QUANTUM MATERIALS

PHYS-C0250

Erkki Thuneberg

Department of Applied Physics
Aalto University
2022

version May 28, 2022

0.1 Introduction

These are lecture notes for the course Quantum Materials. The web page of the course is https://mycourses.aalto.fi/course/view.php?id=33560

The web page contains the lecture material (these notes), the exercises and other material. This is the second and last time this course is held in the present form. The material has mostly been prepared 2021 but there are small modifications this year, especially more figures.

Time table 2022

Lectures on Mondays and Wednesdays 12.15-14.00 (20.4.-26.5.) except the first lecture on Tuesday (19.4.) 8.15-10.00.
Exercise sessions on Fridays 10.15-12.00 (22.4.-27.5.).
[The exercise time on Thursdays is cancelled.]
All lectures and exercises in nanotalo room 228.

Teaching assistant: Fereshteh Sohrabi

Advance exercises will be placed on the web page some days prior each lecture, and they should be completed before the lecture. Each week's exercises will be placed in web page soon after Wednesday's lecture, or earlier. Try to solve the problems before the exercise session on Friday, so that during the session you can pose questions to the assistant. You still have time until next Tuesday to show completed exercises.

Exam Wednesday 1.6. at 13.00–16.00

The evaluation criteria are approximately as follows: advance assignments 15%, exercises 25%, examination 60%.

Relation to other courses

This course is part of quantum technology bachelor program. It is aimed to be an introduction to quantum physics.

ELEC-C9420 Introduction to Quantum Technology. This has given you a flash of several topics in quantum physics.

ELEC-C9430 Electromagnetism. Here you have learned basics about electric and magnetic fields, and electromagnetic radiation.

The present course, Quantum Materials. The aim is to learn basics of quantum mechanics, on the level of the book YF (see below). Since there is no preceding course on general wave motion, we have a discussion of that first.

PHYS-A0140 Structure of Matter (TFM). This course is similar in topic to the present course, but it is lectured in Finnish. In spite of similar topics, this course differs in many respects from the present one. For example, the exercises are completely different.

PHYS-C0252 Quantum Mechanics. This course will tell you more deeply about the formalism of quantum mechanics the next autumn.

PHYS-C0256 Thermodynamics and Statistical Physics. At some places we need some basic results of thermal physics, which will later be more properly explained in this course.

Topics

- Wave motion: basic quantities, wave equation, wave velocity, energy, reflection and transmission, waves in 3 D, Snell's law, interference, diffraction, dispersion, group velocity
- Particle properties of light: photoelectric effect, X-ray generation and diffraction, Compton scattering, pair production, wave-particle duality, uncertainty relation
- Wave properties of particles: Bohr model of hydrogen atom, wave nature of electron
- Quantum mechanics of single particle: Schrödinger equation, potential wells, quantum tunneling, measurement, 3 D deep potential well, hydrogen atom, angular momentum, hydrogen probability distributions, hydrogen energy levels, electron spin
- Many-particle systems: many-body problem, atomic structure and the periodic table, Pauli principle as symmetry, chemical bond, molecular rotations and vibrations
- Solid state: energy bands, conductors and insulators, free electron model of metal, superconductivity, insulators, semiconductors

Learning Outcomes (from Sisu): After the course, the student
1. knows the basics of wave propagation
2. can explain the formation of intensity distribution in interference and diffraction patterns
3. knows the Heisenberg's uncertainty principle and the wave-particle dualism in quantum mechanics
4. can solve simple quantum mechanical systems using Schrödinger's equation
5. can identify properties of atoms and condensed matter based on quantum mechanics

Content (from Sisu): Interference, diffraction, dispersion, group- and phase velocity. Heisenberg's uncertainty principle, wave-particle dualism. Schrödinger's equation, statistical nature of quantum mechanics. Quantum numbers, Pauli exclusion principle, spin and periodic table of elements.

Prerequisites

Basic knowledge of classical mechanics (Newton's equation of motion, work and energy, harmonic oscillator).

Mathematical skills: elementary functions, complex algebra and basic functions, derivation, Taylor series, integration, differential equations, vector algebra, partial derivatives.

Book

- Young and Freedman, University Physics (14th Edition), Pearson (YF). Ebook, other links on MyCourses page. This is a vast book with 1600 pages. Only small part of if will be covered in the lectures. We will concentrate on chapters 15 "Mechanical waves", 33 "Nature and propagation of light", 35 "Interference", and Modern physics covered in chapters 38-42. The lecture notes below contain references to sections in YF book that contain similar discussion. At some points the lectures contain topics that are not properly discussed in the book, for example, Eq. (33), waves in 3 dimension, group velocity, solution of Schrödinger equation for hydrogen, many-body problem, and superconductivity.

Notation, constants and formulas

(figure) this means figure to be drawn in the lectures, which is not in these notes

$a/bc = \frac{a}{bc}$ notation used in these notes

h, Planck's constant, $\hbar = h/2\pi = 1.054 \times 10^{-34}$ Js

$c = 299\,792\,458\,\mathrm{m/s}$, velocity of light in vacuum

$e = 1.602 \times 10^{-19}$ C, the elementary charge, absolute value of the electron charge. Unfortunately, e is also used to denote the Euler's number $e = 2,718\ldots$. What is meant, should be deduced from the context.

$k_\mathrm{B} = 1.380 \times 10^{-23}$ J/K, Boltzmann constant

$N_\mathrm{A} = 6.022 \times 10^{23}$ 1/mol, Avogadro constant

$\mathrm{u} = (0.001\ \mathrm{kg/mol})/N_\mathrm{A} = 1.660 \times 10^{-27}$ kg, atomic mass unit

$\epsilon_0 = 8.854 \times 10^{-12}\ \mathrm{C^2/Nm^2}$, electric constant (vacuum permittivity)

$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{N/A^2} = 4\pi \times 10^{-7}\ \mathrm{T^2m^3/J}$, magnetic constant (vacuum permeability)

$m_e = 9.109 \times 10^{-31}$ kg, mass of an electron, in most cases marked by m

$m_p = 1.6726 \times 10^{-27}$ kg, mass of a proton

1 eV$= 1.602 \times 10^{-19}$ J

$\exp(x) = e^x$, $\cosh x = \frac{1}{2}(e^x + e^{-x})$, $\sinh x = \frac{1}{2}(e^x - e^{-x})$, $\tanh x = 1/\coth x = \sinh x/\cosh x$.

$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$
$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$
a^*, complex conjugate of a

$e^{x+iy} = e^x(\cos y + i\sin y)$

1. Wave motion

1.1 Basic wave quantities

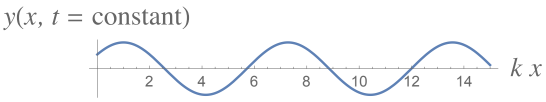
(YF 15.3) Waves are around us everywhere. You can see waves on strings or on liquid surfaces. You can hear sound waves. You use radio frequency electromagnetic waves in a multitude of devices. Light is a wave, as we will soon discuss in more detail. Even all particles have wave nature, which is the principal topic of this course. We start discussing some basic concepts of wave motion.

For a wave we need a field, that depends on location $\boldsymbol{r}$ and time t. In the simplest case it depends on one spatial coordinate x and for the field we use symbol y (following YF). Thus the field is a function $y(x,t)$. For a stretched string, for example, it is the transverse displacement of the string from its equilibrium position.

A particularly simple type of wave is a *sinusoidal wave*. It has the form

$$y(x,t) = A\cos(kx - \omega t) \tag{1}$$

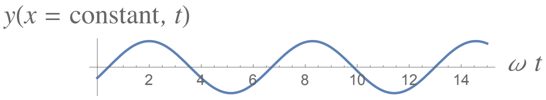
Here A is called the *amplitude* of the wave. At any fixed time t, y is sinusoidal in x.



The *wavelength* λ is the distance between two maxima, $k\lambda = 2\pi$. Thus the *wave number* k in (1) is related to the wave length by

$$k = \frac{2\pi}{\lambda}. \tag{2}$$

At any fixed location x, the wave (1) is sinusoidal in t.



The *period* T is the time interval between two maxima, $\omega T = 2\pi$. One defines the *frequency* f, which is the inverse of the oscillation period, $f = 1/T$. Thus the *angular frequency* ω in (1) is related to the period and frequency by

$$\omega = \frac{2\pi}{T} = 2\pi f. \tag{3}$$

The unit for frequency f is Hz, meaning the number of oscillations per second. For example, the note A_4, common in tuning forks, has frequency 440 Hz. If one wants to express a numerical value of ω, its unit is rad/s (or less precisely 1/s).

What is the velocity of the wave (1)? For that consider first a wave

$$y(x,t) = f(x - vt). \tag{4}$$

Here f is an arbitrary function of a single variable, for example, a step function, a single pulse or a sin function. Drawing this at two instants separated by time Δt, we see it has propagated the distance $\Delta x = v\Delta t$. Thus the velocity of the wave (4) is $\Delta x/\Delta t = v$. We can apply this result to the wave (1). Writing it to a form (4) gives the velocity

$$v = \frac{\omega}{k} = f\lambda. \tag{5}$$

A wave of the form (4), which propagates to one direction only, is called a *traveling wave*.

We see that the wave (1) propagates in the positive x direction, when we assume $\omega > 0$ and $k > 0$. We can also have a wave propagating in the negative x direction by selecting k negative. Sometimes it may be advantageous to keep k positive. In this case, a wave propagating in the negative x direction can we written

$$y(x,t) = A\cos(-kx - \omega t). \tag{6}$$

This can also be written $y(x,t) = A\cos(kx + \omega t)$ because $\cos(-\alpha) = \cos\alpha$.

Using the expressions (2), (3) and (5), the wave (1) can be written in a variety of different forms.

We interpret some properties of the wave (1) assuming it represents a transverse wave in a string. (Corresponding interpretations can be done for other types of waves.) The coordinate y gives the transverse displacement of the string. The displacement depends on the x coordinate and on time t. In order to emphasize that, we often write the arguments explicitly, $y(x,t)$. In order to get the velocity v_y of the string at any fixed value of x, we take the *partial derivative*, where we take the derivative with respect to t while keeping x constant. This is denoted by $\partial/\partial t$. For (1) we get

$$v_y(x,t) = \frac{\partial y}{\partial t}(x,t) = \omega A\sin(kx - \omega t). \tag{7}$$

In order to get the acceleration a_y, we need the second partial derivative with respect to time, marked with $\partial^2/\partial t^2$. This gives

$$a_y(x,t) = \frac{\partial v_y}{\partial t}(x,t) = \frac{\partial^2 y}{\partial t^2}(x,t) = -\omega^2 A\cos(kx - \omega t). \tag{8}$$

The slope is defined by $s = [y(x+\delta,t) - y(x,t)]/\delta$, where we take the limit $\delta \to 0$. This is just the partial derivative with respect to x, and we get for (1)

$$s(x,t) = \frac{\partial y}{\partial x}(x,t) = -kA\sin(kx - \omega t). \tag{9}$$

One can also take mixed partial derivatives like $\partial^2 y/(\partial x\partial t)$ and they are independent of the order in which the two partial derivatives are taken. One can also take higher partial derivatives, like $\partial^4 y/\partial x^4$. [exercise: for

(1) calculate the mixed second partial derivatives, verify their equality and calculate $\partial^4 y/\partial x^4$, which will be needed later.]

