_0 \qquad = R_{21}Y_1^0 \qquad = \sqrt{6}A\cos\theta\,(1-\rho)e^{-\rho}, \tag{4.40d}$$

where A is a constant and ρ is the dimensionless length from Eq. (4.17b).

We can now define the so-called sp^2 hybridization. It consists of the following superpositions of $s(\mathbf{r})$, $p_x(\mathbf{r})$, and $p_y(\mathbf{r})$, i.e., superpositions involving one s and two p functions (leaving p_z as it is),

$$sp_a^2(\mathbf{r}) \equiv \sqrt{\frac{4}{6}}\,s(\mathbf{r}) \;+\; \sqrt{\frac{2}{6}}\,p_x(\mathbf{r}), \tag{4.41a}$$

$$sp_b^2(\mathbf{r}) \equiv \sqrt{\frac{2}{6}}\,s(\mathbf{r}) \;-\; \sqrt{\frac{1}{6}}\,p_x(\mathbf{r}) \;+\; \sqrt{\frac{3}{6}}\,p_y(\mathbf{r}), \tag{4.41b}$$

$$sp_c^2(\mathbf{r}) \equiv \sqrt{\frac{2}{6}}\,s(\mathbf{r}) \;-\; \sqrt{\frac{1}{6}}\,p_x(\mathbf{r}) \;-\; \sqrt{\frac{3}{6}}\,p_y(\mathbf{r}). \tag{4.41c}$$

The shape of the three hybridized sp^2 orbitals is shown in Fig. 4.3. Here in panels (d), (e), and (f) it is seen how sp_a^2 points in the x direction, while sp_b^2 and sp_c^2 are rotated $+120^\circ$ and -120° in the xy plane, respectively. The reason for this result lies in the fact that p_x is positive for $x > 0$ and negative for $x < 0$, as shown in Fig. 4.3(a).[3] When adding this

[3]Similarly for p_y, however rotated 90° in the xy plane (not shown in Fig. 4.3).

angular dependent function to the rotational invariant function s, Fig. 4.3(c), we obtain the assymetric results.

The p_z orbital does not enter the sp^2 hybridization, and it remains a figure eight-like orbital perpendicular to the xy plane containing the three sp^2 orbitals built from s, p_x, and p_y. In Fig. 4.4 is shown two 3D sketches of the resulting sp^2 hybridized orbitals and the non-hybridized orbital p_z.

We only study sp^2 hybridization here, since this is relevant to the formation of graphene planes and carbon nanotubes. However, part of the rich organic chemistry of carbon relates to the fact that also sp^1 and sp^3 can occur. In sp^1 hybridization s and p_x hybridizes into sp^1_x leaving p_y and p_z perpendicular to sp^1 and each other. The uni-directional sp^1 bond is found in compounds containing the linear carbon tripple bond $-\mathrm{C}\equiv\mathrm{C}-$. In sp^3 hybridization all four valence orbitals hybridize an form the characteristic tetragonal structure that each carbon atoms has in the diamond structure.

4.6 Graphene, sigma and pi bonds

As mentioned the formation of chemical bonds is one way for an atom to fulfil the octet rule. It is a general fact of quantum mechanics, that when an orbital from one atom is in close proximity of an orbital from another atom, two new hybridized orbitals can be formed by superposition. We shall use this description to explain the graphene structure, which is a planar hexagonal lattice of carbon atoms. Graphene is the basis for understanding carbon nanotubes.

The following discussion is a simplification of the real quantum orbital calculations. To be specific, consider two carbon atoms 1 and 2. Let atom 1 have the sp^2 orbitals have the orientations $sp^2_{a1} \sim 0°$, $sp^2_{b1} \sim +120°$, and $sp^2_{c1} \sim -120°$ as in Fig. 4.3(d-f), and let atom 2 be in the mirror inverted state such that $sp^2_{a2} \sim 180°$, $sp^2_{b2} \sim -60°$, and $sp^2_{c2} \sim +60°$.

The two orbitals sp^2_{a1} and sp^2_{a2} points towards each other. When the two carb on atoms therefore are close enough, they hybridize and form the two sigma-bonds σ and σ^* with low and high energy, respectively, defined as follows

$$\sigma(\mathbf{r}) = \frac{1}{\sqrt{2}}\left[sp^2_{a1}(\mathbf{r}) + sp^2_{a2}(\mathbf{r})\right], \quad \text{very low energy,} \tag{4.42}$$

$$\sigma^*(\mathbf{r}) = \frac{1}{\sqrt{2}}\left[sp^2_{a1}(\mathbf{r}) - sp^2_{a2}(\mathbf{r})\right], \quad \text{very high energy.} \tag{4.43}$$

The σ bond is the strong covalent bond that makes the carbon compounds very stable. The σ^* orbital has such a high energy that the electrons cannot be excited up into this orbital, it has thus a negligible influence on the molecule and can be disregarded. Hence, it is always the σ bond that matters. All sp^2 orbitals can pairwise form a σ-bond (and an inaccessible σ^*-orbital) with another atom. Each σ bond is always fully occupied with two electrons (spin up and down), one from each atom. The three σ bonds thus accounts for three of the four valence electrons of the carbon atom.

Similarly, the p_{z1} and p_{z2} orbitals from the two carbon atoms can form a superposition. However, since they are parallel to each other their spatial overlap is not as large as the

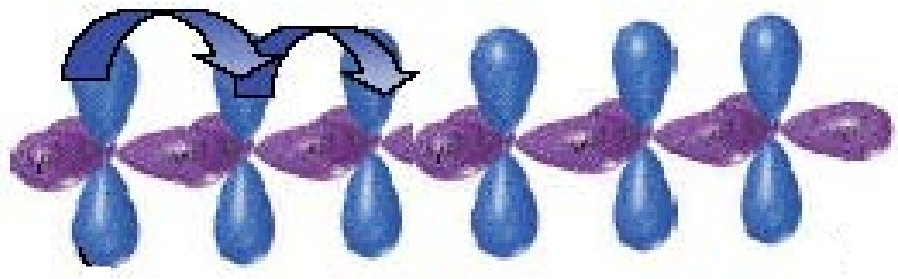


Figure 4.5: The σ and π orbitals in a chain of carbon atoms. The electrically inert σ orbitals, each fully occupied with two electrons (spin up and down) connects the carbon atoms pairwise in the xy plane and they are the reason for the strength of the carbon bonds. The π and π^* bonds between the p_z are only half occupied. Electrons can therefore hop between the π orbitals along the carbon chain, which is the reason for the high conductivity of certain carbon compounds.

two sp_a^2 orbitals that pointed towards each other, so the energies involved are much less. The hybridized bonds are called π-bonds,

$$\pi(\mathbf{r}) = \frac{1}{\sqrt{2}}\left[p_{z1}(\mathbf{r}) + p_{z2}(\mathbf{r})\right], \quad \text{slightly lower energy,} \tag{4.44}$$

$$\pi^*(\mathbf{r}) = \frac{1}{\sqrt{2}}\left[p_{z1}(\mathbf{r}) - p_{z2}(\mathbf{r})\right], \quad \text{slightly higher energy.} \tag{4.45}$$

The small energy difference means that while π orbital is the ground state, the π^* orbital is still accessible to the electrons, thus one expects (and finds) fast electron transfer (equal to large electrical conductivity) between the π orbitals in carbon compounds. This is sketched in Fig. 4.5. The π orbital can hold two electrons (spin up and down), one from each carbon atom. This accounts for the last valence electron.

To summarize: the strength of the carbon compounds is due to the very strong and fully occupied σ bonds that lies in the xy plane. These bonds are electrically inert because the unoccupied σ^* orbitals have too high an energy to be reached. The rather high electrical conductivity of certain carbon compounds is due to the π orbitals because the unoccupied π^* bonds have only a slightly higher energy, and can therefore be reached.

It is now very easy to describe graphene, the planar hexagonal lattice of carbon atoms, that upon stacking forms graphite. In Fig. 4.6(a) is shown a the atomic structure of such a planar sheet of graphene. The small circular disks represent carbon atoms. The lines between the disks represent the σ bonds. The characteristic 120° angles from Fig. 4.4 are clearly recognized in the regular lattice of Fig. 4.6(a). The mechanical in-plane strength of graphene is large due to the strong σ bonds.

While not shown in Fig. 4.6(a) there is also the π orbitals forming a cloud of electrons just above and just below the plane. Following the previous discussion about the occupied π orbitals and the unoccupied π^* orbitals, it is clear that graphene has a very good electrical conductivity parallel to the plane.

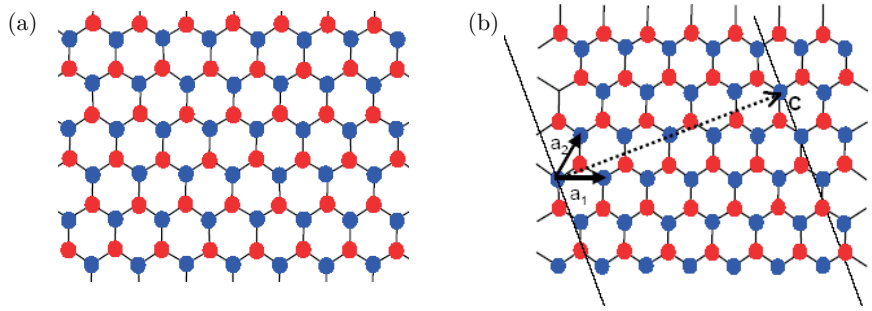


Figure 4.6: (a) The planar hexagonal lattice of carbon atoms (the disks) forming a graphene sheet. The mechanical in-plane strength of graphene originates from the strong σ bonds (the lines connecting the disks). (b) The base vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ defining the unit cell of the hexagonal lattice. The sloped straight lines are the lines to be joined upon rolling up the graphene sheet into a carbon nanotube perpendicular to the chiral vector $\mathbf{c} = n\mathbf{a}_1 + m\mathbf{a}_2$.

Graphite is formed by stacking a huge number of graphene sheets on top of each other, and consequently it is a highly anisotropic material. Graphite yields easily to shear stresses, since each graphene sheet is only weakly bound to each other. The lack of free electron orbitals make van der Waals forces (see Sec. 5.4.2) the only binding perpendicular to the sheets. Electrically graphite is a poor conductor perpendicular to the graphene sheets, but it has a decent conductivity along the sheets.

4.7 Carbon nanotubes

Carbon nanotube are extraordinary macromolecules containing only carbon. They are formed by rolling up graphene sheets as illustrated in Figs. 4.6 and 4.7. There exist two categories of nanotubes, the single-wall nanotube (SWNT) as shown in Figs. 4.6 and 4.7 and multi-wall nanotubes (MWNT), which consist of several concentric single-wall nanotubes. In experiments one often finds that the nanotubes collect in so-called bundles or ropes.

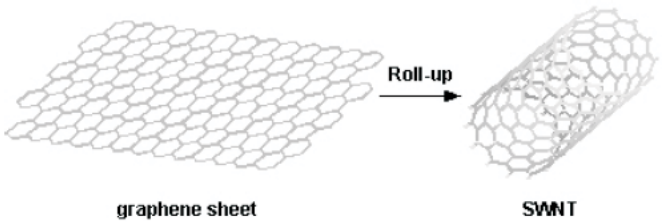


Figure 4.7: The formation of a single-wall carbon nanotube (SWNT) by rolling up a sheet of graphene. The chiral vector for such a roll-up is defined in Fig. 4.6(b).

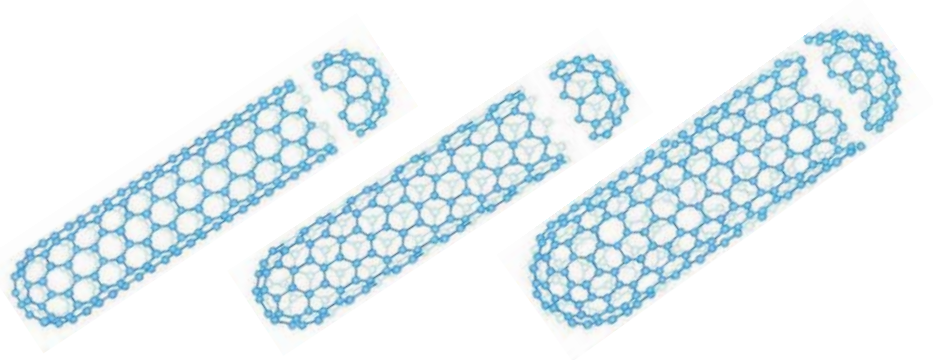


Figure 4.8: The three single-wall carbon nanotubes $(5,5)$, $(9,0)$, and $(10,5)$.

There exist an infinity of different single-wall nanotube the reason being that it is possible to roll-up a graphene sheet in many ways. A single-wall nanotube is uniquely defined by its chiral vector (see Fig. 4.6(b)) or equivalently its two-integer index (n,m),

$$\text{nanotube } (n,m) \text{ has the chiral vector } \mathbf{c} = n\mathbf{a}_1 + m\mathbf{a}_2, \tag{4.46}$$

where $textbf{a}_1$ and $\mathbf{a}_2$ are the base vectors of the unit cell in the hexagonal graphene lattice. Examples of three different carbon nanottube structures are seen in Fig. 4.8

Whereas the mechanical properties a single-wall carbon nanotubes is only weakly on the chirality (n,m), it turns out that the electrical properties are strongly dependent on the chirality. It is beyond the scope of these notes to prove the following result, but it is a straightforward, rather tedious, exercise in single-electron tight binding theory,

$$\begin{array}{lcll} n-m & = & 3\times\text{integer} & \Rightarrow \quad \text{metallic,} \\ n-m & \neq & 3\times\text{integer} & \Rightarrow \quad \text{semiconducting.} \end{array} \tag{4.47}$$

The diameter d and the chiral angle θ of a single-wall carbon nanotube are given by the C=C bond length a and the chiral index (n,m) by

$$d = \frac{a}{\pi}\sqrt{n^2+m^2+nm}, \tag{4.48}$$

$$\tan\theta = \frac{\sqrt{3}\,m}{2n+m}. \tag{4.49}$$

Typical density	1400 kg/m^3	Typical diameter	1.3 nm
Max tensile strength	30 GPa	Bond length C=C	0.142 nm
Young's modulus	1000 GPa	Optical gap	0.5 eV
Thermal conductivity	2000 W/m/K	Max current density	10^{13} A/m^2

Table 4.1: Some physical parameters of single-wall carbon nanotubes.

Chapter 5

Atomic force microscopy (AFM)

The atomic force microscope (AFM) is one of the many recently developed types of nanometer-resolution microscopes or scanning probe microscopes (SPM). The AFM has a much wider range of applicability than the other SPMs: it can be used to observe and manipulate nanometer-sized objects of both conductive and insulating nature in both vacuum, air, gasseous, and liquid environments. The following sections provide a more detailed treatment of the physcical principles behind the AFM.

5.1 The basic principles of the AFM

A sketch of the layout of a typical AFM set-up is shown in Fig. 5.1. The central component is the nanometer-sized tip mounted on the elastic cantilever. By use of the piezo-electric

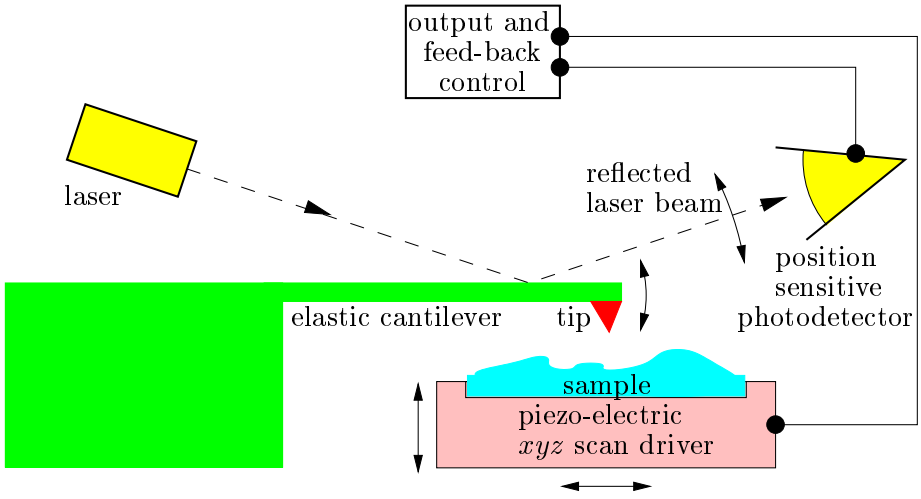


Figure 5.1: A sketch of the layout of a typical AFM set-up showing the key components described in the text.

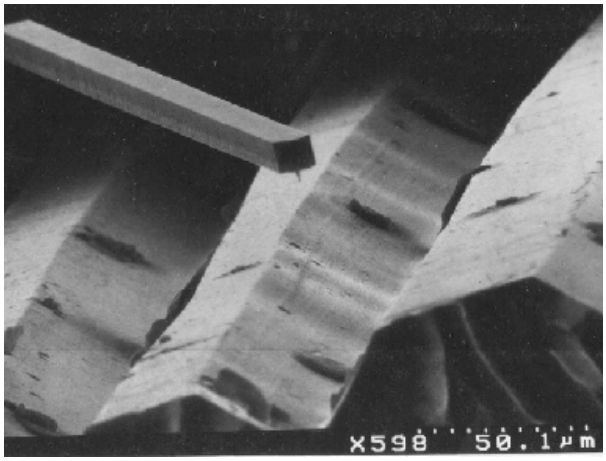


Figure 5.2: A scanning electron microscope picture of a silicon based AFM cantilever with a minute tip. The cantilever is shown on a Simon & Garfunkel vinyl record (J. Brugger, IMT Neuchatel, 1994).

xyz scan driver the tip is scanned across the sample. The distance between the tip and the sample can be anywhere between 0 and 100 nm. Dependent on the topography and contents of the surface of the sample the force acting on the tip changes during the scan resulting in a position dependent deflection of the elastic cantilever. The deflection is monitored by reflecting a laser beam on the cantilever and recording the movement of the reflected beam with a position sensitive photodetector. The signal from the photodetector can be used in a feed-back loop with the xyz scan driver to control the motion of the tip.

To build a well-functioning AFM several requirements has to be fulfilled:

1. The spring constant of the cantilever should be small enough to allow detection of minute atomic forces.
2. The resonance frequency of the cantilever should be as high as possible to minimize sensitivity to external mechanical vibrations.
3. The tip should be as sharp as possible to allow atomic resolution.
4. The tip should be as narrow as possible to allow penetration into deep troughs on the surface.

We shall look into these requirements in the following sections, where the different modes of operations are described.

5.2 The cantilever: spring constant and resonance frequency

The actual design geometry of an AFM cantilever can vary; for one specific example see Fig. 5.2. For simplicity we shall in the following analysis just consider a beam of length

L and rectangular cross section with height h and width w as shown in Fig. 5.3a. Due to interactions with the surface atoms of the sample an external force $F_{\rm tip}$ acts on the tip and bends the cantilever as shown in Fig. 5.3b.

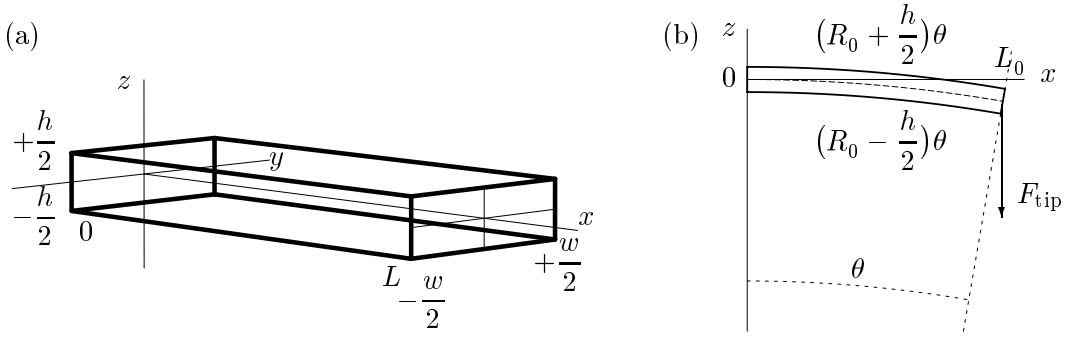


Figure 5.3: (a) An AFM cantilever of dimensions $L\times w\times h$. (b) The approximately circular bending of the cantilever due to the external force $F_{\rm tip}$ acting on the tip. The radius of curvature of the centerline is R_0. The bending has ben exaggerated on the drawing to be visible. Normally, the bending of the 100 μm long cantilever is less than 10 nm.

It is a simple exercise of applied elasticity theory to find the bending Δz of the tip due to the force $F_{\rm tip}$. Hooke's law for elastic solids relates the force per area (the stress) σ_{xx} with the relative change in length (the strain) $\Delta L/L_0$,

$$\sigma_{xx} = Y\,\frac{\Delta L}{L_0}, \tag{5.1}$$

where Y is Young's modulus, a material parameter we already encountered and calculated an estimate of in Exercise 3.7. Due to the bending the cantilever is compressed on the lower surface, $L < L_0$ at $z = -h/2$, elongated on the upper surface, $L > L_0$ at $z = +h/2$, and unchanged at the center plane, $L = L_0$ at $z = 0$. Thus the length $L(z)$ of the cantilever is a function of z. We assume that the bended cantilever has the circular geometry depicted in Fig. 5.3b. This is an approximation that will allow us to obtain decent results without too much math. The math is simplified due to the fact that a circle has a constant radius of curvature. The exact results are quoted along the way.

At Fig. 5.3b we see that there is an angle θ between the end planes at the two ends at $x = 0$ and $x = L$. If we denote the radius of the bended but unstretched center plane $z = 0$ by R_0 we have $L_0 = R_0\theta$ and deduce that

$$L(z) = (R_0 + z)\theta = \Big(1 + \frac{z}{R_0}\Big)L \qquad \Rightarrow \qquad \frac{\Delta L}{L_0} = \frac{L - L_0}{L_0} = \frac{z}{R_0}. \tag{5.2}$$

The external force $F_{\rm tip}$ acts on the tip and bends the cantilever. However, even in this situation the cantilever is in equilibrium, and consequently the external force torque $L_0F_{\rm tip}$ around the line $(x, z) = (0, 0)$ at the left end of the cantilever must balance the internal

torque set up by the internal stress forces σ_{xx} times the arm z:

$$L_0 F_{\rm tip} = \int_{-\frac{w}{2}}^{+\frac{w}{2}} dy \int_{-\frac{h}{2}}^{+\frac{h}{2}} dz\, z\, \sigma_{xx} = w \int_{-\frac{h}{2}}^{+\frac{h}{2}} dz\, z\, Y \frac{z}{R_0} = \frac{1}{12}\, Y\, \frac{w h^3}{R_0} = \frac{1}{6}\, Y\, \frac{w h^3}{L_0^2}\, \Delta z, \quad (5.3)$$

where Δz is the vertical displacement of the cantilever at $x = L$. It is given by $\Delta z = R_0(1 - \cos\theta) \approx \frac{1}{2} R_0 \theta^2 = \frac{L_0^2}{2R_0}$ or $1/R_0 = 2\Delta z / L_0^2$. So the force $F_{\rm tip}$ can be expressed by the displacement Δz of the tip and a spring constant K:

$$F_{\rm tip} = K\, \Delta z, \quad (5.4a)$$

$$K = \frac{1}{6} \left(\frac{h}{L}\right)^3 w Y \quad \text{(the circular shape approximation)}, \quad (5.4b)$$

$$K = \frac{1}{4} \left(\frac{h}{L}\right)^3 w Y \quad \text{(the exact result)}. \quad (5.4c)$$

From the spring constant we can estimate the (smallest) resonance frequency ω_0 of the cantilever (corresponding to a quarter wave-length oscillation) when it is oscillating freely. It is given by $\omega_0 = \sqrt{K/m_{\rm eff}}$. Note that it is not the cantilever mass m that appears in the formula but an effective mass $m_{\rm eff}$. This is due to the fact that the various parts of the cantilever do not oscillate with the same amplitude. For $x = 0$ the amplitude is zero, while for $x = L$ it is the maximal Δz. In Exercise 5.2 we show how the circular shape approximation leads to $m_{\rm eff} = m/5$ and thus

$$\omega_0^2 = \frac{K}{m_{\rm eff}} = \frac{5}{6}\, \frac{Y w}{m} \Big(\frac{h}{L}\Big)^3 = \frac{5}{6}\, \frac{Y w}{\rho w h L} \Big(\frac{h}{L}\Big)^3 = \frac{5}{6}\, \frac{Y}{\rho}\, \frac{h^2}{L^4}, \quad (5.5)$$

where we have introduced the mass density ρ. The approximate and exact results are

$$\omega_0 = 0.91 \sqrt{\frac{Y}{\rho}}\, \frac{h}{L^2} \quad \text{(the circular shape approximation)}, \quad (5.6a)$$

$$\omega_0 = 1.02 \sqrt{\frac{Y}{\rho}}\, \frac{h}{L^2} \quad \text{(the exact result)}. \quad (5.6b)$$

Eqs. (5.4c) and (5.6b) are central to the design of an AFM. The ratio h/L should be small to give a small spring constant and hence a good force sensitivity, but h/L^2 should be large enough to ensure a high resonance frequency that minimizes the sensitivity to external mechanical vibrations. The solution is to make small cantilevers. One good method is to make them in silicon since one then can take advantage of standard silicon processing techniques perfected by the microelectronics industry.

A typical choise is $L = 100$ μm, $w = 10$ μm, and $h = 1$ μm. A deflection $\Delta z \approx 1$ Å is easily detected and since for silicon $Y = 1.6 \times 10^{11}$ Pa and $\rho = 2.33 \times 10^3$ kg m^{-3}, we obtain the following spring constant K, tip-force $F_{\rm tip}$, and resonance frequency ω_0 from

Eqs. (5.4a), (5.4b), and (5.6a):

$$K = 0.40 \text{ N m}^{-1} \tag{5.7a}$$
$$F_{\text{tip}} = \ 40 \text{ pN} \tag{5.7b}$$
$$\omega_0 = \ 927 \text{ kHz}. \tag{5.7c}$$

These numbers are good news. The typical atomic forces acting on an AFM-tip is of the order 1 nN, and mechanical noise above 1 kHz is strongly suppressed.

5.3 Contact mode

When operating in the so-called contact mode the AFM tip is forced down into the surface until repulsive contact with the core electrons of the surface atoms is obtained. During a scan the AFM tip simply follows the topographic features of the sample with atomic resolution (if the tip is sharp enough).

To illustrate the physics of the contact mode it is sketched in Fig. 5.4a how the electron orbitals get deformed as two atoms approach each other. At long range the weak attractive van der Waals force is the dominant force. This force will be treated in the next section. At intermediate range the cloud of valence electrons surrounding the atom get in contact and become deformed as a diatomic molecule is formed. At the closest range also the core electron clouds of the two atoms touch each other and get deformed. At this point the energy increases rapidly as a function of decreasing inter-atomic distance. It is extremely costly energy wise to compress the tightly bound core electrons further.

In general it is very difficult to calculate the exact form of the energy versus distance. There are no small parameters in the problem and many atomic orbitals must be taken into account in a serious calculation. It is customary, however, to parametrize the potential by the Lennard-Jones model potential

$$E(r) = B\Big[\Big(\frac{A}{r}\Big)^{12} - \Big(\frac{A}{r}\Big)^{6}\Big]. \tag{5.8}$$

The r^{-6}-term is well understood as we shall see shortly. The r^{-12}-term on the other hand is just an ansatz. There is no particular reason why the power of 12 has been chosen except that it is bigger than 6 and mimmicks the very strong repulsion observed in many experiments.

The energy versus distance curve $E(r)$ in Fig. 5.4a ressembles the simple energy versus size-of-atom curve $E(a)$ that we studied in Exercise 2.3. For comparison $E(a)$ is depicted in Fig. 5.4b.

5.4 Non-contact mode

When operating in the non-contact mode the AFM-tip is not touching the surface. Instead it is influenced by the long range van der Waals forces. The advantage of the non-contact mode is that it does not destroy even the softest samples. But what is the origin of the van der Waals forces? This question will be answered now.

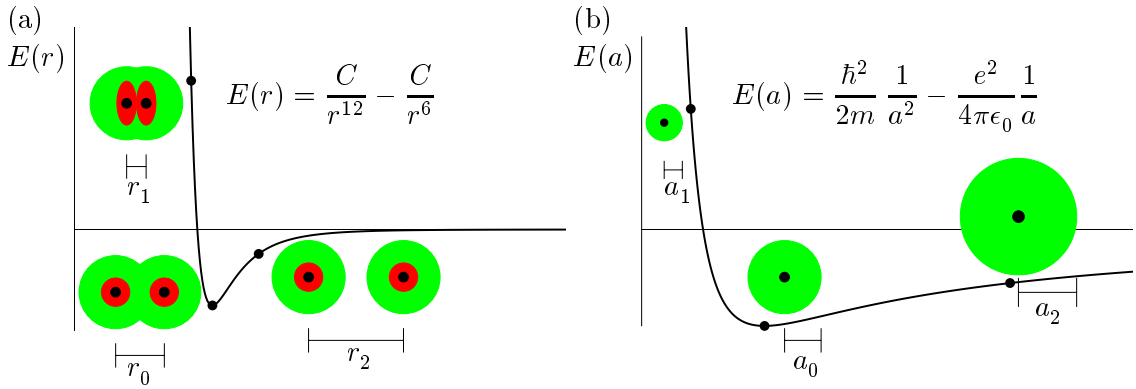


Figure 5.4: Similarities and differences in the energy of diatomic molecules and single atoms. In the three particular situations shown in each of the two panels light gray represents valance electrons, dark gray core electrons, and black the atomic nuclei. (a) The Lennard-Jones model-potential of the energy $E(r)$ in a diatomic molecule as a function of the distance r between the nuclei. (b) The simple model of the ground state energy $E(a)$ of the hydrogen atom as a function of the average distance a between the electron and the proton (see Eq. (2.14) and Exercise 2.3).

5.4.1 Atomic polarization

The first step in understanding the van der Waals forces is a study of atomic polarization. When an electric field $\mathcal{E}$ is applied to a charge neutral atom the positive and negative charges are displaced slightly from their equlibrium positions: the positive charges in the direction of the $\mathcal{E}$-field and the negative charge the opposite way. But how big a displacement d are we talking about? Well, an estimate can be obtained from the simple model of the hydrogen atom given in Eq. (2.14) (see also Fig. 5.4b). We make a Taylor-expansion of $E(a)$ around the minimum $a = a_0$ to second order in $(a - a_0)$ to find the spring constant of the electron-proton system (compare this with Exercise 3.7):

$$E(a) \approx E(a_0) + \frac{1}{2}\, E''(a_0)\, (a - a_0)^2. \tag{5.9}$$

The spring constant in this expression is easily found to be

$$K = E''(a_0) = 3\,\frac{\hbar^2}{m a_0^2}\,\frac{1}{a_0^2} - 2\,\frac{e^2}{4\pi\epsilon_0 a_0}\,\frac{1}{a_0^2} = \frac{2E(a_0)}{a_0^2} = 1.55 \times 10^3 \text{ N m}^{-1}. \tag{5.10}$$

The displacement d between the electron and proton is therefore found by simple force balance between the spring force Kd and the electric force $e\mathcal{E}$:

$$e\mathcal{E} = Kd \qquad \Rightarrow \qquad ed = \frac{e^2}{K}\,\mathcal{E}. \tag{5.11}$$

The quantity ed is called the dipole moment of the atom. An electric dipole consists of an equal amount of positive and negative charge q that has been displaced the distance

d (in general a vector pointing from $-q$ to $+q$) from one another. The dipole moment of the dipole is just the product qd.

What is then the electrostatic energy $U(r)$ of the dipole qd in the electric field $\mathcal{E}(r)$? The electric field is due to some potential $V_{\mathcal{E}}(r_0)$, that is $\mathcal{E}(r_0) = -V'_{\mathcal{E}}(r_0)$. We just study the case where the dipole is parallel to the electric field, so we place $+q$ at $r_0 + d/2$ and $-q$ at $r_0 - d/2$. We then obtain the energy $U(r)$ as

$$U(r_0) = +qV_{\mathcal{E}}\big(r_0+\frac{d}{2}\big) - qV_{\mathcal{E}}\big(r_0-\frac{d}{2}\big) \approx q\big(+\frac{d}{2}\big)V'_{\mathcal{E}}(r_0) - q\big(-\frac{d}{2}\big)V'_{\mathcal{E}}(r_0) = -qd\mathcal{E}(r_0). \quad (5.12)$$

We have seen that an electric field can induce a dipole moment, but the converse is also true: a dipole moment induces an electric field. Consider in 1D a dipole at the center of the coordinate system, $+q$ at $d/2$ and $-q$ at $-d/2$. What is then the electric potential $V(r)$ and the corresponding electric field $\mathcal{E}(r)$? Due to the opposite charges we can write their respective potentials as $+V_0(r - \frac{d}{2})$ and $-V_0(r - \frac{d}{2})$, where $V_0(x) = q/(4\pi\epsilon_0 x)$ is the usual Coulomb potential. For $V(r)$ we thus get

$$V(r) = V_0(r - \frac{d}{2}) - V_0(r + \frac{d}{2}) \approx -V'_0(r)\frac{d}{2} - V'_0(r)\frac{d}{2} = -d\,V'_0(r) = \frac{dq}{4\pi\epsilon_0}\,\frac{1}{r^2}. \quad (5.13)$$

The electric field is just the derivative of the potential, so

$$\mathcal{E}(r) = -V'(r) = \frac{2dq}{4\pi\epsilon_0}\,\frac{1}{r^3}. \quad (5.14)$$

We see here the characteristic r^{-3}-dependence of the electric field from a dipole.