More generally than in (1), an additional *phase shift* ϕ_0 could be added to the argument of cosine. Such a form can also be represented in alternative forms

$$\begin{aligned} y(x,t) &= A\cos(kx-\omega t+\phi_0) \\ &= B\cos(kx-\omega t)+C\sin(kx-\omega t) \\ &= \Re[De^{i(kx-\omega t)}]. \end{aligned} \tag{10}$$

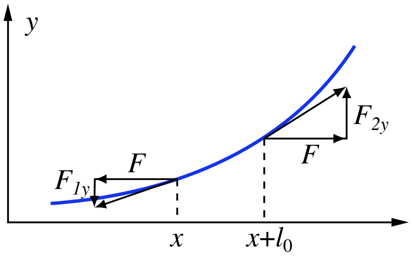
Here A, B, C and ϕ_0 are real quantities. In the first form one can always choose $A \geq 0$. This is possible since $-\cos(\phi) = \cos(\phi+\pi)$. Only by choosing $A > 0$ does $y(x,t)$ determine ϕ_0 uniquely modulo 2π. (That is, ϕ_0 is determined uniquely in the range from 0 to 2π, or any range of length 2π.) The last representation in (10) uses a complex valued $D = \Re D + i\Im D$, where $\Re$ and $\Im$ denote the real and imaginary parts, i is the imaginary unit ($i^2 = -1$), and the complex exponential function $e^{i\phi} = \cos\phi + i\sin\phi$. Each of the alternative forms in (10) can be most convenient, depending on the application. [Exercise: verify the equivalence of the three forms (10) and find relations between A, B, C, D and ϕ_0.] In using the last form, it is convenient to do calculations using a complex valued $y = De^{i(kx-\omega t)}$, with the understanding that the physical solution is only the real part of that.

1.2 Wave equation

(YF 15.4, 15.6) As an example, let us consider a string under tension, a guitar string, for example. The string has mass per length μ and is stretched by tension force F. It is deviated from its equilibrium position by $y(x,t)$, which is assumed to be small. We formulate a Newton's equation of motion

$$m\boldsymbol{a} = \Sigma \boldsymbol{F} \tag{11}$$

to a small section of the string between x and $x + l_0$.



The tension is mostly in the x direction, but because of the deviation $y(x,t)$, there is also a component of the force in the y direction

$$F_y = F\frac{\partial y}{\partial x}. \tag{12}$$

Considering the forces acting on the mass element, there is force

$$F_{1y} = -F\frac{\partial y}{\partial x}(x,t) \tag{13}$$

acting on the element from the left hand side, and force

$$F_{2y} = F\frac{\partial y}{\partial x}(x+l_0,t) \tag{14}$$

on the right hand side. The net force is

$$\sum_i F_{iy} = F\frac{\partial y}{\partial x}(x+l_0,t) - F\frac{\partial y}{\partial x}(x,t) \approx Fl_0\frac{\partial^2 y}{\partial x^2}. \tag{15}$$

On the left hand side of the Newton's equation (11) we have

$$ma_y = \mu l_0 \frac{\partial^2 y}{\partial t^2}. \tag{16}$$

Then we get

$$\mu\frac{\partial^2 y}{\partial t^2} = F\frac{\partial^2 y}{\partial x^2}. \tag{17}$$

Now it is our hope that this equation has a solution of the wave form (1). By calculating the second partial derivatives and substituting into (17) verify that this indeed is the case assuming the velocity of the wave

$$v = \sqrt{\frac{F}{\mu}}. \tag{18}$$

We see that the velocity of the wave depends only on the tension and on the mass density of the string. In particular, the velocity is independent of the frequency (or the length) of the wave.

It should be noted that we have not considered any friction forces. As a consequence, the waves we consider here are not damped. In many cases this is a good approximation, and it simplifies our discussion. When needed, the damping can be taken into account, as discussed in the Electromagnetism course, for example.

Let us rewrite (17) using (18):

$$\boxed{\frac{\partial^2 y}{\partial t^2} - v^2\frac{\partial^2 y}{\partial x^2} = 0.} \tag{19}$$

This is known as the *wave equation.* More precisely it is the wave equation in one spatial dimension. We derived it above for small transverse oscillations of a stretched string, but it can be derived for quite many cases. For example, sound waves in fluids and solids, transverse waves in solids, electromagnetic waves in vacuum. (electrocmagnetic waves: YF 32.2)

We discuss two important properties of the wave equation (19).

- It is a *linear equation.* This means that if $y_1(x,t)$ and $y_2(x,t)$ are two solutions of the equation, then also $y(x,t) = ay_1(x,t) + by_2(x,t)$ is a solution (where a and b are arbitrary real numbers). This is also known as the *principle of superposition.*

- The velocity of the wave is the same for all waves. That is, v (5) is constant that is independent of the frequency f (or the wave length λ) of the wave. Alternatively, this can be expressed that general wave from (4) is a solution of the wave equation. (exercise)

The former property means that waves of different frequencies propagate as if other frequencies were not present at all. In other words, waves with different frequencies do not interact with each other, the total $y(x,t)$ is the simply the sum of waves at different frequencies. More generally, there could be nonlinear terms that should be added to the linear wave equation (19). If the amplitude of the wave is large, such nonlinear terms become important. For example, the waves on a see break on strong wind, and with strong laser light, a wave with one frequency can generate waves of different frequency in some materials. But with small amplitude, the linear wave equation is sufficient, and this is the case we will confine ourselves in this course.

The latter property, the constancy of the propagation velocity of the waves, holds in many cases for mechanical and electromagnetic waves. However, there are also deviations. A prism breaks white light into different colors. This is a result that light of different frequency has different velocity in a material medium, a property called *dispersion*. Waves on water surface have similar property. The waves describing material particles in quantum mechanics are also dispersive. The equations governing dispersive waves are more complicated than (19). We consider dispersive waves later in this course.

1.3 Velocities of some waves

(YF 16.2, 32.3) Sound in a fluid, like air, is compressional wave. The wave is *longitudinal*, as the motion is in the same direction as the propagation of the wave. When a fluid is compressed, its pressure p increases. For small compressions this change of pressure is described quantitatively by bulk modulus,

$$B = -V\frac{dp}{dV}, \tag{20}$$

where V is the volume of the fluid. The increasing pressure in a fluid element in the wave causes a force that tends restore the uniform density. Similarly to the case of stretched string, one can write a Newton's equation of motion. Skipping the details, this again leads to the wave equation (19), but with the velocity given by

$$v = \sqrt{\frac{B}{\rho}}, \tag{21}$$

where ρ is the density of the fluid.

In order to make (21) more concrete, let us consider sound in a gas like air, which nearly behaves like ideal gas at normal conditions. Ideal gas obeys the equation of state $pV = nRT$, where n is the number of moles, $R = 8.31$ J/(K mole) the gas constant and T the absolute temperature. If temperature were constant, the condition $pV =$ constant would give $B = p$, the ambient pressure. However, the gas heats in compression. The compression in a typical sound wave is so fast that there is no time for the heat to flow anywhere. Such an *adiabatic process* is governed by equation $pV^\gamma =$ constant, where $\gamma = C_p/C_V$ is the ratio of the heat capacities at constant pressure and constant volume. This gives $B = \gamma p$ and sound velocity $v = \sqrt{\gamma p/\rho}$. Air at normal conditions has

$$\gamma = 1.4,\ \ p = 1.01\times 10^5\ \mathrm{N/m^2},\ \ \rho = 1.23\ \mathrm{kg/m^3}, \tag{22}$$

which gives $v = 339$ m/s, in accordance with direct measurements.

Another case is electromagnetic radiation. You are supposed to know this from the course Electromagnetism. Starting from Maxwell's equations, one can derive the wave equation (19) for the transverse electric field (or also for the transverse magnetic field). The velocity is

$$v = \frac{1}{\sqrt{\epsilon\mu}}. \tag{23}$$

Here $\epsilon = \epsilon_r\epsilon_0$ is the permittivity and $\mu = \mu_r\mu_0$ the permeability of the material. In vacuum $\epsilon_r = \mu_r = 1$, so that the *velocity of light in vacuum*

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}}. \tag{24}$$

1.4 Energy of a wave

(YF 15.5) What is transported in a wave? For example, consider a sinusoidal wave (1) in the stretched string. There is no obvious transport of mass since the there is equal amount of motion up and down. However, there is transport of energy. We can calculate it by considering the work done, at some location x by the left hand side on the right hand side. The force on the right hand side is given by (13)

$$F_y = -F\frac{\partial y}{\partial x}(x,t).$$

This force does work $dW = F_y dy$ and the power is

$$P(x,t) = \frac{dW}{dt} = F_y\frac{\partial y}{\partial t} = -F\frac{\partial y}{\partial x}\frac{\partial y}{\partial t}. \tag{25}$$

This expression is valid for any wave form. Applying it to the sinusoidal wave (1) gives

$$P(x,t) = -F\frac{\partial y}{\partial x}\frac{\partial y}{\partial t} = Fk\omega A^2\sin^2(kx-\omega t). \tag{26}$$

We see that this is always non-negative (for positive k and ω). That is, energy is transported in the direction of the wave. The average power can be calculated by noticing that the time average of $\sin^2(kx-\omega t)$ is $\frac{1}{2}$. Using also (5) and (18) gives the average power

$$P_{\mathrm{av}} = \frac{1}{2}\sqrt{\mu F}\omega^2 A^2. \tag{27}$$

The quantity $\sqrt{\mu F}$ is known as *characteristic impedance* Similar calculation can be done for a sound wave. For sound we can define the *intensity*, which is the average power P_{av} per cross sectional area S of the sound wave. The result is (YF 16.3)

$$I = \frac{P_{\mathrm{av}}}{S} = \frac{1}{2}\sqrt{\rho B}\omega^2 A^2, \tag{28}$$

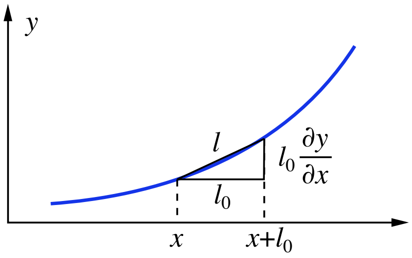
where A is the amplitude of the longitudinal oscillation. For electromagnetic radiation (YF 32.4)

$$I = \frac{P_{\mathrm{av}}}{S} = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}A^2, \tag{29}$$

where A is the amplitude of the linearly polarized electric field. The noteworthy feature in all (27)-(29) is that the power is proportional to square of the amplitude.