5.4.2 van der Waals forces

The van der Waals forces are now deduced as follows. If two charge neutral hydrogen atoms are placed at a distance r from one another they can gain energy by spontaneously generate dipole moments. Consider for example what happens if atom no. 1 by quantum fluctuations generates a dipole moment d_1. According to Eq. (5.14) this results in an electric field $\mathcal{E}_1$ at the site of atom no. 2. But Eq. (5.11) tells us that atom no. 2 as a consequence will generate a dipole moment ed_2,

$$ed_2 = \frac{e^2}{K}\,\mathcal{E}_1, \quad (5.15)$$

and thus, according to Eqs. (5.12) and (5.14), gain the energy U

$$U(r) = -ed_2\,\mathcal{E}_1(r) = -\frac{e^2}{K}\Big(\mathcal{E}_1(r)\Big)^2 = -\frac{e^2}{K}\Big(\frac{2d_1 e}{4\pi\epsilon_0}\,\frac{1}{r^3}\Big)^2 = -\Big(\frac{e^2}{4\pi\epsilon_0}\Big)^2\frac{(2d_1)^2}{K}\,\frac{1}{r^6}. \quad (5.16)$$

We can make the following simple estimate for $U(r)$. First we note that the only length scale relevant for the charge displacement $2d_1$ is the size of the atom, a_0. Thus we take

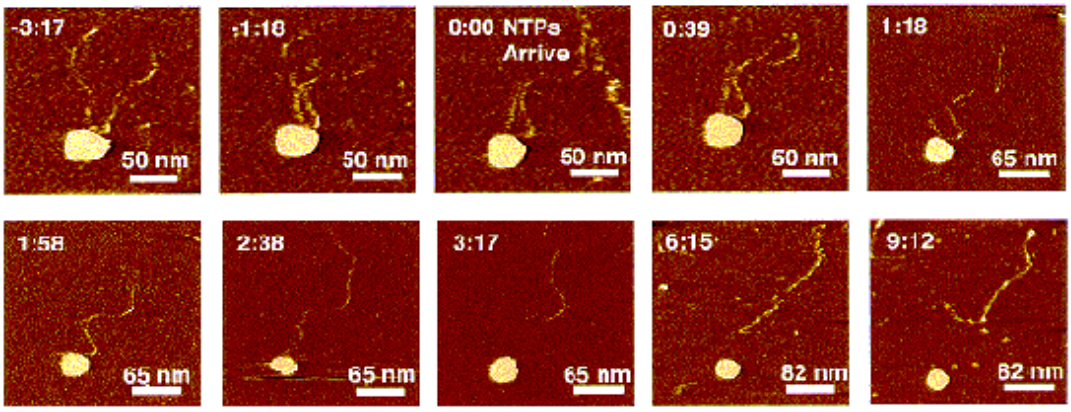


Figure 5.5: An AFM picture of an ongoing RNA polymerase. A string of DNA is transcribed by the enzyme RNA (white spot). The RNA binds to the DNA (thin line). In panel (3) a NTP molecule arrives an activates the RNA, which begins to move along the DNA string. In panel (8) the RNA reaches the end of the DNA string and falls off.

$2d_1 \approx a_0$. With the usual abreviation $E_0 = E(a_0)$ we have $2E_0 = e^2/(4\pi\epsilon_0 a_0)$, and finally use Eq. (5.10) and write $K = 2E_0/({a_0}^2)$. Thus we arrive at

$$U(r) \approx -\Big(2E_0 a_0\Big)^2 \frac{(a_0)^2}{2E_0/(a_0^2)} \frac{1}{r^6} = -2E_0 \Big(\frac{a_0}{r}\Big)^6. \tag{5.17}$$

Since force is the derivate of potential energy we obtain the following expression for the van der Waals force $F_{\rm vdW}(r)$:

$$F_{\rm vdW}(r) = -\frac{dU(r)}{dr} = 12\,\frac{E_0}{a_0}\,\Big(\frac{a_0}{r}\Big)^7. \tag{5.18}$$

To see a typical number we insert an atomic diameter $r = 2a_0$ and find

$$F_{\rm vdW}(2a_0) = 5.8 \text{ nN}. \tag{5.19}$$

5.5 Tapping mode

The tapping mode combines the good sides of the contact and the non-contact mode of AFM-operation. The AFM-cantilever is forced to oscillate with an amplitude of the order 50 nm in such a way that the tip gently touches the sample once during each cycle. The benefits are threefold. First, the tapping mode has nearly the same resolution as the contact mode. Second, the tapping mode is almost as gentle, so it can be used for soft samples, see Fig. 5.5. Third, the tapping mode is in fact a simple and roboust way of operating the AFM.

Chapter 6

Transport in nanostructures

The ability to transport charge (electric current), spin (magnetic current), and energy (heat, sound) through nanostructures is central to both scientific studies and technological applications. The quantum nature of matter on the nanometer scale leads to new transport phenomena and necessitates the development of new theoretical descriptions. In this chapter we focus on the electrical transport properties and treat some of the most important new quantum phenomena in nanostructures: electron waves, tunneling, quantized conductance, and Coulomb blockade.

6.1 Nanostructures connected to electron reservoirs

In Fig. 6.1 is shown a standard set-up for measuring the transport properties of a nano-

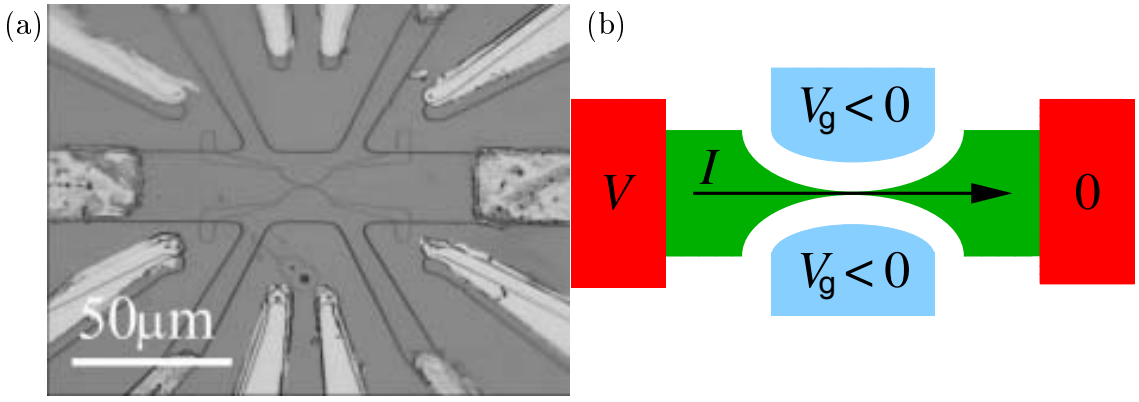


Figure 6.1: (a) A nanostructure (a quantum point contact) etched in the surface of a GaAs-GaAlAs heterostructure. Current is passed through gold wires attached to gold contacts that are alloyed into the heterostructure. (b) A sketch of the main elements in the set-up. The electrical current I is fed through the nanostructure from metal contacts (electron reservoirs) biased with some voltage V. The nanostructure may be influenced by a voltage V_g applied to nearby gate electrodes.

structure. The electrons that flow through the nanostructure is supplied from macroscopic metal contacts. The energy and electron number of the large reservoirs are not affected by the presence of the nanostructure. It is thus natural to denote the metal contacts as electron reservoirs.

In the reservoirs the electrons behave more or less as classical particles. This is mainly due to the large volume $\mathcal{V}$ of the reservoirs, since a large $\mathcal{V}$ leads to a vanishingly small energy separation $\Delta\varepsilon$ between the quantum states in the reservoir, $\Delta\varepsilon \approx \hbar^2/2m\mathcal{V}^{\frac{2}{3}}$. Thus given even the slightest perturbation the electrons will easily and frequently jump back and forth between quantum states whereby the distinct quantum nature is washed out. A given reservoir is characterized by its temperature T and its chemical potential μ, i.e. the free energy of the lastly added electron. The chemical potential can be changed in a controlled manner by applying a voltage V:

$$\mu = \mu_0 - eV. \tag{6.1}$$

The probability n_{F} that a given quantum state with energy ε is occupied by an electron is determined by the Fermi-Dirac distribution as described in Secs. 2.9.3 and 3.6:

$$n_{\mathrm{F}}(\varepsilon, \mu_0, V, T) = \frac{1}{\exp\left[\frac{\varepsilon - \mu_0 + eV}{k_{\mathrm{B}}T}\right] + 1}\,. \tag{6.2}$$

In the nanostructure between the reservoirs the electrons must however be described using quantum physics. This is due to the small transverse size w of the nanostructure which leads to appreciable energy gaps $\Delta\varepsilon$ given by $\Delta\varepsilon \approx h^2/2mw^2$. It now requires a substantial perturbation to force an electron to jump from one quantum state to the next. It is therefore very probable that a given electron stays in the same quantum state while passing through the nanostructure. The quantum properties are therefore not washed out and it can therefore be concluded that although the electrons are fed into the nanostructure as classical particles from a large reservoir the transport through the nanostructure must nevertheless be described in terms of the electron wave picture of quantum theory.

6.2 Current density and transmission of electron waves

We begin our analysis of electric conductance of nanostructures by studying simple 1D waves in piecewise constant potentials. We discuss the electron waves, then their current density, and finally the concept of transmission and reflection coefficients $\mathcal{T}$ and $\mathcal{R}$.

6.2.1 Electron waves in constant potentials in 1D

For an electron with energy ε moving in a given constant potential V the solution $\psi_\varepsilon(x)$ of the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\,\frac{\partial^2\psi_\varepsilon(x)}{\partial x^2} + V\,\psi_\varepsilon(x) = E\,\psi_\varepsilon(x) \tag{6.3}$$

is given by

$$\psi_\varepsilon(x) = e^{\pm ikx}, \quad k = \frac{1}{\hbar}\sqrt{2m(\varepsilon - V)}. \tag{6.4}$$

Note that $\varepsilon = \frac{\hbar^2 k^2}{2m} + V$. The two possible signs '+' and '−' correspond to waves travelling to the right and to the left, respectively. This can be seen by writing out the full time-dependent solutions $\psi_{+k}(x,t)$ and $\psi_{-k}(x,t)$ as given in Eq. (2.30):

$$\psi_{+k}(x,t) = e^{i(+kx - \frac{\varepsilon}{\hbar}t)}, \quad \text{(right-moving wave)}, \tag{6.5a}$$

$$\psi_{-k}(x,t) = e^{i(-kx - \frac{\varepsilon}{\hbar}t)}, \quad \text{(left-moving wave)}. \tag{6.5b}$$

In the following we suppress the trivial time-dependence $\exp(-i\frac{\varepsilon}{\hbar}t)$ and work only with the time-independent Schrödinger eqaution and the time-independent solutions $\psi_{\pm k}(x)$. But of course the concept of left- and right-moving waves remains the same:

$$\psi_{+k}(x) = e^{+ikx}, \quad \text{(right-moving wave)}, \tag{6.6a}$$

$$\psi_{-k}(x) = e^{-ikx}, \quad \text{(left-moving wave)}. \tag{6.6b}$$

6.2.2 The current density J

The classical expression for current density is $\mathbf{J} = n\,\mathbf{v} = n\,\mathbf{p}/m$, n being the density of particles. In quantum theory the density of a single particle is $n = \psi^*\psi$, and a naïve guess for the quantum expression for $\mathbf{J}$ would thus be $\mathbf{J} = \psi^*\psi\,\mathbf{p}/m$. However, as argued in Sec. 2.8.2 the true result is slightly more complicated owing to the fact that in quantum theory $\hat{\mathbf{p}}$ is a differential operator and not a number. Moreover, we need to ensure that $\mathbf{J}$ is real, so as stated in Eq. (2.45) we end up with

$$\mathbf{J}(\mathbf{r}) = \frac{1}{m}\,\mathrm{Re}\Big[\psi^*(\hat{\mathbf{p}}\psi)\Big] = \frac{\hbar}{m}\,\mathrm{Im}\Big[\psi^*(\boldsymbol{\nabla}\psi)\Big]. \tag{6.7}$$

The electric current I passing through the yz plane in the x direction is given by

$$I = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\,(-e)\mathbf{J}_x = -\frac{e\hbar}{m}\int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\,\mathrm{Im}\Big[\psi^*\Big(\frac{\partial\psi}{\partial x}\Big)\Big]. \tag{6.8}$$

In the case of simple product wavefunctions described in terms of some quantum numbers k in the x direction and α in the yz directions,

$$\psi_{k\alpha}(\mathbf{r}) = \chi_k(x)\,\phi_\alpha(y,z), \tag{6.9}$$

we find that the current $I_{k\alpha}$ carried by the quantum state $\psi_{k\alpha}$ is

$$I_{k\alpha} = -\frac{e\hbar}{m}\,\mathrm{Im}\Big[\chi_k(x)^*\,\frac{\partial}{\partial x}\big(\chi_k(x)\big)\Big]. \tag{6.10}$$

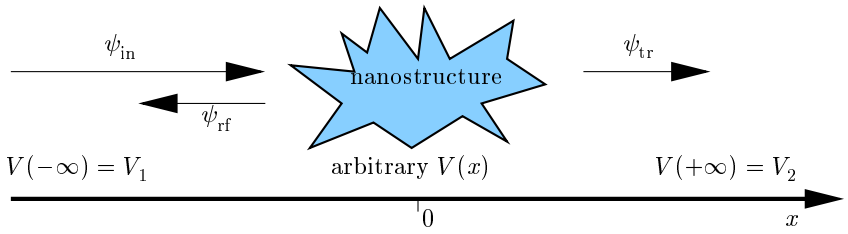


Figure 6.2: Transport through a nanostructure formulated as an electron-wave scattering problem. The incoming wave $\psi_{\rm in}$ is either reflected as $\psi_{\rm rf}$ or transmitted as $\psi_{\rm tr}$. Far away from the nanostructure the potential is constant, V_1 to the left and V_2 to the right. At the nanostructure it is arbitrary.