From the examples above we can see that for each type of wave, there are two parameters that describe the medium. For the stretched string they are the mass density μ and tension F. For the sound wave they are ρ and B. For electromagnetic wave they are ϵ and μ (or more symmetric with the two other cases, $1/\mu$). These determine both the wave velocity and (together with amplitude and frequency) the energy flow. Let us try to understand why it is so. For that, let us again consider the stretched string as an example.

We consider a short section between x and $x+l_0$ of a stretched string.



In the small amplitude approximation, the motion is purely transverse. The length of the section of the string is

$$\begin{aligned} l &= \sqrt{l_0^2 + [y(x+l_0)-y(x)]^2} \approx \sqrt{l_0^2 + \left(l_0\frac{\partial y}{\partial x}\right)^2} \\ &= l_0\sqrt{1+\left(\frac{\partial y}{\partial x}\right)^2} \approx l_0\left[1+\frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right]. \end{aligned} \tag{30}$$

Here the last approximation is based on Taylor expansion, which applied to $(1+x)^n$ gives $(1+x)^n \approx 1+nx$ for $x \ll 1$. The length increase by $\Delta l = \frac{1}{2}l_0(\partial y/\partial x)^2$ implies additional potential energy $F\Delta l$ stored to the section by the tension force F. The total potential energy is the sum of all sections. We now take the limit that the section length is made infinitesimal, $l_0 \to dx$. This means the sum is replaced by the integral,

$$E_{\mathrm{pot}} = \int \frac{1}{2}F\left(\frac{\partial y}{\partial x}\right)^2 dx. \tag{31}$$

The integration is over the length of the string.

There is also kinetic energy. The mass of the section is μl_0 and it is moving with velocity $v_y = \partial y/\partial t$ giving the kinetic energy $\frac{1}{2}\mu l_0(\partial y/\partial t)^2$. We again make the section length infinitesimal, so that the total kinetic energy

$$E_{\mathrm{kin}} = \int \frac{1}{2}\mu\left(\frac{\partial y}{\partial t}\right)^2 dx. \tag{32}$$

The total energy is

$$E = E_{\mathrm{kin}} + E_{\mathrm{pot}} = \int \left[\frac{1}{2}\mu\left(\frac{\partial y}{\partial t}\right)^2 + \frac{1}{2}F\left(\frac{\partial y}{\partial x}\right)^2\right] dx. \tag{33}$$

The quantity in square parenthesis can be interpreted as energy density (energy per length).

We now see that the quantities μ and F in the wave speed (18) and energy flow (27) are the coefficients of the kinetic and potential energies in (33). Precisely analogous result can be obtained for the longitudinal sound wave with coefficients ρ and B. The electromagnetic field is slightly different because there one cannot call the terms kinetic and potential energies, but ϵ and $1/\mu$ are the coefficients of electric and magnetic field energies.

The energy (33) allows to get the following results. 1) The kinetic and potential energy densities in a wave are always equal. 2) The energy transfer (power) in a wave is equal to wave velocity times the energy density. These results are valid for an arbitrary wave form (4). Verify these as an exercise.

Finally, we point out that in addition to energy, a wave carries also momentum and mass. In order to see this would require deeper analysis of the wave than we want to do in this course.

1.5 Transmission and reflection

(YF 15.6-8) Above we have considered waves in a homogeneous medium. Let us now look at what happens at the boundary of two different media. We again use the stretched string as an example of results which are common for all wave motion.

We assume medium a at $x<0$, medium b at $x>0$ and their boundary at $x=0$. (figure) Consider an incident wave from medium a. In general, part of the wave is *transmitted* to medium b and part is *reflected* back to a at the boundary. A medium is described by the material coefficients μ and F in (33), which generally have different values, μ_a and F_a and μ_b and F_b, in the two media. (Having $F_a \neq F_b$ requires some arrangement to prevent the boundary point from moving in the x direction. Let us not worry about this experimental problem here.) As a consequence, the speed of the transmitted wave $v_b = \sqrt{F_b/\mu_b}$ will be different from the speed $v_a = \sqrt{F_a/\mu_a}$ of the incident and reflected waves. Assuming the incident wave is sinusoidal (1), the transmitted and reflected waves are also sinusoidal with the same angular frequency ω. This implies by (5) that the wave numbers k_a and k_b (and thus the wave lengths λ_a and λ_b) are different.

In order to determine the amplitudes of the transmitted and reflected waves, one needs *boundary conditions.* A reasonable boundary condition for the string is that it is continuous. That is, $y(0^-,t) = y(0^+,t)$ for all t. Here 0^- and 0^+ denote a small negative and positive quantity, respectively. A second boundary condition is that y component of the force (12) is continuos. That is, $F_a(\partial y/\partial x)(0^-,t) = F_b(\partial y/\partial x)(0^+,t)$ for all t. Otherwise, the boundary point with infinitesimal mass would have infinite acceleration. These two conditions should be enough to determine the amplitudes and phases of the transmitted and reflected waves, for a given amplitude of the incident wave. In fact, you have done the corresponding calculation in the Electromagnetism course.

Instead of analyzing the general case, we consider two limiting case. If there is no change of material, there is no reflection and everything is transmitted. The opposite extreme is that medium b is very heavy, $\mu \to \infty$. Then we expect that nothing is transmitted leading to the effective boundary condition $y(0,t) = 0$. In order to achieve this for an incident wave $A\cos(kx - \omega t)$, the reflected wave is $-A\cos(kx + \omega t)$. The total wave in medium a is

$$A\cos(kx - \omega t) - A\cos(kx + \omega t) = 2A\sin(kx)\sin(\omega t). \tag{34}$$

This is known as *standing wave*, since the time dependence appears only as multiplicative factor $\sin(\omega t)$. (figure) At locations $\sin(kx) = \pm 1$, $x = \pi(n + \frac{1}{2})/\lambda$ (with integer n) the two traveling waves propagating in opposite directions are in phase and amplify each other. At locations $\sin(kx) = 0$, $x = \pi n/\lambda$, they are in opposite phase, and cancel each other. The standing wave is an example of *interference*, which we will discuss soon more.

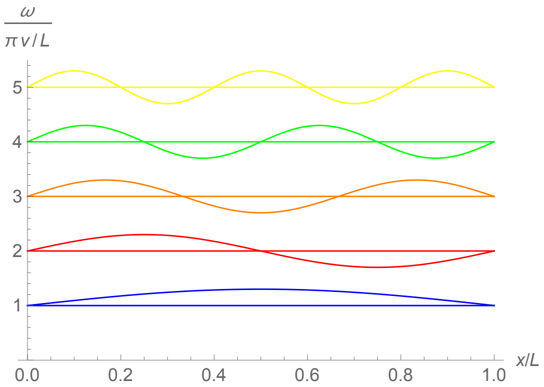
Consider a string of length L with rigid boundary conditions, $y(0,t) = 0$ and $y(L,t) = 0$. For example, a guitar string. This allows standing waves with only discrete wave numbers

$$k = \frac{\pi n}{L}, \quad n = 1, 2, 3, \ldots \tag{35}$$

As a result, the frequencies

$$\omega = v k_n = \frac{\pi n v}{L}, \quad n = 1, 2, 3, \ldots \tag{36}$$

are also discrete. These are called *eigenmodes*,



The frequencies and the wave forms of the five lowest eigenmodes in a string of length L.

1.6 Waves in 3 dimension

(YF 33.1) Above we has discussed waves of type (1), where the dependence is only one spatial coordinate x. Such a wave in 3 dimensional (3D) space is called *plane wave.* That is, using the rectangular coordinates x, y and z, the wave is same (independent of the coordinates) in the y-z plane. We can easily consider a plane wave moving in an arbitrary direction, described by unit vector $\hat{\boldsymbol{n}}$. We define the *wave vector*

$$\boldsymbol{k} = k\hat{\boldsymbol{n}} = \frac{2\pi}{\lambda}\hat{\boldsymbol{n}} \tag{37}$$

Instead of (1) we write the plane wave as

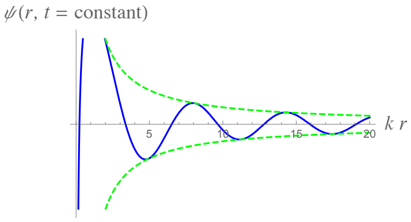
$$\psi(\boldsymbol{r},t) = A\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t). \tag{38}$$

Here $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$ is the location vector and $\cdot$ denotes the scalar product of the two vectors. [Verify that (1) is a special case of this.] Note instead of y we now use symbol ψ for the wave field.

By the principle of superposition, we can form more general wave states by summing plane waves of different $\hat{\boldsymbol{n}}$ and ω. In fact, an arbitrary wave state can be represented as a sum of plane waves. The mathematical formalism for that is known as *Fourier analysis.* In some cases it is advantegeous to use some other wave forms than plane waves. For example, a sinusoidal wave emanating from a point source (at origin) can be represented

$$\psi(\boldsymbol{r},t) = \frac{C}{r}\cos(kr - \omega t), \tag{39}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the radial coordinate. The point to notice here is that the amplitude is not constant, but decays as one over the distance, $A(r) = C/r$. (Verify that this is consistent with energy conservation.) At sufficiently small region outside the origin, the spherical wave (39) can be approximated by a plane wave.



The mathematics of waves in 3D can be based on 3D version of the wave equation (19). It is

$$\frac{\partial^2\psi}{\partial t^2} - v^2\nabla^2\psi = 0. \tag{40}$$

Here ∇^2 is the Laplace operator, which is the square of the gradient operator

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}, \tag{41}$$

$$\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{42}$$

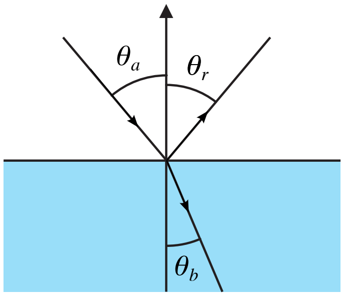
[exercise: check that the plane wave (38) and the spherical wave (39) are solutions of the wave equation (40).]