6.2.3 The transmission and reflection coefficients $\mathcal{T}$ and $\mathcal{R}$ in 1D

We now focus on 1D transport which is relevant if there is complete translation invariance in the yz plane or if there is only one quantum state $\phi_\alpha(y,z)$ available for the yz part of the electron motion. The generic 1D scattering set-up is shown in Fig. 6.2. We imagine that far to the left we have a constant potential V_1 and similarly another constant potential V_2 far to the right. Consider an incoming wave $\psi_{\rm in}(x)$ at energy $\varepsilon > V_1, V_2$ from the far left. According to Eq. (6.4) we have

$$\psi_{\rm in}(x) = \frac{1}{\sqrt{L}}\, e^{ik_1 x}, \quad k_1 = \frac{1}{\hbar}\sqrt{2m(\varepsilon - V_1)}, \tag{6.11}$$

where we now have inserted the normalization factor $1/\sqrt{L}$ taking the length L of the system into account. When this wave hits the scattering potential two things can happen: either it is reflected with some amplitude B back to the left in the state $\psi_{\rm rf}(x)$ given by

$$\psi_{\rm rf}(x) = \frac{B}{\sqrt{L}}\, e^{-ik_1 x}, \quad k_1 = \frac{1}{\hbar}\sqrt{2m(\varepsilon - V_1)}, \tag{6.12}$$

or it is transmitted with some amplitude C to the right in the state $\psi_{\rm tr}(x)$ given by

$$\psi_{\rm tr}(x) = \frac{C}{\sqrt{L}}\, e^{ik_2 x}, \quad k_2 = \frac{1}{\hbar}\sqrt{2m(\varepsilon - V_2)}. \tag{6.13}$$

Each of these states carries a current. From Eq. (6.10) we obtain

$$I_{\rm in} = -\frac{e\hbar}{mL}\,{\rm Im}\,[ik_1] \qquad = -\frac{e\hbar}{mL}\,k_1, \tag{6.14a}$$

$$I_{\rm rf} = -\frac{e\hbar}{mL}\,{\rm Im}\,[ik_1]\,|B|^2 = -\frac{e\hbar}{mL}\,k_1\,|B|^2, \tag{6.14b}$$

$$I_{\rm tr} = -\frac{e\hbar}{mL}\,{\rm Im}\,[ik_2]\,|C|^2 = -\frac{e\hbar}{mL}\,k_2\,|C|^2. \tag{6.14c}$$

In steady-state charge conservation requires all of the incoming current to be either reflected or tranmitted. Charge is not accumulated anywhere. It is therefore natural to

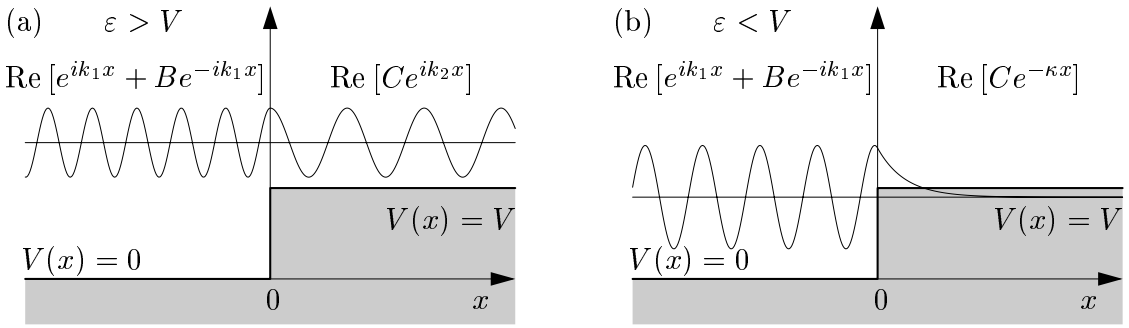


Figure 6.3: A potential step of size V. (a) An electron wave with energy $\varepsilon > V$, and (b) with energy $\varepsilon < V$. Note the exponentially decaying tunneling into the potential barrier.

define transmission and reflection coefficients $\mathcal{T}$ and $\mathcal{R}$ as follows:

$$\mathcal{T} = \frac{|I_{\rm tr}|}{|I_{\rm in}|} = \frac{k_2}{k_1}\,|C|^2, \tag{6.15a}$$

$$\mathcal{R} = \frac{|I_{\rm rf}|}{|I_{\rm in}|} = |B|^2. \tag{6.15b}$$

In terms of $\mathcal{T}$ and $\mathcal{R}$ charge conservation takes the form

$$\mathcal{T} + \mathcal{R} = 1. \tag{6.16}$$

6.3 Electron waves and the simple potential step

As the first example of treating transport with electron waves we treat the simple 1D potential step depicted in Fig. 6.3 and given by

$$V(x) = \begin{cases} 0, & \text{for } x < 0, \\ V, & \text{for } x > 0. \end{cases} \tag{6.17}$$

Consider first the situation shown in Fig. 6.3a where the energy ε of the electron is larger than the potential step V. Classically, we would expect full transmission when $\varepsilon > V$, but this is not so in quantum physics. Following Sec. 6.2.3 we let a wave come in from the left and scatter on the potential step. The wave function takes the form

$$\psi_\varepsilon(x) = \begin{cases} e^{ik_1x} + Be^{-ik_1x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} & \text{for } x < 0, \\ Ce^{ik_2x}, & k_2 = \dfrac{1}{\hbar}\sqrt{2m(\varepsilon - V)} & \text{for } x > 0. \end{cases} \tag{6.18}$$

The coefficients B and C are found by demanding continuity of $\psi_\varepsilon(x)$ and $\psi'_\varepsilon(x)$ at $x = 0$:

$$\psi_\varepsilon(0^-) = \psi_\varepsilon(0^+) \qquad \Rightarrow \qquad 1 + B = C, \tag{6.19a}$$

$$\psi'_\varepsilon(0^-) = \psi'_\varepsilon(0^+) \qquad \Rightarrow \qquad ik_1(1 - B) = ik_2C. \tag{6.19b}$$

From this follows

$$B = \frac{k_1 - k_2}{k_1 + k_2}, \qquad \text{and} \qquad C = \frac{2k_1}{k_1 + k_2}. \tag{6.20}$$

From Eqs. (6.15a) and (6.15b) we then calculate the transmission and reflection coefficient $\mathcal{T}$ and $\mathcal{R}$:

$$\mathcal{T} = \frac{k_2}{k_1} |C|^2 = \frac{4k_1 k_2}{(k_1 + k_2)^2}, \tag{6.21a}$$

$$\mathcal{R} = |B|^2 = \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2}. \tag{6.21b}$$

We note that $\mathcal{T} + \mathcal{R} = 1$ as required by the charge conservation Eq. (6.16). But we also note that even though $E > V$ we have $\mathcal{T} < 1$, i.e. less than full transmission. Only in the limit of high energies $\varepsilon \gg V$, where $k_2/k_1 \to 1$, do we obtain $\mathcal{T} \to 1$.

Consider then the situation of Fig. 6.3b where the energy ε of the electron is less than the potential step V. Classically, we would expect full reflection when $\varepsilon < V$, and this we also obtain in quantum physics. The wave function now takes the form

$$\psi_\varepsilon(x) = \begin{cases} e^{ik_1 x} + Be^{-ik_1 x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} \qquad \text{for } x < 0, \\ Ce^{-\kappa x}, & \kappa = \dfrac{1}{\hbar}\sqrt{2m(V-\varepsilon)} \quad \text{for } x > 0. \end{cases} \tag{6.22}$$

Note how k_2 now is written as $i\kappa$ due to the sign change under the square root. As before the coefficients B and C are found by demanding continuity of $\psi_\varepsilon(x)$ and $\psi'_\varepsilon(x)$ at $x = 0$:

$$\psi_\varepsilon(0^-) = \psi_\varepsilon(0^+) \qquad \Rightarrow \qquad 1 + B = C, \tag{6.23}$$

$$\psi'_\varepsilon(0^-) = \psi'_\varepsilon(0^+) \qquad \Rightarrow \qquad ik_1(1 - B) = -\kappa C. \tag{6.24}$$

From this follows

$$B = \frac{k_1 - i\kappa}{k_1 + i\kappa}, \qquad \text{and} \qquad C = \frac{2k_1}{k_1 + i\kappa}. \tag{6.25}$$

Since in this case $k_2 = i\kappa$ we find from Eq. (6.14c) that $I_{\text{tr}} \propto \text{Im}\,[ik_2] = \text{Im}\,[-\kappa] = 0$ and consequently

$$\mathcal{T} = \frac{|I_{\text{tr}}|}{|I_{\text{in}}|} = 0. \tag{6.26}$$

For the reflection coefficient we can use Eq. (6.15b):

$$\mathcal{R} = |B|^2 = \frac{|k_1 - i\kappa|^2}{|k_1 + i\kappa|^2} = 1. \tag{6.27}$$

We note that again $\mathcal{T} + \mathcal{R} = 1$ as required by charge conservation. We further note that as in the classical case we have total reflection when $\varepsilon < V$. Finally, we see that in spite of the total reflection there is in fact a finite probability to find the electron inside the potential barrier, a probability that is decaying exponentially with the distance from the potential step. This is a manifestation of the so-called quantum mechanical tunneling effect, an effect we study further in the following section.

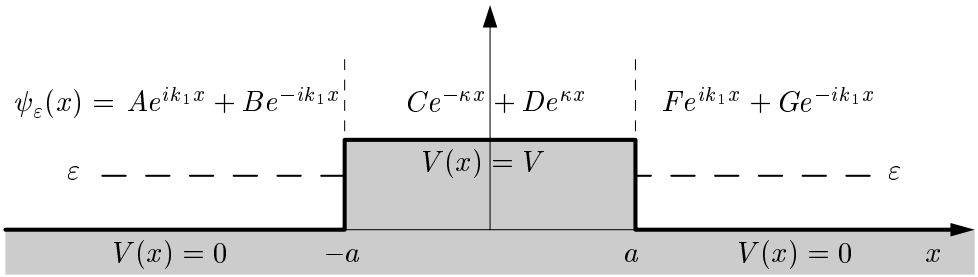


Figure 6.4: A rectangular potential barrier of height V and width $2a$. The general form in each of the three potential regions of a wavefunction $\psi_\varepsilon(x)$ at energy $\varepsilon < V$ is also given.

6.4 Tunneling through a potential barrier

Consider a rectangular potential barrier as the one sketched in Fig. 6.4. The barrier is defined by

$$V(x) = \begin{cases} 0, & \text{for } x < -a, \\ V, & \text{for } -a < x < a, \\ 0, & \text{for } x > a. \end{cases} \tag{6.28}$$

We focus our attention on the general form of an eigenstate $\psi_\varepsilon(x)$ at the sub-barrier energy $\varepsilon < V$. As a straightforward generalization of Eq. (6.22) we write

$$\psi_\varepsilon(x) = \begin{cases} Ae^{ik_1x} + Be^{-ik_1x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} & \text{for } x < -a, \\ Ce^{-\kappa x} + De^{\kappa x}, & \kappa = \dfrac{1}{\hbar}\sqrt{2m(V-\varepsilon)} & \text{for } -a < x < a, \\ Fe^{ik_1x} + Ge^{-ik_1x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} & \text{for } a < x. \end{cases} \tag{6.29}$$

In analogy with Eqs. (6.23) and (6.24) we demand continuity of $\psi_\varepsilon(x)$ and $\psi'_\varepsilon(x)$ at $x = -a$. We write the two resulting equations as one matrix equation:

$$\begin{pmatrix} e^{-ik_1a} & e^{ik_1a} \\ ik_1e^{-ik_1a} & -ik_1e^{ik_1a} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} e^{\kappa a} & e^{-\kappa a} \\ -\kappa e^{\kappa a} & \kappa e^{-\kappa a} \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix}. \tag{6.30}$$

We then multiply to the left with the inverse of the left matrix and get

$$\begin{pmatrix} A \\ B \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (1 + \frac{i\kappa}{k_1})e^{\kappa a + ik_1a} & (1 - \frac{i\kappa}{k_1})e^{-\kappa a + ik_1a} \\ (1 - \frac{i\kappa}{k_1})e^{\kappa a - ik_1a} & (1 + \frac{i\kappa}{k_1})e^{-\kappa a - ik_1a} \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix}. \tag{6.31}$$

In the same way continuity of $\psi_\varepsilon(x)$ and $\psi'_\varepsilon(x)$ at the other potential edge $x = a$ leads to a matrix equation for (C, D) and (F, G):

$$\begin{pmatrix} C \\ D \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (1 - \frac{ik_1}{\kappa})e^{\kappa a + ik_1a} & (1 + \frac{ik_1}{\kappa})e^{\kappa a - ik_1a} \\ (1 + \frac{ik_1}{\kappa})e^{-\kappa a + ik_1a} & (1 - \frac{ik_1}{\kappa})e^{-\kappa a - ik_1a} \end{pmatrix} \begin{pmatrix} F \\ G \end{pmatrix}. \tag{6.32}$$

Finally, by inserting Eq. (6.32) into Eq. (6.31) and multiplying the two matrices we obtain a relation between the wavefunction coefficients to the left and to the right of the barrier:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \big(\cosh 2\kappa a + \frac{i\gamma}{2}\sinh 2\kappa a\big)e^{i2k_1 a} & \frac{i\eta}{2}\sinh 2\kappa a \\ -\frac{i\eta}{2}\sinh 2\kappa a & \big(\cosh 2\kappa a - \frac{i\gamma}{2}\sinh 2\kappa a\big)e^{-i2k_1 a} \end{pmatrix} \begin{pmatrix} F \\ G \end{pmatrix}, \tag{6.33}$$

where we have introduced the abbreviations

$$\gamma = \frac{\kappa}{k_1} - \frac{k_1}{\kappa}, \qquad \text{and} \qquad \eta = \frac{\kappa}{k_1} + \frac{k_1}{\kappa}. \tag{6.34}$$

From Eq. (6.33) we are able to calculate the transmission and reflection coefficients $\mathcal{T}$ and $\mathcal{R}$. We simply let Fig. 6.4 correspond to the scattering set-up sketched in Fig. 6.2, i.e. we let a wave of amplitude $A = 1$ come in from the left and obtain a reflected and a transmitted wave of amplitude B for $x < -a$ and F for $x > a$, respectively. In this situation no wave comes in from the right towards the barrier, i.e. $G = 0$. With $A = 1$ and $G = 0$ Eq. (6.33) leads to

$$1 = F\,\big(\cosh 2\kappa a + \frac{i\gamma}{2}\sinh 2\kappa a\big)e^{i2k_1 a}, \tag{6.35a}$$

$$B = -F\,\frac{i\eta}{2}\sinh 2\kappa a, \tag{6.35b}$$

and therefore

$$F = \frac{e^{-i2k_1 a}}{\cosh(2\kappa a) + \frac{i\gamma}{2}\sinh(2\kappa a)}, \tag{6.36a}$$

$$B = \frac{-\frac{i\eta}{2}\sinh(2\kappa a)\; e^{-i2k_1 a}}{\cosh(2\kappa a) + \frac{i\gamma}{2}\sinh(2\kappa a)}. \tag{6.36b}$$

From this and using Eqs. (6.15a) and (6.15b) with $k_2 = k_1$ and with F instead of C we obtain the transmission and reflection coefficients:

$$\mathcal{T} = |F|^2 = \frac{1}{\cosh^2(2\kappa a) + \frac{\gamma^2}{4}\sinh^2(2\kappa a)} = \frac{1}{1 + \frac{\eta^2}{4}\sinh^2(2\kappa a)} \tag{6.37a}$$

$$\mathcal{R} = |B|^2 = \frac{\frac{\eta^2}{4}\sinh^2(2\kappa a)}{\cosh^2(2\kappa a) + \frac{\gamma^2}{4}\sinh^2(2\kappa a)} = \frac{\frac{\eta^2}{4}\sinh^2(2\kappa a)}{1 + \frac{\eta^2}{4}\sinh^2(2\kappa a)}. \tag{6.37b}$$

We see that the charge conservation condition $\mathcal{R} + \mathcal{T} = 1$ is fulfilled.

6.4.1 Transmission below the barrier

The expression Eq. (6.36a) for the transmission amplitude F, combined with $G = 0$ and the matrix equation Eq. (6.32), gives us a direct way to calculate the amplitudes C and D

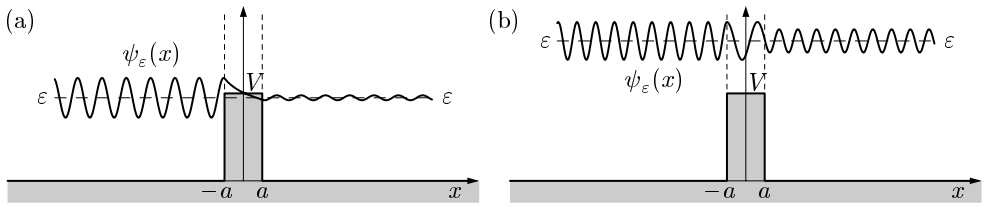


Figure 6.5: Explicit examples of electron waves with eigenenergies ε in the potential forming a rectangular barrier of height V and width $2a$. (a) An electron wave with energy $\varepsilon < V$ coming from the left is partially reflected and partially transmitted by a tunneling proces. Note the exponential decay in the barrier region. (b) An electron wave with energy $\varepsilon > V$ coming from the left. Again partial reflection and transmission is observed, but no tunneling is involved. Note the change of wavelength in the barrier region.

in the barrier region. We thus have expressions for all amplitudes, and hence by Eq. (6.29) we can calculate the wavefunction $\psi_\varepsilon(x)$. An example of such a calculation in shown in Fig. 6.5a. It is clearly demonstrated how a tiny portion of the incoming wave tunnels through the barrier and continue its propagation on the other side of the barrier. Note the exponential decay of the amplitude in the barrier region.

An explicit expression for $\mathcal{T}$ can be derived by first obtaining η^2 from Eqs. (6.29) and (6.34),

$$\eta^2 = \Big(\frac{k_1}{\kappa} + \frac{\kappa}{k_1}\Big)^2 = \frac{k_1^2}{\kappa^2} + \frac{\kappa^2}{k_1^2} + 2 = \frac{\varepsilon}{V-\varepsilon} + \frac{V-\varepsilon}{\varepsilon} + 2 = \frac{V^2}{\varepsilon(V-\varepsilon)}. \tag{6.38}$$

so that

$$\mathcal{T}(\varepsilon) = \Big(1 + \frac{V^2}{4\varepsilon(V-\varepsilon)}\,\sinh^2(2\kappa a)\Big)^{-1}, \qquad \text{valid for } \varepsilon < V. \tag{6.39}$$

Eq. (6.39) gives the transmission coefficient $\mathcal{T}(\varepsilon)$ for all electron energies $\varepsilon < V$. In the limit of a high, $V \gg \varepsilon$, and wide, $\kappa a \gg 1$, barrier we can simplify the expression. Since $\sinh(x) = \frac{1}{2}[\exp(x) - \exp(-x)] \to \frac{1}{2}\exp(x)$, for $x \to \infty$, we find

$$\mathcal{T}(\varepsilon) \approx 16\,\frac{\varepsilon}{V}\,e^{-4\kappa a} \approx 16\,\frac{\varepsilon}{V}\,e^{-4\sqrt{V/E_a}}, \quad \text{for } \varepsilon \ll V \text{ and } 1 \ll \kappa a. \tag{6.40}$$

In the second quasi-equality we have used the abreviation $E_a = \hbar^2/2ma^2$. The expression reveals that the transmission coefficient decays exponentially with increasing barrier thickness a and with the square root of increasing barrier height, $\sqrt{V}$.

This result is truely remarkable. It tells us that if we take the mathematical formulation of quantum theory at face value, then we are forced to accept that particles can tunnel through potential barriers with a non-zero probability, a process that classical physics would rule out completely. It is an experimental fact that tunneling processes do occur in Nature. Let us mention a few examples of phenomena where tunneling is essential:

(1) The α-decay of heavy nuclei is explained in terms of α-particles that tunnel out from the attractive nuclear force potential through the Coulomb-potential barrier.

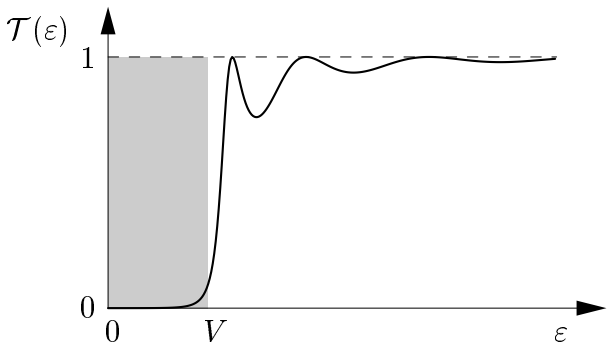


Figure 6.6: The transmission coefficient $\mathcal{T}(\varepsilon)$ for a rectangular potential barrier of height V and width $a = \sqrt{10\,\hbar^2/2mV}$. For electron energies below the barrier, $\varepsilon < V$, we see the exponentially suppressed transmission. Above the barrier, $\varepsilon > V$, the transmission approaches full transmission in a transient manner. Several resonance peaks are clearly seen; they are broadened and tend to vanish as the energy increases.

(2) The ionization of atoms in external electric fields weaker than the break-down field is explained as electrons tunneling out through the Coulomb potential barrier to the region of lower potential created by the external field.

(3) Modern, high precision voltage standards are based on the Josephson effect where pairs of electrons tunnel from one superconductor into another through an insulating oxide layer.

(4) The scanning tunneling microscope, a key instrument in nanotechnolgy, is based on single-electrons tunneling from electron orbitals of the sample into the metal tip of the probe.

6.4.2 Transmission above the barrier

To finish the discussion of the rectangular potential barrier we now turn to the transmission for energies higher than the barrier, $\varepsilon > V$. In this situation the wavefunction $\psi_\varepsilon(x)$ from Eq. (6.29) becomes

$$\psi_\varepsilon(x) = \begin{cases} Ae^{ik_1 x} + Be^{-ik_1 x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} & \text{for } x < -a, \\ Ce^{ik_2 x} + De^{-ik_2 x}, & k_2 = \dfrac{1}{\hbar}\sqrt{2m(\varepsilon - V)} & \text{for } -a < x < a, \\ Fe^{ik_1 x} + Ge^{-ik_1 x}, & k_1 = \dfrac{1}{\hbar}\sqrt{2m\varepsilon} & \text{for } a < x. \end{cases} \tag{6.41}$$

The exponential terms $\exp(\mp\kappa x)$ are now oscillatory $\exp(\pm ik_2 x)$, i.e. $k_2 = i\kappa$. We could now go through the same calculations for above-barrier transmission as for below-barrier transmission, i.e. match the wavefunction and its derivative at $x = \pm a$, to find the transmission coefficient $\mathcal{T}(\varepsilon)$. Instead we will simply use the substitution $\kappa \to -ik_2$ in Eq. (6.39)

and the identity $\sinh(ix) = \frac{1}{2}[\exp(ix) - \exp(-ix)] = i\sin(x)$ to obtain

$$\mathcal{T}(\varepsilon) = \left(1 + \frac{V^2}{4\varepsilon(\varepsilon - V)} \sin^2(2k_2 a)\right)^{-1}, \qquad \text{valid for } \varepsilon > V. \tag{6.42}$$

6.4.3 The complete transmission function $\mathcal{T}(\varepsilon)$

Taken together Eqs. (6.39) and (6.42) give the complete transmission function $\mathcal{T}(\varepsilon)$ for any positive vaule of ε. A specific graph of $\mathcal{T}(\varepsilon)$ is shown in Fig. 6.6. The exponential suppression of the transmission is clearly seen in the gray barrier region $\varepsilon < V$.

As ε crosses over V and the above-barrier region is entered, the transmission starts to rise steaply. We note that $\mathcal{T}(\varepsilon) \to 1$ for $\varepsilon \to \infty$, i.e. full transmission as expected for $\varepsilon \gg V$. We also note that for values of ε only slightly larger than V we generally do not obtain $\mathcal{T}(\varepsilon) = 1$; there is above-barrier reflection. However, when the resonance condition $2k_2 a = n\pi$, where $n = 1, 2, 3, \ldots$, is fulfilled peaks reaching full transmission appear.

The graph of $\mathcal{T}(\varepsilon)$ demonstrates that one can encounter highly non-trivial transmission functions in quantum physics. The resonance structure can be expected each time sharp potential drops or well-defined quantum states, e.g. molecular orbitals, are present in the nanostructure. In other cases the transmission coefficient rises more monotonically from zero to unity.

6.5 Transfer and scattering matrices

The matrix equations Eqs. (6.31) and (6.32) are examples of the so-called transfer matrix method. In 1D one can collect the wave amplitudes to left, say (A_{N-1}, B_{N-1}), and those to the right, say (A_N, B_N), and then connect them by a matrix, the transfer matrix $\mathbf{M}_{N-1,N}$,

$$\begin{pmatrix} A_{N-1} \\ B_{N-1} \end{pmatrix} = \mathbf{M}_{N-1,N} \begin{pmatrix} A_N \\ B_N \end{pmatrix}. \tag{6.43}$$

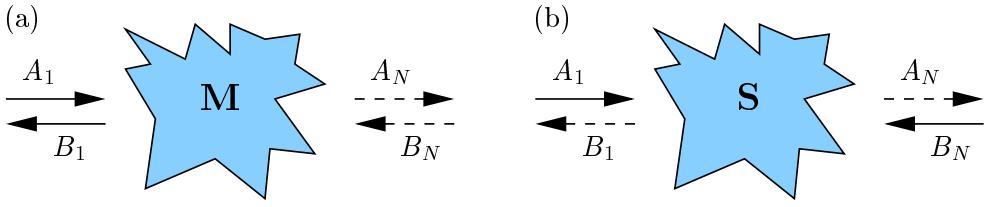


Figure 6.7: Two possible ways to connect the wave amplitudes A_1, B_1, A_N, and B_N. (a) In 1D the transfer matrix $\mathbf{M}$ is often used to connect the amplitudes to the left, A_1 and B_1, (full arrows) with those to the right, A_N and B_N, (dashed arrows). (b) In more general cases it is more convenient to use the scattering matrix $\mathbf{S}$ that connects the incoming amplitudes A_1 and B_N (full arrows) with the outgoing ones B_1 and A_N (dashed arrows).

For a 1D system consisting of N falt potential regions, one can find the connection between the amplitudes of the first and the last region simply by succesively apply the transfer matrices between neighboring regions. The result is

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \mathbf{M}_{1,2}\mathbf{M}_{2,3}\ldots\mathbf{M}_{N-2,N-1}\mathbf{M}_{N-1,N}\begin{pmatrix} A_N \\ B_N \end{pmatrix} = \mathbf{M}\begin{pmatrix} A_N \\ B_N \end{pmatrix}, \tag{6.44}$$

where $\mathbf{M}$ is the total transfer matrix of the system. A useful application of this formula is the following algorithm for calculating the transmission coefficient for any arbitrarily shaped potential. The given potential is approximated to any desired accuracy by N piecewise constant potential regions. We then follow Fig. 6.2 and connect a wave of the type $(1, B_1)$ to the left with $(A_N, 0)$ to the right,

$$\begin{pmatrix} 1 \\ B_1 \end{pmatrix} = \mathbf{M}\begin{pmatrix} A_N \\ 0 \end{pmatrix}. \tag{6.45}$$

According to Eq. (6.15a) $\mathcal{T}$ is, in terms of the 1,1–component $\mathbf{M}_{11}$ of the transfer matrix,

$$\mathcal{T} = \frac{k_N}{k_1}|A_N|^2 = \frac{k_N}{k_1}\frac{1}{\mathbf{M}_{11}}. \tag{6.46}$$

For completeness we end this section by noting that in the most general scattering geometries, e.g. in higher dimensions than 1, it is often more convenient to work with the so-called scattering matrix, or simply S-matrix, instead of the transfer matrix. The scattering matrix relates the outgoing amplitudes (A_N, B_1) with the incoming amplitudes (A_1, B_N),

$$\begin{pmatrix} A_N \\ B_1 \end{pmatrix} = \mathbf{S}\begin{pmatrix} A_1 \\ B_N \end{pmatrix}. \tag{6.47}$$

The S-matrix is useful because it is more directly related to fundamental concepts like conservation of charge and time-reversal symmetry. And in contrast to the transfer matrix the scattering matrix can alway be defined in any scattering set-up. The scattering matrix is also fundamental in formulating conductance properties of quantum systems such as nanostructures. This is the topic of the following section.

6.6 Conductance and scattering matrix formalism

We now return to the problem of finding the conductance $G = I/V$ of a generic nanostructure such as the one sketched in Fig. 6.1b. We write the eigenstates of the electrons in the straight part of the sample between the nanostructure and the electron reservoirs in the form given by Eq. (6.9)

$$\psi_{k\alpha}(\mathbf{r}) = \frac{1}{\sqrt{L}}e^{ikx}\,\phi_\alpha(y,z). \tag{6.48}$$

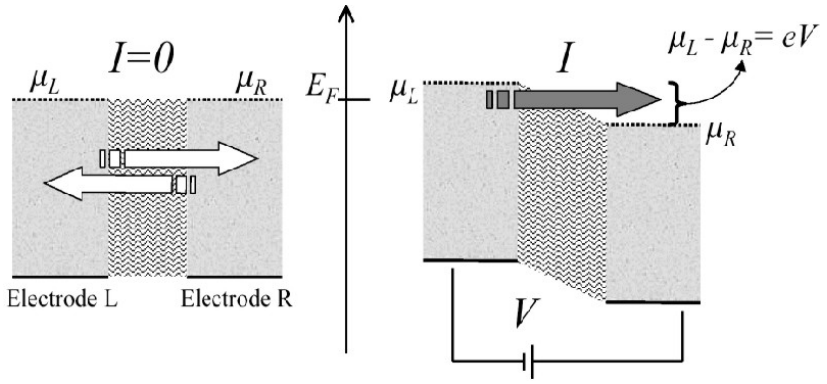


Figure 6.8: Electrical currents through a nanostructure at zero temperature with injection of electrons from a left and a right electron reservoir. Without an applied bias voltage V we have $\mu_L = \mu_R$, and the left- and right-going currents cancel each other yielding a zero net current. For $V > 0$ we have $\mu_L = \mu_R + eV$ and the current carried by right-going electrons with energies larger than $\mu_L - eV$ and smaller than μ_L is not compensated, thus resulting in a non-zero net current.

6.6.1 Electron channels

For a given quantum number α there are many possible quantum states, namely two (spin up and spin down) for each k. One therefore talks about the electron channel α. The total energy of a state in a given channel α is denoted $\varepsilon_{k\alpha}$. It consists of two contributions: one denoted ε_α is the quantum energy of the transverse wavefunction $\phi_\alpha(y,z)$, while the other is the usual kinetic energy $\varepsilon_k = \hbar^2 k^2/2m$ associated with the motion in the x direction. From a simple energy consideration we derive an expression for k:

$$\varepsilon_{k\alpha} = \varepsilon_k + \varepsilon_\alpha \qquad \Rightarrow \qquad k = \frac{1}{\hbar}\sqrt{2m(\varepsilon_{k\alpha} - \varepsilon_\alpha)}. \tag{6.49}$$

Note that the lowest possible energy for electrons in channel α is ε_α obtained when filling the state $\psi_{0\alpha}$.

6.6.2 Current, reservoirs, and electron channels

We now analyze the current flow sketched in Fig. 6.8 starting with the current $I_{k\alpha}$ carried by an electron in state $\psi_{k\alpha}$ in channel α. It is given by Eq. (6.10), with a factor of 2 for spin inserted,

$$I_{k\alpha} = -\frac{2e\hbar}{Lm}\,\mathrm{Im}\Big[e^{-ikx}\,\frac{\partial}{\partial x}\big(e^{ikx}\big)\Big] = -\frac{2e\hbar k}{Lm} = -2\,\frac{e}{L}\,v_{k\alpha}, \tag{6.50}$$

where $v_{k\alpha}$ is the velocity of an electron in the state $\psi_{k\alpha}$.

It is now easy to derive an expression for the current $I_{L\to R}$ running from the left reservoir L into the right reservoir R through the nanostructure. We consider the energy

$\varepsilon = \varepsilon_{k\alpha}$ and multiply three factors: (1) the Fermi-Dirac probability $n_{\rm F}^L(\varepsilon)$ that the right-moving state $\psi_{k\alpha}$ is occupied from the left reservoir; (2) the current $I_{k\alpha}$ carried by that state; and (3) the transmission probability $\mathcal{T}_\alpha(\varepsilon)$ of an electron at energy ε making it through to the reservoir to the right. Then we sum over all possible electron channels α and wavenumbers k:

$$I_{L\to R} = \sum_\alpha \sum_k \big[n_{\rm F}^L(\varepsilon)\big] \times \big[-2\,\frac{e}{L}\,v_{k\alpha}\big] \times \big[\mathcal{T}_\alpha(\varepsilon)\big]. \tag{6.51}$$

First we convert the k-sum into a k-integral by use of Eq. (3.10), which in 1D reads $\sum_k \to (L/2\pi)\int dk$. Then we convert the k-integral to an ε-integral: $\int dk \to \int d\varepsilon\,\frac{\partial k}{\partial \varepsilon}$. But from Eq. (2.8b) we know that $\frac{\partial k}{\partial \varepsilon} = \frac{1}{\hbar v}$. So collecting all this yields $\sum_k \to (L/2\pi)\int d\varepsilon\,\frac{1}{\hbar v}$ and Eq. (6.51) becomes

$$I_{L\to R} = -\,\frac{2e}{h}\,\sum_\alpha \int_{-\infty}^{\infty} d\varepsilon\, n_{\rm F}^L(\varepsilon)\,\mathcal{T}_\alpha(\varepsilon). \tag{6.52}$$

Note how all references to geometry, i.e. L, and velocity have vanished from the expression.