An additional thing that one has to consider for transverse waves is *polarization*. For example, electromagnetic radiation is oscillation of the electric field $\boldsymbol{E}$ (and magnetic field) perpendicular to the propagation direction, $\boldsymbol{E}(\boldsymbol{r}, t) \perp \boldsymbol{k}$. For $\boldsymbol{k} \parallel \hat{\boldsymbol{x}}$, there are two independent waves (of the same frequency) with $\boldsymbol{E}$ either in the y or in the z direction. By the principle of superposition, any other combination of the two polarizations is also possible.

1.7 Reflection and refraction

(YF 33.2, 33.5, 33.7) Consider again two media a and b, but with an oblique incidence to a plane surface. We are most familiar with this case in connection with visible light. Thus we use below language that is most common for light, for example "ray". In spite of that, the physics applies to all forms of wave motion.

We define the angles of incidence θ_a, reflection θ_r and *refraction* θ_b. The last one is the angle of the transmitted wave. These angles are measured between the ray and the surface normal of the plane boundary.



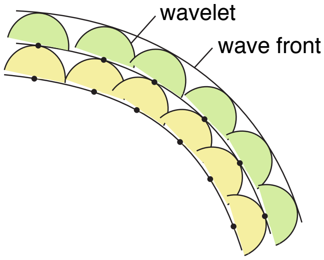
Long before the nature of light was understood, the following results were obtained experimentally.

1. the incident, reflected and transmitted rays are all in the same plane as the surface normal.
2. the reflection angle is equal to the incident angle, $\theta_r = \theta_a$.
3. the refracted and incident angles satisfy *law of refraction* or *Snell's law*

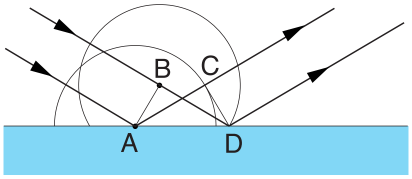
$$n_b \sin\theta_b = n_a \sin\theta_a. \tag{43}$$

   Here the constants n_a and n_b, known as *refractive indices*, depend on the nature of the two media. (For their frequency dependence, see later.)

These laws can be understood based on the wave nature of light. As an intermediate step, we introduce *Huygen's* principle. According to it, a wave propagates in the form of wave fronts. The fronts are formed by points $\boldsymbol{r}$ on which the wave has the same phase. It states that every point on the wave front, serves as a starting point of a new wavelet that that spreads in all directions with the local wave velocity. After some time Δt, the new wave front is obtained as the envelope of the wavelets. Effectively, the wavelets are summed over, and in all other directions except the front direction, the wavelets cancel each other.



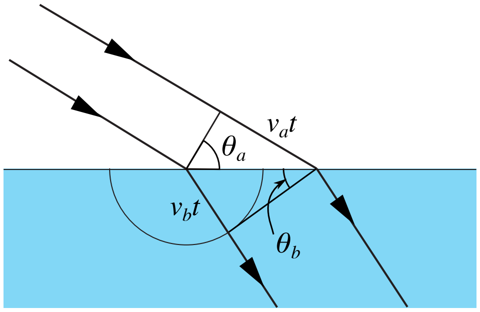
The Huygen's principle allows to understand the law of reflection. In the figure below the lines AC and BD are of equal length.



The Huygen's principle also allows to understand the law of refraction, under the supposition that the speed of the wave in a material is inversely proportional to the refractive index. More precisely, for electromagnetic radiation we write the wave speed (23) using the refractive index as

$$v = \frac{c}{n} \tag{44}$$

Comparing with (23) we get $n = \sqrt{\epsilon_r\mu_r}$. (For transparent materials $\mu_r = 1$ to a good approximation so that we can simplify $n = \sqrt{\epsilon_r}$.) For example, air has $n = 1.00$ and water $n = 1.33$ at normal conditions. The fact that the law of refraction follows from (44) and Huygen's principle is conveniently presented by drawing the figure below.

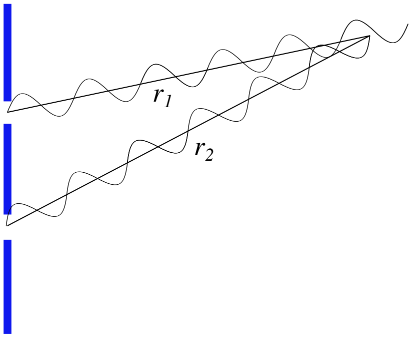


There are several consequences of the law of refraction. For example, the water in a pool seems shallower than it is. Total internal reflection, lenses.

The Huygen's principle is a convenient tool to understand wave propagation. It can be derived mathematically from the wave equation (40). Based on the wave equation and on boundary conditions, one can also find out the amplitudes of the reflected and refracted waves. For electromagnetic radiation one needs to consider also polarization. In general, the reflection is stronger for polarization that is perpendicular to the incident plane including the surface normal. Filtering such polarization is used in sun glasses. At a special incident angle (Brewster angle), the reflected wave is completely linearly polarized. (This happens when $\theta_b + \theta_r = \pi/2$.)

1.8 Interference

(YF 16.6, 35.1-4) Consider two coherent sources of waves. In the acoustic case, this could be two speakers driven by the same sinusoidal electric signal. With electromagnetic radiation in the optical range, it could be two narrow holes or slits in a screen, which are illuminated by a single monochromatic light source behind the screen. Let us consider a point that has distance r_1 and r_2 to the two sources.



At this point the wave is superposition of the partial waves originating from the two sources,

$$\psi = \psi_1 + \psi_2 = A_1 \cos(kr_1 - \omega t) + A_2 \cos(kr_2 - \omega t). \quad (45)$$

Supposing the amplitudes at the sources are the same and the distance difference is small, $r_1 - r_2 \ll r_1, r_2$, we can approximate $A_1 = A_2 = A$.

In order to proceed it is simplest to express the phases of the two wave components

$$\phi_1 = kr_1 - \omega t, \quad \phi_2 = kr_2 - \omega t \quad (46)$$

using their difference and average values,

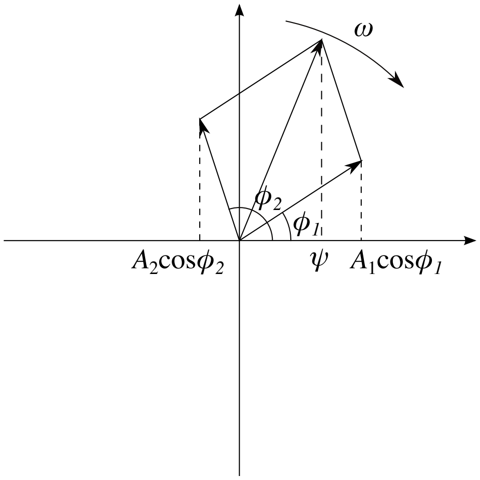
$$\begin{aligned} \phi &= \phi_2 - \phi_1 = k(r_2 - r_1) \\ \Phi &= \frac{\phi_2 + \phi_1}{2} = \frac{k(r_1 + r_2)}{2} - \omega t, \\ \psi &= A[\cos(\Phi - \phi/2) + \cos(\Phi + \phi/2)] \quad (47) \end{aligned}$$

Using the trigonometric identity
$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ we get

$$\begin{aligned} \psi &= 2A \cos\frac{\phi}{2} \cos\Phi \\ &= 2A \cos\frac{k(r_2 - r_1)}{2} \cos\left(\frac{k(r_1 + r_2)}{2} - \omega t\right). \quad (48) \end{aligned}$$

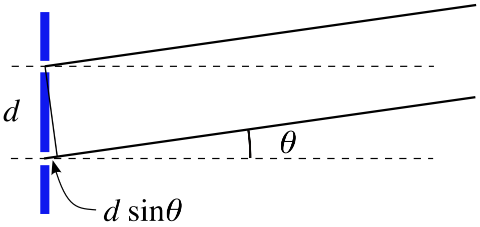
We can interpret the result as follows. The latter cosine factor in (48) oscillates sinusoidally in time, whereas the former cosine is time independent. Thus we can interpret the factor $2A\cos(\phi/2)$ as the amplitude of the oscillation. Depending on the phase difference ϕ, the absolute value of the amplitude can vary between $2A$ and 0. [If the amplitude is negative, the minus sign means a π shift in phase, $-\cos\Phi = \cos(\Phi + \pi)$.] The maximum value $2A$ is obtained with phase difference $\phi = 0 \pmod{2\pi}$. This is called *constructive interference*, where the partial waves are in phase. The opposite extreme of zero amplitude is obtained with phase difference $\phi = \pi \pmod{2\pi}$. This is called *destructive interference*, where the partial waves have "opposite phases".

The above can be visualized by a *phasor diagram.* There one draws 2D vectors with components $(A_1\cos\phi_1, A_1\sin\phi_1)$ and $(A_2\cos\phi_2, A_2\sin\phi_2)$. The resulting wave is constructed as vector sum of the two partial waves, and its amplitude depends on the relative angle of the partial waves. All the phasors rotate in the plane at angular frequency ω.



Note that the intensity of the wave is proportional to square of the amplitude [(27)-(29)]. The same result holds also for the energy density in the wave. Thus at constructive interference of two sources, the intensity if four times the intensity resulting from a single source, and the intensity vanishes at destructive interference.

Let us look the interference from the point of distance difference $r_2 - r_1$. We see from above that $\phi = k(r_2 - r_1) = (2\pi/\lambda)(r_2 - r_1)$. Thus constructive interference means that the path difference is an integer number n of wave lengths, $r_2 - r_1 = n\lambda$. Destructive interference means that the path difference is a half integer number of wave lengths, $r_2 - r_1 = (n + \frac{1}{2})\lambda$. If the distance between the two sources is d, and we are measuring at a large distance compared to d, $r_2 - r_1 = d \sin\theta$, where θ is the angle of the observation point from the normal of plane at the two slits.