The current $I_{R\to L}$ flowing from the right to the left is obtained in a similar way,

$$I_{R\to L} = -\,\frac{2e}{h}\,\sum_\alpha \int_{-\infty}^{\infty} d\varepsilon\, n_{\rm F}^R(\varepsilon)\,\mathcal{T}_\alpha(\varepsilon). \tag{6.53}$$

The total current I flowing in the system is therefore

$$I = I_{L\to R} - I_{R\to L} = -\,\frac{2e}{h}\,\sum_\alpha \int_{-\infty}^{\infty} d\varepsilon\,\mathcal{T}_\alpha(\varepsilon)\,\big[n_{\rm F}^L(\varepsilon) - n_{\rm F}^R(\varepsilon)\big]. \tag{6.54}$$

6.6.3 The conductance formula for nanostructures

We now have a relatively simple expression for the current, so to obtain the conductance we need to find the voltage dependence. In our set-up in Fig. 6.1b the voltage V is applied to the left reservoir. If we restrict ourselves to consider only small voltages (the linear response limit) it follows from Eq. (6.2) that

$$n_{\rm F}^L(\varepsilon) - n_{\rm F}^R(\varepsilon) = n_{\rm F}(\varepsilon,\mu_0 - eV) - n_{\rm F}(\varepsilon,\mu_0) \approx \left.\frac{n_{\rm F}}{\partial\mu}\right|_{\mu_0} (-eV) = \Big(-\frac{n_{\rm F}}{\partial\varepsilon}\Big)\,(-eV). \tag{6.55}$$

With this Eq. (6.54) becomes

$$I = \frac{2e^2}{h}\,\sum_\alpha \int_{-\infty}^{\infty} d\varepsilon\,\mathcal{T}_\alpha(\varepsilon)\,\Big(-\frac{n_{\rm F}}{\partial\varepsilon}\Big)\,V. \tag{6.56}$$

Finally, we obtain the temperature dependent conductance $G(T)$,

$$G(T) = \frac{I}{V} = \frac{2e^2}{h}\,\sum_\alpha \int_{-\infty}^{\infty} d\varepsilon\,\mathcal{T}_\alpha(\varepsilon)\,\Big(-\frac{n_{\rm F}}{\partial\varepsilon}\Big). \tag{6.57}$$

In the limit of very low temperature (temperatures around 1 K are routinely obtained in the lab cooling with liquid helium) Eqs. (3.28a) and (3.29) state that $-\frac{n_{\rm F}}{\partial\varepsilon} = \delta(\varepsilon_{\rm F} - \varepsilon)$. Inserting this in Eq. (6.56) yields the zero temperature conductance

$$G(T=0) = \frac{2e^2}{h} \sum_\alpha \mathcal{T}_\alpha(\varepsilon_{\rm F}). \tag{6.58}$$

Eq. (6.57) is the conductance formula for nanostructures. We derived it by considering scattering states, i.e. the scattering matrix formalism. We note that the integral is dimensionless, so that means that the prefactor $G_0 = e^2/h$, depending entirely on universal constants, is some kind of conductance quantum, a natural unit for measuring conductance,

$$G_0 = \frac{e^2}{h} = 3.87404614 \times 10^{-5}\ \text{S}, \qquad R_0 = \frac{1}{G_0} = \frac{h}{e^2} = 2.58128056 \times 10^4\ \text{k}\Omega. \tag{6.59}$$

The zero temperature conductance Eq. (6.58) is simply the conductance quantum $\frac{2e^2}{h}$ times the sum of the transmission coefficients for each channel α evaluated at the Fermi energy set by the metal reservoirs. All information about the nanostructure lies in the transmission coefficient $\mathcal{T}_\alpha(\varepsilon)$.

In the following section we shall see a spectacular consequence of the scattering wave nature of the conductance of nanostructures.

6.7 Quantized conductance

The conductance quantum is directly observable in the beautiful experiment shown in Fig. 6.9 based on the quantum point contact depicted in Fig. 6.1a. In the following we give a simple explanation of the observed quantization of the conductance.

The quantum point contact is fabricated on a GaAs-GaAlAs heterostructure. As explained in Exercise 3.4 all conduction electrons of this structure are bound to move at the interface between GaAs and GaAlAs, all having the same wavefunction $\zeta_0(z)$ in the z direction (perpendicular to the interface). By etching techniques a narrow wire is created in the x direction. As can be seen in Fig. 6.1a the width w of the wire in the y direction varies with position, $w(x)$, being of the order 100 nm at the narrowest point. The actual width can be controlled during the experiment by changing the gate voltage V_g on the side electrodes as sketched in Fig. 6.9a.

Let us model the potential $V_x(y)$ of the wire at each position x as a simple potential box stretching $w/2$ on both sides of the x axis:

$$V_x(y) = \begin{cases} 0, & \text{for } -\frac{1}{2}\,w(x) < y < \frac{1}{2}\,w(x), \\ \infty, & \text{for } |y| > \frac{1}{2}\,w(x). \end{cases} \tag{6.60}$$

We further imagine that the change of width as we move along the x axis is so slow that a simple product wavefunction fitting to the width at any given position is a good solution

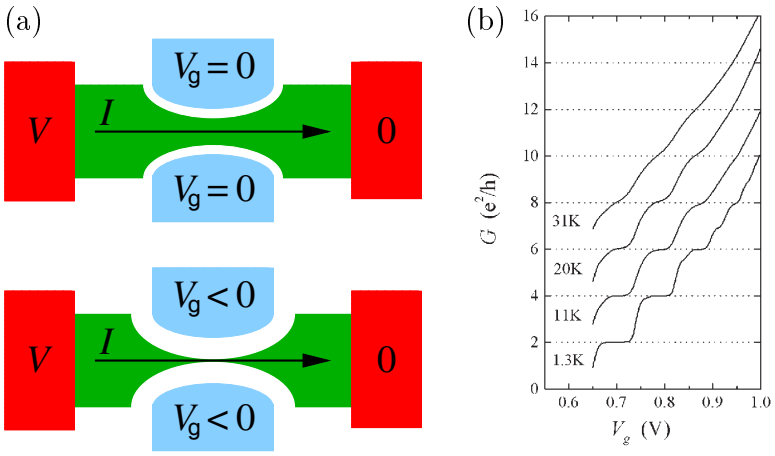


Figure 6.9: (a) The principle in measuring the quantized conductance of a point contact. The channel width is controlled during the experiment by tuning the voltage V_g on the gate electrodes. (b) Measurements of the conductance G versus gate voltage V_g at $T = 1.3, 11, 20,$ and 31 K. The curves are displaced vertically for clarity.

to the Schrödinger equation. This approximation which effectively neglects the derivatives $w'(x)$ and $w''(x)$ is known af the adiabatic approximation. The explicit wavefunctions are

$$\psi_{\varepsilon n0}(x,y,z) = \frac{1}{\sqrt{L}} e^{ik_{\varepsilon n}(x)x} \sqrt{\frac{2}{w(x)}} \sin\left(n\pi\left[\frac{y}{w(x)} + \frac{1}{2}\right]\right) \zeta_0(z). \tag{6.61}$$

The wavefunction is fully characterized by the energy ε, the number n of standing half-waves in the wire and the number 0 reminding us that all electrons have the same wave $\zeta_0(z)$ in the z direction. The index α used prior is here a double index $\alpha = n0$. The total energy ε of the state $\psi_{\varepsilon n0}$ can be written as

$$\varepsilon = \varepsilon_k + \varepsilon_n(x) + \varepsilon_z = \frac{\hbar^2 k_{\varepsilon n}(x)^2}{2m} + \frac{\pi^2\hbar^2}{2m}\frac{n^2}{w(x)^2}. \tag{6.62}$$

Here we have chosen to put the zero point of the energy scale at the value ε_z of the quantum energy of $\zeta_0(z)$, i.e. $\varepsilon_z = 0$. We can interpret Eq. (6.62) as the energy eqaution for a particle moving along the x-axis with a position dependent kinetic energy $\varepsilon_k = \frac{1}{2m}\hbar^2 k_{\varepsilon n}(x)^2$ and a position dependent potential energy $\varepsilon_n(x)$ given by

$$\varepsilon_n(x) = \frac{\pi^2\hbar^2}{2m}\frac{n^2}{w(x)^2}. \tag{6.63}$$

As the particle moves towards the narrowest point of the point contact it must convert more and more of its kinetic energy in the forward x direction into potential energy (really kinetic energy in the transverse y direction). To find out whether the particle actually has enough kinetic energy to pass through it is of course now highly relevant to know what

is the maximal potential energy $\varepsilon_n^{\max}$ along the wire. The smallest width of the wire is denoted $w_{\min}$, then according to Eq. (6.63) we have

$$\varepsilon_n^{\max} = \frac{\pi^2\hbar^2}{2m}\,\frac{n^2}{w_{\min}^2}. \tag{6.64}$$

The x-dependent wavenumbe $k_{\varepsilon n}(x)$ is given by

$$k_{\varepsilon n}(x) = \sqrt{\frac{2m}{\hbar^2}\Big(\varepsilon - \varepsilon_n(x)\Big)}. \tag{6.65}$$

We see that in our simple model we obtain full transmission if $k_{\varepsilon n}(x)$ remains real throughout the passage, i.e. if $\varepsilon > \varepsilon_n^{\max}$. If however we arrive at $\varepsilon < \varepsilon_n^{\max}$ then $k_{\varepsilon n}(x)$ becomes imaginary corresponding to tunneling through the effectigve potential barrier $\varepsilon_n(x)$. As a rough model for the transmission coefficient $\mathcal{T}_n$ of channel n, i.e. all electrons having a transverse wave consisting of n half-waves, we simply take

$$\mathcal{T}_n(\varepsilon) = \begin{cases} 1, & \text{for } \varepsilon > \varepsilon_n^{\max} \\ 0, & \text{for } \varepsilon < \varepsilon_n^{\max}. \end{cases} \tag{6.66}$$

With the transmission coefficient Eq. (6.66) inserted into Eq. (6.58) we obtain an explanation of the zero temperature conductance of the quantum point contact. The resulting formula is namely

$$G(T=0, V_g) = \frac{2e^2}{h}\,\sum_n \mathcal{T}_{n,V_g}(\varepsilon_{\rm F}), \tag{6.67}$$

where we explicitly write that the conductance must depende on the gate voltage V_g that according to Fig. 6.9a controls the actual minimal width of the wire, i.e. $w^{\min} = w^{\min}(V_g)$. If we initially have a large negative V_g the channel will be pinched off entirely. No electrons can pass and $G = 0$. As we gradually increase V_g the wire gets wider and wider. At some point electrons in channel $n = 1$ will be able to pass, since they have the lowest threshold potential $\varepsilon_{n=1}^{\max}$, and the conductance jumps from 0 up to $2e^2/h$. As V_g is increased further we reach a point where also the electrons in the $n = 2$ channel can pass, and the conductance increases from $2e^2/h$ to $4e^2/h$. This process goes on as we increase V_g more and more, and we thus trace out a conductance curve like the lowest one ($T = 1.3$ K) in Fig. 6.9.

This result is a dramatic deviation from classical physics. The conductance does not increase linearly with decreasing width, but instead it increases in discrete steps of height $2e^2/h$, the conductance quantum. Another deviation from Ohm's law is achieved if one places two quantum point contacts in series close to one another. If we assume that point contact 1 is more narrow than point contact 2, then electron waves can only pass through the second if they are able to pass through the first. So the total conductance G of the system is entirely determined by one of the point contacts, $G = \min\{G_1, G_2\} = G_1$. In terms of the electrical resistance R we have $R = \max\{R_1, R_2\}$. But this is radically

different form the usual classical formula $R = R_1 + R_2$. We only recover the classical result if the distance between the point contacts is so large that the quantum waves are destroyed in the region between them, and the whole electron wave picture thus breaks down.

Chapter 7

Scanning Tunneling Microscopy (STM)

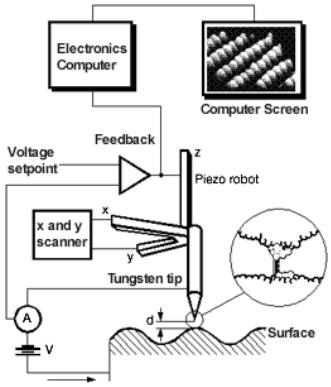
Although treated first in these lecture notes, the AFM was actually preceded by the scanning tunneling microscope (STM). Invented in 1981 the STM was the first of many scanning probe microscopes (SPM). In 1986 the inventors Gert Binnig and Heinrich Rohrer from the IBM Research Laboratory in Rüschlikon (Switzerland), were awarded the Nobel Prize in Physics. Nowadays, SPMs can be found in many academic and industrial physics, chemistry and biology laboratories, and as we have already seen, they are used both as standard analysis tools and as high-level research instruments.

7.1 The basic principle of the STM

The basic principle of the STM is sketched in Fig. 7.1. In contrast to the atomic force microscope the STM can only scan conducting surfaces. The reason is that the instrument relies on the small tunneling current, typically in the sub-nA regime, that runs between the substrate and the metal (tungsten) tip. Through a computer controlled feed-back loop and a piezo-element this current controls the height d in the z direction over the substrate. Scans across the surface in the xy plane are controlled by a piezo-electric element. The feed-back voltage controlling the height in the z direction together with the xy scan voltage is used as output.

With the STM one can achieve a resolution of around 0.1 nm. This is better than the AFM, and the reason for this is the exponential dependence of the distance d in the transmission coefficient, $\mathcal{T}(\varepsilon) \approx \exp[-4\kappa d]$, that were derived in Eq. (6.40). For a typical parameter values the tunneling current reduces by a factor 10 for every 0.1 nm increase in d. This means that over a typical atomic diameter of e.g. 0.3 nm, the tunneling current changes by a factor 1000. This is what makes the STM so sensitive. The tunneling current depends so strongly on the distance that it is dominated by the contribution flowing between the last atom of the tip and the nearest atom in the specimen.

The scan range for a STM is typically up to 1 μm with a scan speed of the order mm/min. The scan speed depends on the mode used for the STM.



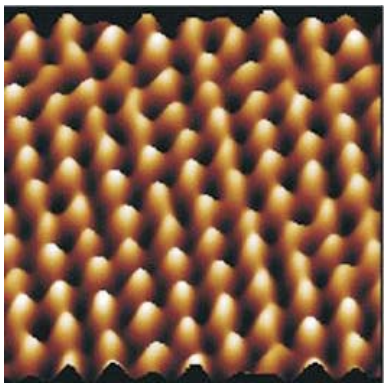


Figure 7.1: The principle of the STM. The instrument relies on the tunneling current between the conducting substrate and the metallic scanning tip. Through a computer controlled feed-back loop and a piezo-element this current controls the height d over the substrate. Scanning across the surface is controlled by a piezo-robot. The feed-back voltage is used as output. To the right is shown a 3 nm by 3 nm surface scan with atomic resolution of a graphite surface.

The constant height mode is the fastest scan mode. Here the tip is kept at a fixed vertical position during the scan. The changes in the tunneling current thus reflects the electronic topology of the surface. Using this mode there is a risk that the tip will bump into unexpected high regions on the surface and be destroyed.

The constant current mode is perhaps the most widely used STM mode. In this mode the feed-back mechanism ensures that the tunneling current is kept constant by displacing the tip vertically as the scan proceeds. This is slower than the constant height mode, but without its risk.

7.2 The piezo-electric element and spectroscopy

The principle of the piezo-electric element is sketched in Fig. 7.2. The principal effect is that for crystal lacking inverison symmetry an applied voltage in one direction makes the element longer or shorter in another direction.

A simple tripod allowing for control of the tip in all three spatial directions can be made by mounting three piezo-rods perpendicular to each other Fig. 7.2(b).

However, in most modern scanning probe microscopes, one uses a tube geometry, like the one shown on Fig. 7.2(c). Each of the four indicated sections can be made longer or shorter individually. If all four sections are made longer or shorter by the same amount, the tip moves in the z direction. If the $X+$ side is made longer, and at the same time the $X-$ side is made shorter by the same amount, the tube tilts a little bit to one side, as indicated. For small deformations, this makes the tip move primarily in the x direction. The same can be done in the y direction.

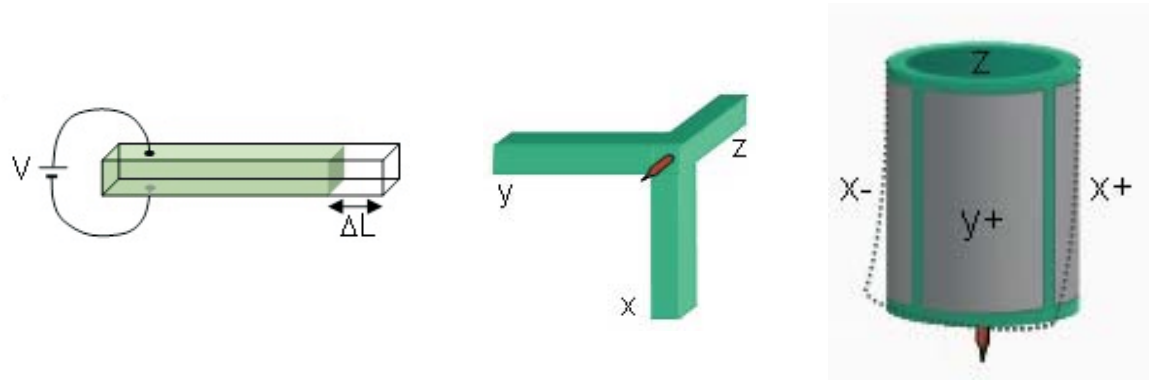


Figure 7.2: The principle of the piezo-element. (a) Elongation of a simple piezo-electric by applying a voltage. (b) The tripod consisting of three mutually perpendicular piezo-rods. (c) The tube geometry that is widely used in many STM setups.

7.3 The local electronic density of states

It is important to realize that the STM does not measure the topology but rather in some sense the electronic topology. From the discussion of quantum conductance in Sec. 6.6 we know that the current running between two reservoirs is given by counting how many electrons that are transferred per unit time. However, in our previous analysis we tacitly assumed that there always were states available in both reservoirs. This is true if we are dealing with bulk metals. This assumption is not generally true, and in fact we need to introduce the concept of the local electronic density of states $d(\mathbf{r}, \varepsilon)$.