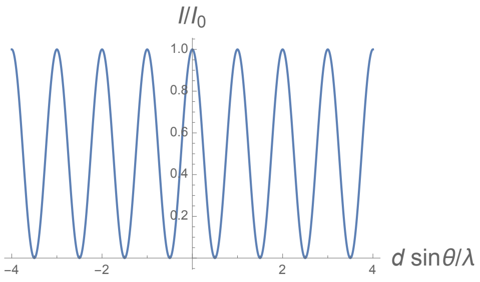


Thus maxima of the intensity are obtained at

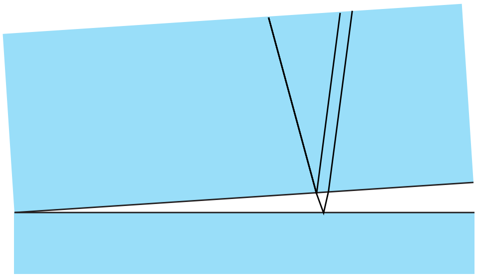
$$d \sin\theta = n\lambda, \quad n = 0, 1, 2, 3, \ldots \tag{49}$$

and minima at half integer values. The intensity is

$$I = I_0 \cos^2 \frac{k(r_2 - r_1)}{2} = I_0 \cos^2 \left(\frac{\pi d \sin\theta}{\lambda} \right), \tag{50}$$



As another example of interference, we discuss reflection of light from a thin film. You have seen this as colors when light is reflected from an oil film on water surface or from soap bubbles, for example. Let us consider the following simple case: a narrow horizontal wedge of air between two glass plates is illuminated from straight above by monochromatic light.



We study the interference of light reflected from the upper and lower surfaces. Their distance is denoted by d. Thus the ray reflecting from the lower surface travels an extra distance $2d$. We thus could expect that there is constructive interference between the two reflections when $2d$ equals an integer number of wave lengths. This would mean bright reflection at the thin edge of the wedge, where $d \to 0$. Doing the experiment gives the opposite result!

In order to understand this result, we recall the discussion of reflection leading to equation (34). We saw that reflection from rigid wall led that the reflected wave had negative amplitude (compared to the incident amplitude), $A_r = -A_i$. As pointed out in connection of (10), another way to express this is that the there is a phase shift by π. In the optical regime this phase shift appears when light is reflected from a material that has larger refractive index, but not in the opposite case. In the example above, there will be π phase shift from the reflection at the lower surface but not at the upper. Thus the condition for constructive interference is $2d = (n + \frac{1}{2})\lambda$ with integer n.

What is not reflected is transmitted (as we neglect dissipation). Making layers of different materials one can tailor the reflection and transmission. This can be done for one wavelength, but it is more difficult to achieve for a range of wave lengths.

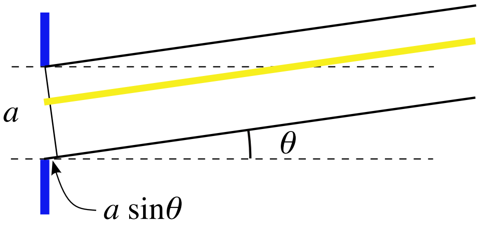
Above we have considered a sharp boundary between two media. As alternative, one can consider a gradual change of the parameters describing the medium. In this case, there is only little reflection. For example, waves on a gently sloping beach.

These considerations can be generalized to 3D. For example, a sound wave in a tube is partly reflected back at the open end of the tube. In order to make reflection less, the opening can be made gradual, as in a horn.

1.9 Diffraction

(YF 36.1-3, 36.5) By diffraction we mean interference in the case that there are more than just a few interfering waves. A classic example is the shadow formed behind a straight sharp edge. Instead of a sharp shadow, there are several intensity maxima and minima. (figure) The diffraction pattern depends on the distance of the screen from the edge. Such a case is known as *Fresnel diffraction*. In contrast, when the screen is far so that the rays from the object can be considered parallel, is called *Fraunhofer diffraction*.

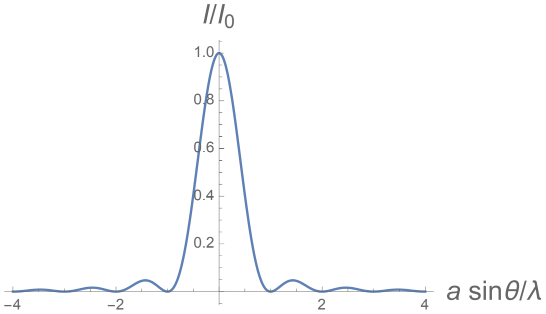
Another case is a slit, which is narrow but not small compared to the wave length. Let a be the width of the slit and θ the observation angle.



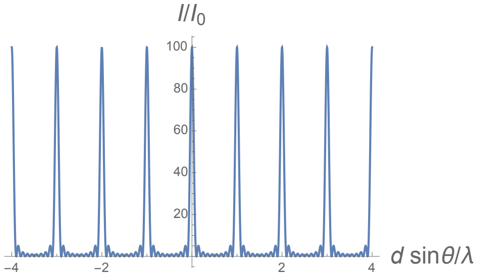
The diffraction can be understood by interference of light coming via different slices of the slit. We consider the special case of $a \sin\theta = \lambda$. In this case the slices at the lower edge and at the center are in opposite phases ($r_2 - r_1 = \lambda/2$), and they interfere destructively. The same applies to any two slices that are equally shifted upwards from the lower edge and the center. Thus there will be intensity minimum at this angle. More generally there are minima

$$a \sin\theta = n\lambda, \quad n = 1, 2, 3, \ldots \tag{51}$$

There is a central maximum at $\theta = 0$ and smaller maxima at angles between the minima (51). The whole intensity profile could be determined by analyzing the phasor diagram.



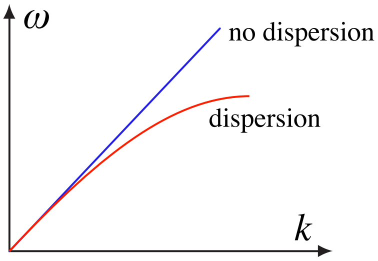
An important application of diffraction is diffraction grating, which has a large number of narrow slits. For this the maxima obey the same condition (49) as for two slits. The difference to two slits is that at intermediate values of θ, there is always destructive interference between the slits. Thus one observes only sharp peaks at values given by (49).



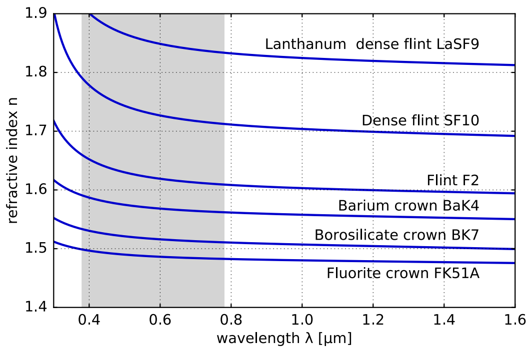
The figure is calculated for diffraction grating with $N = 10$ slits, and the maximal intensity is N^2 times the single slit maximum intensity.

1.10 Dispersion

(YF 16.6, 33.4) Let us recall the sinusoidal wave (1). It has angular frequency ω and wave number k. For a given k there is a definite ω, depending on the medium. It is useful to draw a plot with k on the horizontal axis and ω on the vertical. According to (5) the velocity of the wave is $v = \omega/k$. In the k-ω plane this is the slope of the line from the origin to the point (k, ω). We found that in several media v is a constant (independent of k). Examples are transverse waves in a string, sound waves and electromagnetic waves in vacuum. This means that all points (k, ω) for such a medium are on the same straight line, $\omega = vk$.



In some cases, however, a more general dependence $\omega(k)$ is observed. One example is light in materials, where the refractive index n depends on k. Thus $\omega(k) = ck/n(k)$ is not a straight line. This is known as *dispersion*.



The figure shows the dependence of the refractive index on the wave length for some materials. (λ in the figure refers to the wave length in air, likely obtained using a diffraction grating. The wave length in the material would be shorter, by how much?) The visible range of light is shown by shading. (Figure by Geek3 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=65418142)

Because of dispersion, the angle of refraction in Snell's law (43) is different for different frequencies. This makes white light to separate in different colors in a prism (a triangular piece made of dispersive material) and in water droplets, as seen in a rainbow.

Another commonly observed case of dispersive waves are waves on the surface of water. Mechanical waves in a solid rod and and high frequency ($\sim 10^{13}$ Hz) oscillations in crystals are also dispersive.

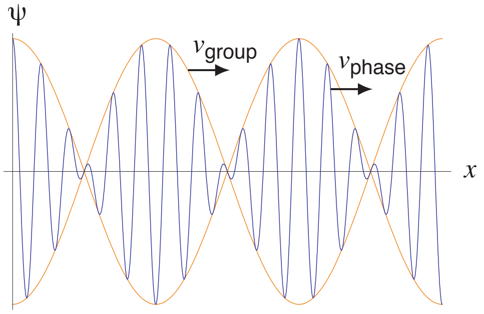
The fact that different sinusoidal waves travel at different velocities means that a wave form does not preserve its shape during propagation, contraty to (4). For example, we consider a superposition of two waves of different k and ω.

$$\psi = \psi_1 + \psi_2 = A\cos(k_1 x - \omega_1 t) + A\cos(k_2 x - \omega_2 t). \tag{52}$$

We do the same mathematics as in the case of interference (48) and obtain

$$\begin{aligned} \psi &= 2A\cos\frac{(k_2-k_1)x - (\omega_2-\omega_1)t}{2} \\ &\quad \times \cos\frac{(k_1+k_2)x - (\omega_1+\omega_2)t}{2}. \end{aligned} \tag{53}$$

This is illustrated in the figure below for the case that the difference $k_2 - k_1$ is small compared to the average $(k_1 + k_2)/2$.



The blue line shows ψ. The rapid oscillation is caused by the latter cosine factor in (53). The amplitude of this oscillation is *modulated* by the slower oscillation of the former cosine factor (orange line). Such an oscillation of magnitude is known as *beating*. This is just interference of two waves with nearly the same wave vectors. Similar effect appears in the time dependence by interference of nearly the same frequencies.

Let us look at the velocities of the rapid oscillation and the modulation. From to latter cosine factor, using (4), we deduce the *phase velocity*

$$v_{\text{ph}} = \frac{\omega_1 + \omega_2}{k_1 + k_2} = \frac{\omega}{k}. \tag{54}$$

For $k_2 - k_1 \ll (k_1 + k_2)/2$, this is to a good approximation the same as we had in (5), as indicated by the latter equality in (54). The velocity of the modulation is obtained similarly from the former cosine factor in (53). This is the *group velocity*

$$v_{\text{g}} = \frac{\omega_2 - \omega_1}{k_2 - k_1} \approx \frac{d\omega(k)}{dk}. \tag{55}$$

We thus have two velocities, (54) and (55). In the the absence of dispersion, $\omega = vk$, the two velocities are identical. In the presence of dispersion they generally are different. The difference means that the wave in the figure above does not move rigidly, but is continuously evolving as the speed of the fast oscillation is different from the speed of the modulation.