The density of state function contains the information about how many quantum states of energy ε there are available in the neighborhood of the point $\mathbf{r}$ on the surface. Therefore the expression for the tunneling current $I(\mathbf{r})$, when the tip is placed at the position $\mathbf{r}$ is modified to

$$I(\mathbf{r}) = C\mathcal{T}(\varepsilon)\, d(\mathbf{r}, \varepsilon) \approx A \exp[-4\kappa d]\, d(\mathbf{r}, \varepsilon). \tag{7.1}$$

Consequently great care must be taken when interpreting STM scan pictures. Imagine using the constant height mode to scan a metallic surface covered with some small non-conducting particles. Such particles have a very small electronic density of states. Therefore according to Eq. (7.1) the tunneling current would be very low when the tip is just above one such particle even though the distance is small, while the current would be much higher for positions of the tip above the metallic substrate, even though the distance is larger. With a naive interpretation of the resulting STM picture one would get the impression that the sample has a lot of holes, completely contrary to the real situation.

With a proper understanding, the appearance of the electronic density of states in Eq. (7.1) can be advantageous. The STM can namely be used as a spectrometer for measuring electronic energies. Consider a STM tip placed at a fixed position $\mathbf{r}_0$. By changing the applied voltage V between the STM tip and the sample the energy ε of the electrons leaving the STM tip is changed by the amount $\varepsilon = \varepsilon_0 - eV$, where ε_0 is the energy at zero applied voltage. In an energy range where the energy dependence of

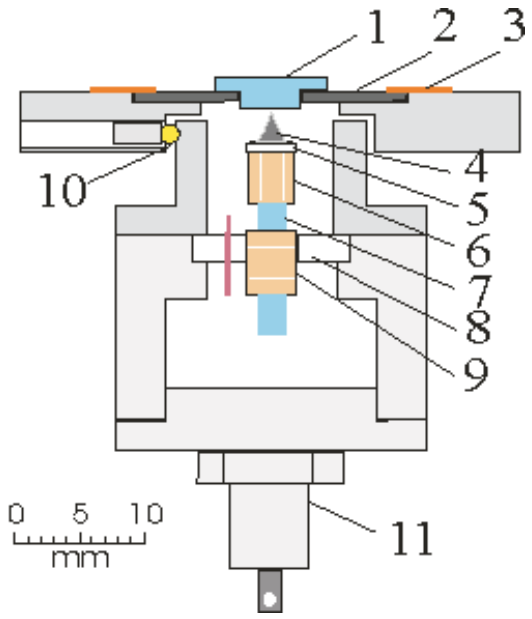


Figure 7.3: An example of a STM from Aarhus University, Denmark. On the figure is shown: (1) Sample, (2) Sample holder, (3) Clamps, (4) Tip, (5) Tip holder, (6) Piezoelectric scanner tube, (7) Approach motor rod, (8) Motor mount, (9) Approach mount, (10) Quartz balls, and (11) Zener Diode. For more explanations see the text in Sec. 7.4

the transmission coefficient is negligible $\mathcal{T}(\varepsilon) = \mathcal{T}$, measuring the tunneling current as a function of aplied voltage V gives directly an energy scan of the density of states,

$$I(\mathbf{r}_0, V) = C\mathcal{T}\, d(\mathbf{r}_0, \varepsilon - eV), \tag{7.2}$$

and thus a direct measurements of at which energies there are electronic states available in the sample. This technique has been used in numerous investigations on materials on the nanometer scale.

7.4 An example of a STM

We end this chapter on the STM by showing an concrete example of a STM. It is taken from the Department of Physics and Astronomy, Aarhus University, Denmark (www.ifa.au.dk/camp/home.htm). A sketch of the instrument is shown in Fig. 7.3

The setup is described as follows. The sample (1) is placed in a tantalum holder (2), which may be removed from the STM, and which is normally held down on the STM top by springs (3). The top plate is thermally and electrically insulated from the STM body by three quartz balls (10). The top plate is mounted on a 0.6 kg aluminum block which may be cooled to $-160°C$ or heated to $+130°C$. The tip (4) is held by a macor holder (5), which is affixed to the top of the scanner tube (6). The scanner tube is 4 mm long with an outer/inner diameter of 3.2/2.1 mm and is mounted to the rod (7), which together with

the piezo tube (9) forms a small inchworm motor used for coarse approach. The electrode of the tube is divided into three rings. In the tube two bearings are placed under the upper and the lower electrode with an extremely good fit to the rod (7).

Applying a positive voltage to an electrode will clamp that electrode to the rod whereas a negative voltage will free that electrode from the rod. A voltage applied to the center electrode will cause it to elongate or contract. With the right sequence of voltages applied to the three electrodes the rod will move up or down since the tube is fixed to the STM body by the macor ring (8). The motor may work in steps of down to 2 Å, but at full speed it moves around 2 mm/min. The scan range is up to +-1 um when using antisymmetrical scan voltages of +-200 V.

The Zener diode BZY93C75 (11) is used to counter-heat the STM body during cooling.

Appendix A

Exercises

Exercises for Chap. 1

Exercise 1.1

Moore's law for lengths of gate electrodes. An exponential law cannot be maintained indefinitely. According to Moore's law in Fig. 1.3 when will the length of a gate electrode equal the diameter of a hydrogen atom? Discuss the result.

Exercise 1.2

The minimum photolithographic line width. The usual diffraction limit for resolution of light with wavelength λ is roughly given by $w_{\text{diff}} \approx \lambda$. Discuss this in comparison with the minimum photolithographic width w_{min} given by Eq. (1.1).

Exercise 1.3

Exposure time of electron beam lithography. Use Eq. (1.3) to estimate the exposure time for an electron beam lithography process on a 100 mm diameter wafer with a pattern density of 10%, when the clearing dose is $D = 250\ \mu\text{C cm}^{-2}$ and the beam current is $I = 20$ pA.

Exercise 1.4

The de Broglie wavelength of an electron. Derive the expression Eq. (1.2) for the de Broglie wavelength λ of an electron accelerated by a voltage drop U. Use the information given just above the equation and verify the result $\lambda = 0.012$ nm for $U = 10$ kV.

Exercises for Chap. 2

Exercise 2.1

Frequency, energy, and temperature. Calculate the energy in J and eV and the

corresponding temperatures in kelvin of the characteristic photons coming from (a) a FM-radio transmitter, (b) a mobile phone, (c) a candle, and (d) an X-ray machine at the dentist. Use that $hf = E = k_{\mathrm{B}}T$, where h is Planck's constant and k_{B} is Boltzmann's constant.

Exercise 2.2

de Broglie wave lengths. From $p = h/\lambda$ find the de Broglie wave lengths for the following particles: (a) a nitrogen molecule in air at room temperature, (b) an electron in copper given a velocity of 1.57×10^6 m/s, (c) an electron in gallium-arsenide, given a velocity of 1.12×10^5 m/s and an effective mass of $m^* = 0.067m_e$, and (d) a rubidium atom in a cold atomic trap with a temperature of 40 nK.

Exercise 2.3

A simple estimate of the radius of the hydrogen atom. Due to the wave nature of particles it costs energy to localize them. This can be used to estimate the size of a hydrogen atom. (a) Convince yourself that the classical energy given in Eq. (2.14) is correct. (b) Use the Heisenberg uncertainty principle in the extreme quantum limit Eq. (2.11) $p = \hbar/a$ to derive the quantum energy given in Eq. (2.14). (c) Verify that the Bohr radius a_0 indeed is the radius that minimizes the quantum energy. (d) Verify the expression for the ground state energy $E_0 = E(a_0)$ given in Eq. (2.16). Although the estimates for a_0 and E_0 are based on an approximate quantum theory, Iit turns out that by chance the estimates are exact.

Exercise 2.4

A simple estimate of the maximal height of mountains. The quantum pressure of particle waves also determines the maximal height of mountains on Earth. A mountain cannot exceed the height H for which the potential energy MgH of the last added molecule, say SiO_2, equals the energy E_{bend} that it takes to bend the electrons orbitals forming the mineral structure at the bottom of the mountain. Now, to remove an electron completely from a hydrogen atom costs $E_0 = 13.6$ eV, but electrons in minerals are bound more loosely, typically with an energy $0.1E_0$, and furthermore we only need to bend the electron orbitals not destroy them alltogether, so we take $E_{\mathrm{bend}} \approx 0.1 \times 0.1 \times E_0$. Given this, find an estimate for the maximal height H of mountains.

Exercise 2.5

Size-quantization. The minimal quantized kinetic energy for a particle in a cubic box is given in Eq. (2.12). (a) Find the similar expression for a box with unequal side lengths L_x, L_y, and L_z. (b) In general the standing wave in the box can contain any integer number n_x, n_y, and n_z of half wavelengths in each of the three directions. Find the general expression for the size-quantized kinetic energy $E(n_x, n_y, n_z)$.

Exercise 2.6

The energy spectrum of a particle in a box. (a) Write down the triplets (n_x, n_y, n_z) corresponding to the lowest 10 energy levels. (b) If a number g of triplets lead to the same energy, that energy level is said to be g-fold degenerate or to have a degeneracy of g. What are the degenracies of the 10 lowest energy levels? (c) Find the difference in energy between the two lowest states for an electron in each of the three cubes with side lengths $L = 0.1$ nm, $L = 10$ nm, $L = 1$ cm. (d) What are the corresponding temperatures?

Exercise 2.7

A semiconductor nanowire. In modern computer chips the smallest wires are only 100 nm wide. Consider a silicon nanowire in the x-direction with a square cross section $L_y = 100$ nm and $L_z = 100$ nm. The effective mass of an electron in silicon is $m^* = 0.2m_e$. (a) What is the energy difference between the two lowest quantum states for an electron in the nanowire with $p_x = 0$ kg m/s? (b) Same question for a 10×10 nm^2 quantum wire.

Exercise 2.8

The harmonic oscillator. The classical 1D harmonic oscillator with the equation of motion $m\ddot{x} = -Kx$ has the energy $E = \frac{p^2}{2m} + \frac{1}{2}Kx^2$. (a) Express the oscillation frequency ω in terms of m and K. (b) Now use quantum theory and write down the time-independent Schrödinger equation for the oscillator. (c) Show that $\psi_0(x) = \exp[-x^2/(2\ell^2)]$ is a solution to this equation if the characteristic length ℓ is chosen properly. (d) Determine the solutions $\psi_1(x) = (x + B)\exp[-x^2/(2\ell^2)]$ and $\psi_2(x) = (x^2 + Bx + C)\exp[-x^2/(2\ell^2)]$ keeping ℓ the same. (e) Express the corresponding eigenenergies E_0, E_1, and E_2 in terms of ω.

Exercise 2.9

1D wavefunctions. Draw the three lowest eigenstates for the 1D box and the 1D harmonic oscillator. Discuss similarities and differences between the two sets of wavefunctions.

Exercises for Chap. 3

Exercise 3.1

The momentum of a standing wave. Calculate and discuss the expectation value $\langle\hat{p}\rangle_k$ of the momentum operator $\hat{p}$ for the standing wave $\tilde{\psi}_k = \sqrt{\frac{2}{L}}\sin\left(\pi\frac{x}{L}\right)$.

Exercise 3.2

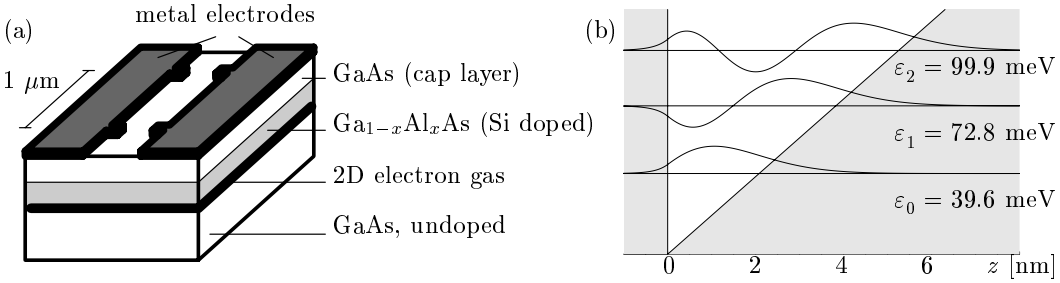
Kinetic energy and momentum of a travelling wave. Prove that the travelling wave Eq. (3.6) indeed is an eigenstate of both the kinetic energy operator and the momentum operator with the appropriate eigenvalues as stated in Eqs. (3.7a) and (3.7b).

Exercise 3.3

The microscopic parameters of metallic iron. Iron (Fe) in its metallic state has valence II, and X-ray measurements have revealed that it forms a body-centered-cubic (BCC) crystal. Each cubic unit cell in the BCC crystal has side length $a = 0.287$ nm, while one atom sits in each corner and one in the center. Find the density n of the resulting gas of valence electrons, and determine $k_{\rm F}$, $\varepsilon_{\rm F}$, $v_{\rm F}$, and $\lambda_{\rm F}$.

Exercise 3.4

A 2D electron gas. In GaAs-heterojunctions the conduction electrons are caught at the interface between GaAs at one side and GaAlAs at the other. Many nano-devices are based on this structure. The interface lies in the xy plane. The wavefunction of the electrons has the special form $\psi_{k_x k_y n}(\mathbf{r}) = (1/\mathcal{A})e^{i(k_x x + k_y y)}\zeta_n(z)$, where $\mathcal{A}$ is the area and $\zeta_n(z)$ describes the confinement to the interface (see figure). In the ground state all electrons have the same wavefunction $\zeta_0(z)$ in the z-direction so motion in that direction is locked and can be left out of the problem. The electrons are only free to move in the xy-plane described by $(1/\mathcal{A})e^{i(k_x x + k_y y)}$, and one therefore calls this system for the 2D electron gas (2DEG). Eqs. (3.12) and (3.16) relates $k_{\rm F}$ and $E_{\rm kin}/N$ to the density n in 3D. Derive the analogous expressions for the 2DEG using $\psi_{\mathbf{k}}(x, y) = (1/\mathcal{A})e^{i(k_x x + k_y y)}$.



(a) A sketch of a GaAs heterostructure with a 2DEG at the GaAs-$Ga_{1-x}Al_xAs$ interface in the xy plane. (b) The potential profile in the z direction near the interface (GaAs for $z > 0$). The wavefunctions $\zeta_n(z)$ and their energies ε_n for $n = 0, 1$ and 2 are also shown.

Exercise 3.5

Thermal smearing of the Fermi sphere. Consider a Fermi-sphere that contains N electrons and that have the Fermi energy $\varepsilon_{\rm F}$. As the temperature is raised from zero to T some of the ΔN electrons that started out with energies ε in the interval $\varepsilon_{\rm F} - k_{\rm B}T < \varepsilon < \varepsilon_{\rm F}$ can, roughly speaking, receive thermal energy kicks by energies up to $k_{\rm B}T$. Thereby they are thermally excited. Excitations of any of the other electrons in the Fermi sphere are blocked by the Pauli exclusion principle. (a) Express the ratio $\frac{\Delta N}{N}$ of the number ΔN of thermally excitable electrons over the total number N of electrons in terms of $\varepsilon_{\rm F}$ and $k_{\rm B}T$. (b) What is this ratio for copper at room temperature? (c) Do the electrons contribute significantly to the specific heat of copper at room temperature?

Exercise 3.6

The stability of metals. Prove from Eq. (3.23) that the stability point for metals indeed is given by Eq. (3.24a). What is in SI-units the density corresponding to r_s^*?

Exercise 3.7

Youngs modulus for metals in the jellium model. Hooke's law $F = K\,\Delta\ell$ for an ordinary spring states that the reaction force F is proportional to the stretch $\Delta\ell$. The proportionality constant is called the spring constant K. This law is generalized for any elastic solid to a relation between the reaction force per cross section area $\mathcal{A}$ (called stress) and the relative stretch $\Delta\ell/\ell_0$ (called strain):

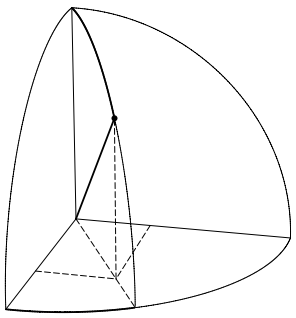
$$\frac{F}{\mathcal{A}} = Y\,\frac{\Delta\ell}{\ell_0},$$

where ℓ_0 is the length before external forces are applied. The "spring constant" Y is called Young's modulus of the material.

Consider in the jellium model the volume $1/n$ that contains one electron. The length can be taken to be $\ell_0 = r_s^*\,a_0$, where $a_0 = 0.053$ nm is the Bohr radius, and the cross section area to be $\mathcal{A} = \frac{4\pi}{3}\ell_0^2$. (a) Expand the energy $E(r_s)/N$ of the electron Eq. (3.23) around r_s^* to second order in $\Delta r_s = r_s - r_s^*$. (b) The resulting expression has the form of the potential energy of a harmonic oscillator. Find an expression for the spring constant. Note that r_s and Δr_s are dimensionless, so it is preferable to introduce $\ell = r_s\,a_0$ and $\Delta\ell = \Delta r_s\,a_0$. (c) Find an expression for Young's modulus of the model by considering $F/\mathcal{A}$. (d) Calculate Young's modulus in SI units and discuss the result in comparison with the table below.

metal	Al	Cu	Fe	Ni	Pb
$Y/[10^{10}$ Pa]	7.0	11.0	21.0	21.0	1.6

Exercises for Chap. 4



Exercise 4.1

Spherical coordinates

(a) Write in the 3D coordinate system to the right the proper definitions of cartesian coordinates (x, y, z) and spherical coordinates (r, θ, ϕ). (b) In cartesian coordinates the volume element is $dx\,dy\,dz$. What is it in spherical coordinates?

Exercise 4.2

The expectation value of 1/r in the hydrogen atom

Calculate the expectation value

$$\langle \frac{1}{r} \rangle_{100} = \int d\mathbf{r}\, \psi^*_{100}(\mathbf{r})\, \frac{1}{r}\, \psi_{100}(\mathbf{r})$$

in the ground state $\psi_{100}(\mathbf{r})$ of the hydrogen atom.

Exercise 4.3

Eigenstates of the angular momentum operator.
Prove by direct calculation that $Y_0^0(\theta,\phi)$, $Y_1^0(\theta,\phi)$, $Y_2^1(\theta,\phi)$ are eigenstates to both $\hat{\mathbf{L}}^2$ and L_z. What are the corresponding eigenvalues?

Exercise 4.4

The angular part of the Laplacian.
Prove Eq. (4.35) using Eq. (4.34).

Exercise 4.5

Radial wavefunctions
Plot the six radial wavefunctions in Eq. (4.25) and discuss the resulting graphs.

Exercise 4.6

Spherical harmonics
Select five of the spherical harmonics in Eq. (4.14) and plot their absolut values $|Y_l^m(\theta)|$ in a polar plots. Discuss the resulting graphs

Exercise 4.7

Counting of orbitals
Go through the argument in Secs. 4.5 and 4.6 leading from a single carbon atom with the $2s^2 2p^2$ configuration to the σ and π orbitals in graphene. Make a list over the orbitals present at each stage and their occupancy.

Exercise 4.8

The chiral vector in carbon nanotube
Consider the chiral vector $\mathbf{c} = n\mathbf{a}_1 + m\mathbf{a}_2$ in Fig. 4.6. What are the values of n and m in this case? Describe the resulting carbon nanotube.

Exercises for Chap. 5

Exercise 5.1

An actual AFM-setup. Use the Internet or the library to find an example of an actual AFM-setup. What kind of tip and cantilever was used? Which mode was used during the AFM-scan? What was the sample?

Exercise 5.2

The effective mass for an oscillating cantilever. In the circular shape approximation the displacement in the z direction of an oscillating cantilever is written as $z(x,t) = z_0(x/L)^2 \cos(\omega t)$. The cantilever has a homogeneous mass density $\rho = m/L$, where m is the total mass of the cantilever and L its length. (a) What is the velocity of the end point $x = L$? (b) If the entire mass was situated at the end point what would then be the kinetic energy? (c) What is the velocity of an arbitrary point between 0 and L? (d) Calculate the kinetic energy of the cantilever and show why it is reasonable to say that the effective mass is $m_{\text{eff}} = \frac{1}{5}\, m$.

Exercise 5.3

A copper cantilever. Young's modulus and the mass density of copper is given by $Y_{\text{Cu}} = 1.1 \times 10^{11}$ Pa and $\rho_{\text{Cu}} = 8.96 \times 10^3$ kg m^{-3}. Consider a rectangular cantilever with the dimensions $L = 100$ μm, $w = 10$ μm, and $h = 1$ μm. Assume the a force F_{tip} is deflecting the tip $\Delta z \approx 1$ Å. (a) Calculate the spring constant K, the tip-force F_{tip}, and the resonance frequency ω_0 for the system. (b) Compare with the results Eqs. (5.7a) – (5.7c) for silicon and discuss the differences.

Exercise 5.4

Atomic Polarization. As stated in Eq. (2.14) and in Fig. 5.4b an approximate expression for the energy $E(a)$ of the hydrogen atom as a function of the average distance a from the electron to the nucleus is given by

$$E(a) = \frac{\hbar^2}{2m}\,\frac{1}{a^2} - \frac{e^2}{4\pi\epsilon_0}\,\frac{1}{a}.$$

Show in detail how to get from this expression to Eq. (5.11) for the atomic dipole ed:

$$ed = \frac{(a_0 e)^2}{2E_0}\,\mathcal{E}.$$

Exercise 5.5

The relative dielectric constant of a gas. In classical electromagnetism the relative dielectric constant ϵ_r is defined by the relation $D = \epsilon_r \epsilon_0 \mathcal{E}$ between the displacement field

D and the electric field $\mathcal{E}$. This can also be expressed in terms of the polarization $\mathcal{P}$, which is nothing but the number N of dipole moments ed per volume $\mathcal{V}$:

$$D = \epsilon_r \epsilon_0 \mathcal{E} = \epsilon_0 \mathcal{E} + \mathcal{P} = \epsilon_0 \mathcal{E} + \frac{N}{\mathcal{V}}\, ed.$$

(a) Combine this with Exercise 5.4 to show that

$$\epsilon_r = 1 + 8\pi N\, \frac{a_0^3}{\mathcal{V}}.$$

(b) In a gas at room temperature and atmospheric pressure Avogadro's number of molecules takes up a volume of 22.4 L. Calculate ϵ_r for hydrogen H_2 and compare with the experimental value $\epsilon_r = 1.00026$.

Exercise 5.6

The index of refraction of a liquid. The speed of light in a liquid is given by $c = c_0/n_{\text{liq}}$, where n_{liq} is the index of refraction of the liquid. We know from electromagnetic theory that the speed of light is given by $c = 1/\sqrt{\mu_r \mu_0 \epsilon_r \epsilon_0}$, where μ_0 is the magnetic susceptibility of vacuum and μ_r the relative susceptibility (which is 1 for non-magnetic media). (a) Check this formula for vacuum where $\mu_r = \epsilon_r = 1$. (b) Show that for a non-magnetic medium we have $n_{\text{liq}} = \sqrt{\epsilon_r}$. (c) Give an estimate of the density $N/\mathcal{V}$ of molecules in a liquid. (d) Calculate n_{liq} assuming that each molecule has the dipole moment ed of Exercise 5.4, and compare the result to the experimental value for water.

Exercise 5.7

The electric dipole. Go carefully through the arguments leading to Eqs. (5.12) and (5.13).

Exercise 5.8

The harmonic oscillator solved with complex numbers. Consider a harmonic oscillator given by the equation of motion $m\ddot{x}(t) = -Kx(t)$, where m is the mass and K the spring constant. (a) Assume $x(t) = A\cos(\omega t)$ and determine the resonance frequency $\omega = \omega_0$ in terms of K and m. (b) Solve the same problem now assuming $x(t) = A\exp(i\omega_0 t)$. (c) What is the relation between the real-number solution and the complex-number solution? (d) Is it possible to determine the amplitude A from the equation of motion?

Exercise 5.9

The damped harmonic oscillator. To the equation of motion of the harmonic oscillator in Exercise 5.8 we now add a friction term $-m\gamma\dot{x}(t)$ that damps the oscillator. (a) Assume the complex solution $x(t) = A\exp(i\omega t)$ and show that ω must fulfill the equation

$$-\omega^2 = -\omega_0^2 - i\gamma\omega.$$

(b) If $\gamma \ll \omega_0$ we obtain $\omega \approx \omega_0$, and therefore

$$\omega^2 \approx \omega_0^2 + i\gamma\omega_0.$$

Use this expression in $x(t) = A\exp(i\omega t)$, and sketch $x(t)$.

Exercise 5.10

The externally driven, damped harmonic oscillator. To the damped harmonic oscillator of Exercise 5.9 we now add an external driving force $F(t) = F_0\exp(i\omega t)$. The equation of motion the becomes

$$m\ddot{x}(t) = -K\,x(t) - m\gamma\,\dot{x}(t) + F_0\exp(i\omega t).$$

(a) Assume the complex solution $x(t) = A\exp(i\omega t)$, which contains the same frequency ω as the driving force, and determine the amplitude A.

(b) Sketch $|A(\omega)|^2$, i.e. the absolute square of the amplitude as a function of the driving frequency ω.

Exercise 5.11

Maximal amplitude and half with of the oscillator resonance line. Consider the externally driven, damped harmonic oscillator of Exercise 5.9. For small damping, $\gamma \ll \omega_0$, the amplitude curve $|A(\omega)|^2$ changes rapidly in a narrow range around ω_0. For frequencies ω in that range we can therefore make the following approximations:

$$\omega_0^2 - \omega^2 = (\omega_0 + \omega)(\omega_0 - \omega) \approx 2\omega_0(\omega_0 - \omega) \qquad \text{and} \qquad \gamma\omega \approx \gamma\omega_0.$$

(a) Use these approximations to show that

$$|A(\omega)|^2 = \frac{F_0^2}{4m^2\omega_0^2}\,\frac{1}{(\omega_0 - \omega)^2 + \frac{1}{4}\gamma^2}.$$

(b) For which ω is $|A(\omega)|$ maximal?
(c) Determine the maximal value of $|A(\omega)|$.
(d) For which values of ω is $|A(\omega)|$ exactly half of its maximal value?
(e) The Q-factor of a resonance peak is defined as $Q = \omega_0/\Delta\omega$, where ω_0 is the resonance frequency and $\Delta\omega$ is the full width at half maximum. Use (d) to write down an expression for the Q-factor of the resonance

Exercise 5.12

A cantilever based mass sensor. In a collaboration between MIC and institutes in Barcelona a cantilever based mass sensor with ultra high sensitivity has been developed (see the course web page for the original papers). The cantilever forms one plate in an on-chip capacitor, while an high frequency AC-voltage can be applied to the other near-by capacitor plate. The AC-voltage forces the cantilever to oscillate. By tuning the

AC-frequency $f = \omega/2\pi$ and observing the amplitude of the oscillation (either directly or through the capacitance of the system) the resonance frequency $f_0 = \omega_0/2\pi$ of the cantilever can be measured.

In an active chemical environment the mass of the cantilever chages as molecules are deposited on its surface. In the following we use the theory in Sec. 5.2 to analyze how sensitive the system is to changes of the mass of the cantilever.

(a) State the relationship between the resonance frequency f_0, the mass m of the cantilever, and the spring constant K (the latter is assume to remain constant as molecules are adhering to the cantilever).

(b) Find, say by differentiation, the relation between a small change dm of mass and the accompanying change df_0 in the resonance frequency.

(c) In one of the papers an cantilever of length $L = 48$ μm, height $h = 1$ μm, and width $w = 780$ nm is studied. Assume that the cantilever is made of pure silicon and that the resonance frequency can be measured with an accuracy of 1 Hz. Find the smallest detectable change of mass, and comment on the result.

(d) In one of the original papers the resonance peak of the cantilever is measured directly using an AFM. It is displayed in panel (e) of the figure. Estimate the observed Q-factor of the system (see Exercise 5.11).

Exercises for Chap. 6

Exercise 6.1

Electron reservoirs
Set $\mu = 1$ and plot the Fermi-Dirac function $n_{\rm F}(\varepsilon, \mu = 1, V = 0, T)$ Eq. (6.2) as a function of ε for $k_{\rm B}T = 0.01, 0.1, 1,$ and 10. Do the same for the derivative $-\partial n_{\rm F}/\partial\varepsilon$. These functions are important when dealing with the electron reservoirs connected to nanostructures.

Exercise 6.2

Right- and left-moving waves
Prove that $\psi_{\pm k}(x,t)$ in Eqs. (6.5a) and (6.5b) are solution to the time-dependent Schrödinger equation Eq. (2.17). Explain why $\psi_{+k}(x,t)$ is a right-moving wave and $\psi_{-k}(x,t)$ a left-moving wave.

Exercise 6.3

The current density J
What are the units of the current density $\mathbf{J}(\mathbf{r}) = \frac{\hbar}{m}\,\mathrm{Im}\Big[\psi^*(\boldsymbol{\nabla}\psi)\Big]$?

Exercise 6.4

The electrical current I
Prove Eq. (6.10) from Eqs. (6.7), (6.8), and (6.9).

Exercise 6.5

The eigenstates for the potential step
Show that the wavefunction $\psi_\varepsilon(x)$ given by Eq. (6.18) with the coefficients B and C specified in Eq. (6.20) is a solution to the time-independent Schrödinger equation Eq. (2.32) with the potential $V(x)$ given by Eq. (6.17).

Exercise 6.6

The transfer matrix for a potential jump
Consider the potential barrier problem defined by Eqs. (6.28) and (6.29). Prove the correctness of the transfer matrix equation Eq. (6.31).

Exercise 6.7

The transmission coefficient in the extreme tunneling regime
Prove expression Eq. (6.40) for the transmission coefficient $\mathcal{T}(\varepsilon)$ by taking the appropriate limit of the general expression Eq. (6.39).

Exercise 6.8

The probability of electron tunneling between metals
Two copper cubes of side length 1 cm are placed 1 nm from each other. Use the simple model of metals in Chap. 3 and the transmission coefficient Eq. (6.40) in the extreme tunneling regime to estimate how often an electron tunnels from one cube to the other.

Exercise 6.9

Above barrier reflection
In Fig. 6.6 it is clearly seen that $\mathcal{T} < 1$ for a slightly above-barrier energy $\varepsilon = V^+$. Find an analytical expression for the reflection coefficient $\mathcal{R}(\varepsilon = V^+)$ by expanding the sin-function in Eq. (6.42) to first order. Compare the result with the graph.

Exercise 6.10

Resonant transmission
Show from Eq. (6.42) that unity transmission is obtained when the resonance condition $2k_2a = n\pi$ is fullfilled. Calculate the position of the first two resonance peaks, $n = 1, 2$ and compare with graph ($E_a = 0.1\,V$).

Exercise 6.11

The unit of the conductance quantum
Show that the unit of the conductance quantum $G_0 = e^2/h$ indeed is siemens (S).

Exercise 6.12

The conductance formula
Go carefully through the proof of the conductance formula Eq. (6.57).

Exercises for Chap. 7

Exercise 7.1

The wavelength of an STM electron
Let the energy of an electron be given by the applied voltage of $U = 1$ V that is typically used in a STM. What is the de Broglie wavelength of this electron? Could we observe atoms using the acceleration voltage in an electron microscope?

Exercise 7.2

Electrons inside tunneling barriers
The work function Φ is the distance from the unperturbed Fermi energy to the top of the barrier that keeps electrons inside a metal. The probability $P(x)$ of an electron to have penetrated the distance x into the barrier is given by the square of its wavefunction $\psi(x)$,

$$P(x) = |\psi(x)|^2 = |\psi(0)|^2\, e^{-2\kappa x}, \quad \text{where} \quad \kappa = \frac{1}{\hbar}\sqrt{2m(V-\varepsilon)}.$$

How far does the electron penetrate into the barrier for platinum, $\Phi(\mathrm{Pt}) = 5.7$ eV, and for tungsten, $\Phi(\mathrm{Pt}) = 4.8$ eV?

Exercise 7.3

Tunneling frequncy and time
How many electrons are present in the STM tunnel gap at the same time? Do they influence each other? Time τ is not an observable in quantum mechanics, so for the barrier $V(x)$ we make a semiclassical WKB analysis,

$$\tau = \int_{x_1}^{x_2} \sqrt{\frac{m}{2[V(x)-\varepsilon]}}\, dx,$$

where x_1 and x_2 are the classical turning points where $\varepsilon = V(x_i)$. Take a rectangular barrier of height V_0. Assume a barrier width of 0.4 nm and an effective height of 4 eV. Compare the result to typical STM currents of 1 nA.

Exercise 7.4

Speed of the STM tip
A typical STM scan covers a square of 10 nm by 10 nm divided into 256 by 256 pixels. The scanning time is of the order $t_{\mathrm{scan}} = 10$ s. What is the average horizontal speed of the tip? And how long time would it take to move the tip 1 m at that speed?

Exercise 7.5

Local vacuum in the STM tunnel gap
Assume a distance $d = 0.5\ nm$ between the tip and the sample. Give an estimate of the number of air molecules present in the gap volume at room temperature and a 1 atm pressure.

Exercise 7.6

The exponential decrease in tunneling current
Use the typical values of the STM parameters from the previous exercises to verify the statement in Sec. 7.1 that by increasing the distance from the substrate by 0.1 nm the tunneling current decreases by a factor of 10.