The significance of the group velocity is that it gives the speed of a pulse. The superposition of two sinusoidal waves gave the beating. By including more waves in the superposition, it is possible to form a train of localized pulses. A single pulse, where the oscillation is limited to a finite region of length and time, can be represented as an integral

$$\psi(x,t) = \Re\int_{k_1}^{k_2} D(k) e^{i[kx - \omega(k)t]} dk. \tag{56}$$

[For alternative representations, see (10).] Such a pulse moves with the group velocity (55). Because of dispersion the shape of the pulse also changes as time elapses.

An example of dispersion are gravity waves (waves on water surface with $\lambda \gg 2$ cm in deep water), where $v_{\text{g}} = \frac{1}{2} v_{\text{ph}}$. By this fact alone, one can explain several peculiar features of waves generated by a boat/ship.

2. Particle properties of light

In the preceding section we have discussed wave motion. We have seen that light has several properties of wave motion, in particular, interference and diffraction. These would be difficult, or impossible, to explain imaging light consisting only particles. Even such a simple thing as partial transmission and reflection, could not be understood by Newton's mechanics by thinking monochromatic light as flow of particles of the same energy. Namely, if a classical particle hits a potential barrier, it is always either reflected (if it kinetic energy is lower than the change in potential energy) or transmitted (if it kinetic energy is higher than the change in potential energy). (figure) Therefore, we have to conclude that light is a wave.

The wave nature of light, however, cannot be the whole story. In the following we discuss properties of light, that are difficult to understand thinking light as a wave only.

2.1 Photoelectric effet

(YF 38.1) We study the *photoelectric effect*. When light is shined on a material surface, electrons are emitted from the surface. In order to study that, we can make an experiment with two electrodes (made of conducting material). (YF figure 38.3) One electrode, the cathode, is lighted and the other, anode, is used to collect the electrons. In order to avoid electron collision with gas molecules, this is done in a vacuum tube. We apply a voltage $V_{\rm AC}$ between the anode and cathode electrodes and measure the electric current. The electric current is called *photocurrent*, as it is caused by light absorption. With positive $V_{\rm AC}$, all the electrons emitted from the cathode are expected to be collected to the anode. For negative $V_{\rm AC}$ there is force on the electrons to return the electrons to the cathode. This allows to determine the maximal kinetic energy of the emitted electrons, $K_{\rm max} = \frac{1}{2}mv^2 = eV_0$, where $V_{\rm AC} = -V_0$ is the voltage at which the current stops. Here V_0 is called *stopping potential*. We can repeat the experiment at different amplitudes and frequencies of the light.

Obviously, it should be the electric field of the electromagnetic wave that should accelerate the electrons in the cathode material, so that they can escape to the vacuum. According to wave theory of light we could make the following predictions

1. The photocurrent is expected to depend on the light intensity (29), but there is no obvious dependence on frequency.
2. For very low intensity of light, it should take longer before the electron is accelerated sufficiently for the escape.
3. With increasing intensity, the electrons are expected to accelerated more, so the stopping potential should depend on the intensity.

What really is observed in the experiment is as follows.

1. The photocurrent depends on the light intensity (29). In addition, it depends on the frequency so that it does not take below a threshold frequency.
2. For very low intensity of light, the current starts immediately.
3. The stopping potential is independent on the intensity (YF fig 38.4). It depends linearly on the frequency of light (YF fig 28.5).

These results were obtained since the first observation of the photoelectric effect in 1877 and they remained a mystery until 1905. In the latter year, Einstein proposed a theory that could account for the observations. Let us now look at that.

We start the story slightly earlier. Maxwell had published his equations in 1860's. In late 1800's one of the problems with electromagnetic radiation was to understand emission from hot bodies. The emission intensity was measured as a function of wave length (YF fig 39.32). It was found that the maximum intensity was at a definite wave length, that depends on the temperature. The classical theory seemed to be in agreement with observation at long wave lengths. The problem was that the classical theory would predict the emission to increase monotonically with decreasing wave length. In 1900 Max Planck suggested a mathematical expression that could fit the whole dependence of intensity on λ. His expression could be interpreted that the energy of the radiation is not a continuous quantity, but consist of packets. They are called *photons* or *quanta*. The energy of a single photon is

$$E = hf. \tag{57}$$

Here f is the frequency of the electromagnetic field, and h the *Planck's constant*,

$$h = 6.63 \times 10^{-34} \text{ Js}. \tag{58}$$

What Einstein realized was that the photon concept gives explanation for the problems presented above. An electron takes the energy of a photon. If this energy is larger than *work function* ϕ of the material, the electron can escape to vacuum, and is counted as photocurrent. This condition $hf > \phi$ explains the threshold frequency $f_t = \phi/h$. The intensity dependence comes that there are more photons to do the same for more electrons.

The possibility of a photon absorption starts immediately as the light can arrive on the cathode, and thus there is no delay.

The energy of the photon goes to the work function and the rest to the kinetic energy of the electron, $hf = \phi + K_{\max}$. Thus the stopping potential $eV_0 = K_{\max} = hf - \phi$ is independent of the light intensity. We can also plot it as a function of frequency (YF fig 38.5). It is linear function intersecting zero at the threshold frequency.

Photon momentum

A photon can be considered as particle with zero mass. As an energy packet of electromagnetic field, it propagates in vacuum at the velocity of light c (24). We claim that in addition to energy, a photon carries also momentum. Its magnitude in vacuum is

$$p = \frac{hf}{c}. \tag{59}$$

Recalling $c = f\lambda$ (5), this can also be written

$$p = \frac{h}{\lambda}. \tag{60}$$

The direction of the momentum is the same as the propagation direction of the electromagnetic wave, $\hat{\boldsymbol{n}}$. Thus the momentum vector $\boldsymbol{p} = (h/\lambda)\hat{\boldsymbol{n}}$.

[We could try to justify (60) as follows. 1) Theory of relativity requires this. 2) Our wave example, transverse wave in a stretched string, has $p = E/v$. Unfortunately, both justifications go beyond the scope of this course.]

2.2 Additional evidence of photons

By now, there is a lot of evidence of photons. The following cases, although historic, are still valuable to know today.

Generation of x rays

(YF 38.2) X-rays are electromagnetic radiation with rather short wave length, λ in the range $10^{-9} \ldots 10^{-12}$ m. They were first found by Wilhelm Röntgen in 1895. He used a device that in basic principle is the same as modern x-ray tubes (YF fig 38.7). It has a heated cathode, from where electrons escape to vacuum. They are accelerated by a high voltage of several thousand volts. Then they hit the anode electrode. In slowing down, they emit electromagnetic radiation, known as *bremsstrahlung* (German, brems for braking and strahlung for radiation). It has a continuous spectrum, and large part of it is in the x-ray range. It is found that the spectrum has a minimum wave length $\lambda_{\min}$ below which no radiation is observed.

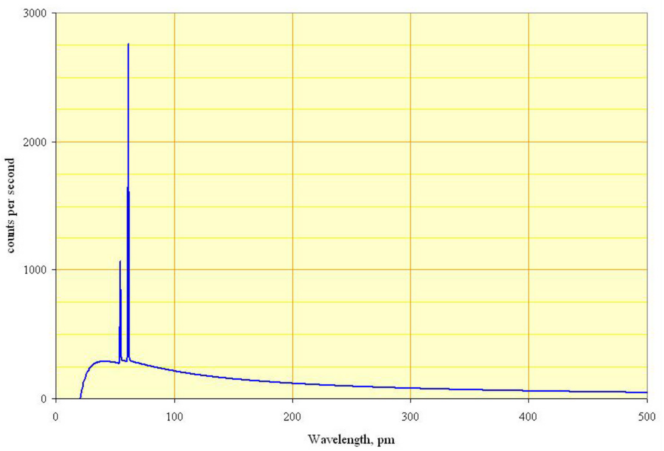


Figure: Spectrum of the X-rays emitted by an X-ray tube with a rhodium target, operated at 60 kV. The smooth, continuous curve is due to bremsstrahlung, and the spikes are characteristic K lines for rhodium atoms. (source Wikipedia)

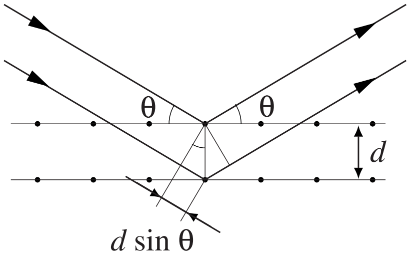
Based on Maxwell's equations one can derive that accelerated charges produce electromagnetic radiation. This should apply to the rapid deceleration of electrons in the anode material. The classical explanation does not produce any clear-cut explanation for the minimum wave length.

In the photon view, an electron looses its kinetic energy by emitting one or more photons. The minimum wave length corresponds to emission of one photon only,

$$eV_{\mathrm{AC}} = K = hf_{\max} = \frac{hc}{\lambda_{\min}}. \tag{61}$$

This formula for $\lambda_{\min}$ is in agreement with experiments. (The initial kinetic energy of the electron is negligible here.) This is strong evidence of the photon picture for emission. (In photoelectric effect we considered absorption.)

(YF 36.6) Side topic: How can we measure the wave length of x rays? It is too short for artificial diffraction gratings. But λ is of the same order of magnitude as the distance between atoms in a crystal. Since x rays penetrate well inside a crystal, we can get reflections from consecutive atomic layers inside the crystal.



The reflections from two neighboring layers with distance d are in phase when the path difference $2d\sin\theta$ equals an integer number of wave lengths. Similar to an ordinary diffraction grating, in all other angles there is destructive interference. Thus a strong reflection is observed when

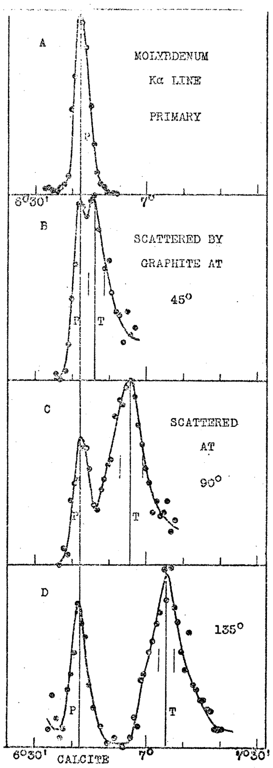
$$2d\sin\theta = n\lambda,\ \ n = 1, 2, 3, \dots \tag{62}$$

This is known as *Bragg condition.* Measuring reflection angles θ and knowing d one can deduce λ. (Note that the definition of the angle θ differs from the one used in the optical case.)

Compton scattering

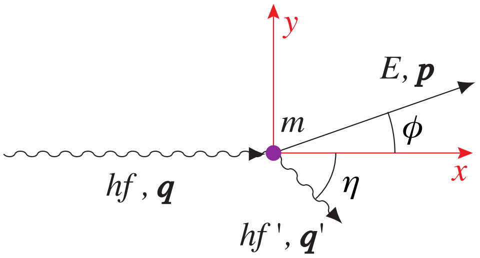
(YF 38.3) Consider scattering of electromagnetic wave from a free electron. In the classical picture the electron would oscillate in the electric field. In oscillation motion it will radiate electromagnetic waves in all directions. That is scattering. In classical theory, the scattered wave has the same frequency as the incident wave.

Experimentally it is found that the frequency of the scattered wave is different from the incident one.



The figure shows the original data published by Arthur Compton [Physical Review **22**, 409 (1923)]. The incident radiation was generated by x-ray tube with Molybdenum target that produces characteristic radiation with $\lambda = 71$ pm. (fig A). The wave scatters from the electrons of graphite and is observed at three different angles η (B: 45°, C: 90°, D: 135°). The spectrum of scattered light is obtained by turning a calcite crystal ($CaCO_3$), from which the wave is reflected according to Bragg condition (62). The horizontal axis in the figures is the turning angle of the crystal θ in (62). We see peaks at the incident frequency as well as a shifted frequency.

In the photon view we consider scattering of a photon from an electron (mass m). We assume that energy and momentum are conserved in the process.



The conservation laws require the electron to take part of the energy of the initial photon. Thus less energy is left for the scattered photon, $f' < f$ and $\lambda' > \lambda$. Working out the mathematics gives

$$\lambda' - \lambda = \frac{h}{cm}(1 - \cos\eta). \tag{63}$$

The quantity $h/cm = 2.4 \times 10^{-12}$ m. This is in good agreement with experiment.

Note that this experiment also tested the photon momentum (60).

Pair production

(YF 38.3) With very high photon energies, it is possible that a photon is converted to an electron-positron pair. Positron is like an electron but positive charge $+e$. The minimum amount of energy required is the double of the rest mass of an electron,
$E_{\text{min}} = hf_{\text{min}} = ch/\lambda_{\text{max}} = 2c^2m = 1.02$ MeV, corresponding to $\lambda_{\text{max}} = 1.21$ pm. This process can occur only in material since momentum conservation requires an atom, for example, to take part of the momentum.

(YF did not explain how to measure such a small length, but apparently the energy can be measured.)

Compton scattering and pair production are important processes in shielding from gamma rays.

2.3 Wave-particle duality

(YF 38.4) Above we have seen that both the wave and the particle properties of electromagnetic radiation. Both these seemingly very different aspects are needed. This is called *Wave-particle duality.* A related concept is *principle of complementarity.* The message is that the two aspects complement each other, one or the other should be used to explain a particular experiment, but trying to use both simultaneously can lead to confusion.

In order to understand that the two aspects are not in contradiction, we should study how interference and diffraction is obtained in the particle point of view. Consider as an example the double slit experiment. The wave picture gave the interference fringes on the screen.

If instead of screen one uses a photomultiplier, it is possible to see the individual photons to arrive. The first arrives at some point. The next some other point, and third still some other point. With a few recorded quanta, there is no particular pattern to be observed. However, looking for longer time, one sees that the photons arrive mostly at the locations of constructive interference of the wave picture. The more one gets the photons, the more closely their distribution is to the intensity distribution produced by the wave picture.

Thus we see that looking macroscopic number of photons, they propagate according to the wave picture.

The experiment above brings clearly out a new element of reality, that is not present in classical mechanics. Namely, there is *randomness* in the locations where the photons arrive. This takes place even if one tries to arrange the initial conditions for each photon as similar as possible.

Uncertainty relation

We now try to understand the uncertainty of the photon motion more quantitatively. Let us consider that we make a narrow beam of light letting light go through a slit. (YF fig. 38.17) We discussed the diffraction above and found that most of the intensity (detailed calculation gives 85 %) go in the central maximum. We use θ_1 to denote the angle of the first minimum. We found in (51) that $\sin\theta_1 = \lambda/a$. We now look the case where a is much larger than λ. This means that we can approximate $\sin\theta \approx \theta$ so that

$$\theta_1 \approx \frac{\lambda}{a} \tag{64}$$

The photon has momentum $p = h/\lambda$ (60). A photon going into angle θ_1 has component of momentum in the transverse y direction,

$$p_y = p\tan\theta_1 \approx p\,\theta_1 \approx \frac{h}{a}. \tag{65}$$

Photons going to different angles θ have somewhat different values of p_y, smaller or larger. However, we can estimate that a typical change of momentum Δp_y of a photon after passing the slit is of the same order of magnitude as (65). We reach the following important estimate

$$\Delta p_y a \approx h. \tag{66}$$

We interpret this as follows: we have tried to confine the photons so that the width of the beam would be a (in the y direction). This introduces uncertainty of the momentum Δp_y (in the y direction). According to (66), the product of these uncertainties is approximately the Planck's constant h (58). The smaller the uncertainty in the coordinate ($\Delta y = a$), the larger is the uncertainty in the momentum (Δp_y), and vice versa.

The uncertainties can be defined more precisely than above. For that one uses standard deviation. Then one can, instead of the estimate (66), prove the uncertainty relation

$$\boxed{\Delta x\,\Delta p_x \geq \frac{\hbar}{2}.} \tag{67}$$

Here we have defined a quantity, also known as Planck's constant, or "h-bar",

$$\hbar = \frac{h}{2\pi} = 1.05\times 10^{-34}\text{ Js.} \tag{68}$$

Using $\hbar$ rather than h simplifies (67) and several other formulas in the following.

Eq. (67) expresses *uncertainty principle*. It was proposed by Werner Heisenberg 1925. It says that when we measure coordinate x and the momentum p_x (in the same direction), there will always be uncertainties in these two, so that the product of their standard deviations is not smaller than $\hbar/2$. For example, in a sinusoidal wave, p is accurately known but Δx is very large. Decreasing Δx makes p less precise. We have now discussed this for photons, but later we claim this result to be valid for electrons and other particles as well.

Let us write the energy (57) and momentum (60) of a photon using ω and k. With definition (68) they take the simple form

$$E = \hbar\omega, \quad \boldsymbol{p} = \hbar\boldsymbol{k}. \tag{69}$$

The latter one is written as a vector relation since the directions of $\boldsymbol{p}$ and $\boldsymbol{k}$ are the same, and magnitudes $p = \hbar k$.

Let us form another approximate justification of the uncertainty relation (67). In section "Dispersion" we studied superposition of two waves and found a beating caused by the former cosine factor in (53). At a fixed time, one beat takes distance Δx so that $\Delta k\Delta x/2 = \pi$, where $\Delta k = k_2 - k_1$. Using the latter of (69) this gives $\Delta x\Delta p = 2\pi\hbar$ This agrees in order of magnitude with the minimum in (67).

Interestingly, we can study the same beat pattern in time, at a fixed x. Similarly as above but using the former of (69) one gets $\Delta E\Delta t = 2\pi\hbar$. More precisely, one can derive *uncertainty relation between time and energy*

$$\Delta t\,\Delta E \geq \frac{\hbar}{2}. \tag{70}$$

Eq. (70) says that when we measure the photon arrival time t and the energy E, there will always be uncertainties in these two, so that the product of their standard deviations is not smaller than $\hbar/2$. For example, in a sinusoidal wave, E is accurately known but Δt is very large. Decreasing Δt makes E less precise.

3. Wave properties of particles

We now turn to study particles, in particular, the electron and the structure of an atom.

3.1 Model of an atom

(YF 39.2-3) In late 1800's it was found that the light emitted by hot gases had certain wave lengths that are characteristic to the substance. The same wave lengths were observed also in absorption. For hydrogen, the lightest of all substances, several spectral lines were found. The wave lengths of these could be fitted by the expression

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \tag{71}$$

where $R \approx 10^7$ 1/m is Rydberg constant, $n_1 = 2$ or 3, and n_2 has integer values ($n_2 > n_1$). The lines corresponding to $n_1 = 2$ are called Balmer series (in the visible and ultraviolet region) and $n_1 = 3$ Paschen series (in infrared region).

The first observation of an electron particle was made by J.J. Thompson 1897. Using experiments with cathode ray tubes, he found negatively charged particles whose mass were less than 1/1000 of that of the hydrogen atom. The next crucial discovery to clarify the structure of an atom was by Ernest Rutherford 1910-1911. Studying scattering of alpha rays (He nuclei) from atoms, he found that the atom has a small nucleus, where the positive charge was concentrated. Based on these, Niels Bohr suggested a model of an atom 1913, which we will now study in more detail

As the positively charged nucleus is very small, the electrons are supposed to be occupying most of the volume of an atom. By Coulomb law there is electrostatic attraction between the electrons and the nucleus. In hydrogen atom the Coulomb force on the single electron is

$$F = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} \tag{72}$$

directed towards the nucleus. Here r is the radial coordinate of the electron. In order for the electron not to drop to the nucleus, it could be circulating the nucleus. For an electron moving with velocity v on a circular orbit, the centripetal acceleration $a = v^2/r$. According to Newton's law $ma = F$,

$$\frac{mv^2}{r} = \frac{1}{4\pi\epsilon_0}\frac{e^2}{r^2} \tag{73}$$

The momentum $p = mv$ and in circular orbit the angular momentum $L = rp$. We can write the condition (73) in the form

$$r = \frac{4\pi\epsilon_0}{e^2 m}L^2. \tag{74}$$

We calculate also the energy. The potential energy corresponding to the Coulomb law (72) is $U = -(1/4\pi\epsilon_0)e^2/r$. [Recall that the force (72) is the gradient of U, $\boldsymbol{F} = -\boldsymbol{\nabla} U$.] The total energy is the sum of kinetic and potential energies, and we get

$$\begin{aligned} E &= \frac{1}{2}mv^2 - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r} \\ &= -\frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\frac{1}{r} \\ &= -\frac{m}{2}\left(\frac{e^2}{4\pi\epsilon_0}\right)^2\frac{1}{L^2} \end{aligned} \tag{75}$$

We see that (74) gives the relation between the orbit radius r and the angular momentum L. Similarly (75) gives the energy corresponding L. According to classical mechanics L can have arbitrary positive values. The same applies to r and the energy absolute value $|E|$ because of (74) and (75).

The model above has a problem because the electron is charged. In the circular orbit the electron is continuously accelerating and would radiate electromagnetic waves. Such radiation would clearly be in contradiction with the observed stability of atoms.

Bohr's idea was to postulate, that instead of arbitrary E (and correspondingly r and L), only some of the energy values, called *energy levels*, are allowed. These correspond to discrete E_n, with $n = 1, 2, 3, \ldots$ and correspondingly r_n and L_n. Instead of the classical electrodynamics, there is radiation only in changes from one energy level to an other. The transition from level n to n' with $E_n > E_{n'}$ takes place by emitting a photon (57)

$$hf = E_n - E_{n'} \tag{76}$$

so that energy is conserved. This implies that in the lowest energy orbit the atom cannot radiate. Correspondingly, the absorption of a photon corresponds to change from one energy level to another, whose energy is higher by the photon energy.

Moreover, Bohr observed that if one takes the energy levels to correspond to integral values of the angular momentum,

$$L = \hbar n, \quad n = 1, 2, 3, \ldots \tag{77}$$

one can explain the measured spectrum of hydrogen (71). This should follow using formulas (75), (76) and $c = \lambda f$ [exercise: calculate the expression for R based on this.]

Besides explaining the absorption/emission lines of hydrogen, the Bohr model makes another important prediction. It is the first theory to predict the size of an atom. Substituting the lowest $L = \hbar$ into (74) gives the *Bohr radius*

$$a = \frac{4\pi\epsilon_0\hbar^2}{e^2 m} = 0.529 \times 10^{-10}\ \text{m}. \tag{78}$$

3.2 Wave nature of electron

(YF 39.1) The Bohr model of hydrogen atom was highly successful, but there was no theoretical justification for the condition (77). The important breakthrough was made by Louis de Broglie in 1924. He suggested that electron, that was considered a particle, has also wave nature. Thus he extended the wave-particle duality from photons to electrons. More quantitatively, he suggested that an electron with momentum p is associated with a wave length λ, known as *de Broglie wave length*, according to the relation

$$p = \frac{h}{\lambda}. \tag{79}$$

Also, the energy of the electron is associated with frequency by the relation

$$E = hf \tag{80}$$

Note that these relations are precisely the same as for photons in (60) and (57). These can conveniently also represented in the form (69). Let us repeat them here for electrons,

$$E = \hbar\omega, \quad \boldsymbol{p} = \hbar\boldsymbol{k}. \tag{81}$$

[Note that for photons we also had formula (59), but that is not valid for electrons.]

Although we have been speaking about electrons, the wave-particle duality above applies to all particles. This includes also composite particles like atoms and molecules.

As the first application of (79), consider the circular electron orbit. In order to get constructive interference after traversing one circle, we should have

$$2\pi r = n\lambda, \quad n = 1, 2, 3 \ldots. \tag{82}$$

Using (79), $L = rp$ and (68) gives precisely the Bohr quantization condition (77).

Besides explaining the stability and spectroscopy of atoms, there is direct evidence of matter waves. For example electron and neutron scattering from crystals is found to obey the Bragg condition (62).

4. Quantum mechanics of a single particle

4.1 Introduction

(YF 40.1) The Bohr model of an atom combined with de Broglie formula (79) explains nicely the spectroscopy of hydrogen. However, several problems remain. For example, why are only circular orbits allowed? Is there diffraction of the electron waves in an atom?

A proper study of diffraction has to be based on the wave equation. We found that several types of waves obey the wave equation, (19) in 1D or (40) in 3D. In this course we did not use directly the wave equation to study diffraction, since that would have been a too mathematical approach. Instead, we used simpler methods, like Huygen's principle, which are based on the wave equation. Also we learned a bit of the usefulness of the 3D wave equation (40) in that we were able to verify that the spherical wave (39) is a solution of the wave equation.

What we should now be looking for is some generalization of wave equation that would describe an electron. As a solution of it we would then have a wave describing an electron in an atom. Following YF, we denote the wave by Ψ (upper case ψ). This would be some kind of generalization of the spherical wave (39).

Let us start the program by writing the (nonrelativistic) energy of an electron. We write the potential energy U but, for the moment, we assume it to be a constant. We also assume motion in 1 dimension only. The energy is

$$E = \frac{1}{2}mv^2 + U = \frac{p^2}{2m} + U, \tag{83}$$

where m is the mass and v the velocity of the particle. The latter form uses momentum $p = mv$ instead of the velocity v. We now replace E and p using ω and k in (81), and get

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} + U. \tag{84}$$

This is the dispersion relation $\omega(k)$ for free electron waves. (figure) Since it is not linear, $\omega \neq ak$ (with any constant a), we conclude that electron waves are dispersive. Such waves do not obey the standard wave equation (19) or (40).

We learned in equations (7)-(9) that making partial derivatives with respect to t and x produces multiplying factors ω and k, respectively. Therefore we could test the possibility that the differential equation

$$a\frac{\partial\Psi}{\partial t} = b\frac{\partial^2\Psi}{\partial x^2} + U\Psi. \tag{85}$$

would produce the dispersion relation (84) with some constants a and b. Substituting $\Psi = A\cos(kx - \omega t)$ into

(85) reveals that it is not a solution. The problem is that on the left hand side we get a sine, but on the right hand side a cosine.

[Let us comment here that although equation (85) does not describe electron waves, it is important for other things. For $U = 0$ this equation is known as *diffusion equation*. In these applications, Ψ can be the concentration of some solute atoms in solution, or Ψ can be the temperature in heat conduction problems.]

In order to avoid the prolem above, we can assume that there are two wave functions Ψ_1 and Ψ_2. We recall the the electromagnetic field has two orthogonal polarizations, and Ψ_1 and Ψ_2 could be something analogous. We propose the pair of equations

$$\begin{aligned} a\frac{\partial \Psi_2}{\partial t} &= b\frac{\partial^2 \Psi_1}{\partial x^2} + U\Psi_1 \\ c\frac{\partial \Psi_1}{\partial t} &= d\frac{\partial^2 \Psi_2}{\partial x^2} + U\Psi_2. \end{aligned} \tag{86}$$

Now, if $\Psi_1 = A\cos(kx - \omega t)$ and $\Psi_2 = A\sin(kx - \omega t)$, all terms in the former equation are proportional to cosine, which factor can be canceled. Similarly in the latter all terms are proportional to sine, which factor can be canceled. We are left with equations

$$\begin{aligned} -a\omega &= -bk^2 + U \\ c\omega &= -dk^2 + U. \end{aligned} \tag{87}$$

We now see that dispersion relation (84) could be produced by selecting $c = -a = \hbar$, $b = d = -\hbar^2/2m$. Using these, the differential equations (86) are

$$\begin{aligned} -\hbar\frac{\partial \Psi_2}{\partial t} &= -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi_1}{\partial x^2} + U\Psi_1 \\ \hbar\frac{\partial \Psi_1}{\partial t} &= -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi_2}{\partial x^2} + U\Psi_2. \end{aligned} \tag{88}$$

Until now we have assumed that the potential U is a constant. We now make the bold generalization that equations (88) describe electron waves also for general $U(x)$. With this generalization (88) is equivalent with the equation that Erwin Schrödinger proposed in 1926 for electron waves. More precisely, this is the one-dimensional version of the equation, whose 3D version will be discussed soon.

Before going on, let us recall some properties of a complex numbers. A complex number z can be represented using two real numbers x and y, so that $z = x + iy$, where i multiplying y is the imaginary unit. $x = \Re z$ is called the real part and $y = \Im z$ the imaginary part of z. The calculations (addition, subtraction, multiplication, division) with i and complex numbers goes similarly as with real numbers but with the additional rule that $i^2 = -1$. For example, $zz' = (x + iy)(x' + iy') = xx' - yy' + i(xy' + x'y)$. From this we see that $\Re(zz') = xx' - yy'$ and $\Im(zz') = xy' + x'y$. The complex conjugate of $z = x + iy$ is marked by z^* and is given by $z^* = x - iy$. The absolute value is marked and defined by $|z| = \sqrt{zz^*} = \sqrt{x^2 + y^2}$.

Why we said "equivalent" above, is that Eqs. (88) can also be represented in another form. Above we assumed Ψ_1 and Ψ_2 to be real valued. We now define a complex valued wave function Ψ so that

$$\Psi = \Psi_1 + i\Psi_2. \tag{89}$$

That is, Ψ_1 and Ψ_2 are the real and imaginary parts of Ψ. We now claim that the pair of equations (88) can be represented as a single equation for Ψ. This is the standard form of the *Schrödinger equation* in 1D

$$\boxed{i\hbar\frac{\partial \Psi}{\partial t}(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi}{\partial x^2}(x,t) + U(x)\Psi(x,t).} \tag{90}$$

[Exercise: verify that taking the real and imaginary parts of (90) gives (88).]

4.2 Basic properties of Schrödinger equation

The Schrödinger equation (90) is our starting point to explore *quantum mechanics*. Quantum mechanics is a theory of Nature, which is based on the wave nature of particles. It replaces *classical mechanics*, by which we mean the theory based on Newton's equation of motion. For microscopic objects the predictions of quantum mechanics differ from the ones of classical mechanics. For larger particles the predictions of the two theories are the same to a good accuracy. Thus classical mechanics is still the best approach to study macroscopic problems, as using quantum mechanics would lead to unnecessary complicated calculations.

The solutions $\Psi(x,t)$ of the Schrödinger equation equation are called *wave functions* or *states*.

An important property of the Schrödinger equation (90) is that it is linear. That is, if $\Psi_1(x,t)$ and $\Psi_2(x,t)$ are two solutions of the equation, then their linear combination $a\Psi_1(x,t) + b\Psi_2(x,t)$ is a solution (where a and b are arbitrary real or complex numbers). This is known as the *principle of superposition*, similarly as for the wave equation (19). For mechanical waves we discussed that the principle is only approximate, valid in the case that the wave amplitude is not too large. In the case of Schrödinger equation the principle is exact.

Previously we used a complex wave (10) simply as a mathematical tool to represent real solutions of real equations. Now the Schrödinger equation is complex equation, as it contains i, and also the imaginary part of the complex wave is important as well. As a test, consider the case of $U =$ constant and a wave

$$\Psi(x,t) \;=\; Ae^{i(kx-\omega t)}$$

$$= \quad Ae^{ikx}e^{-i\omega t} \tag{91}$$

We call (91) a sinusoidal wave. In the second form we have used the property $e^{a+b} = e^a e^b$. In (91) the exponential is complex, $e^{i(kx-\omega t)} = \cos(kx - \omega t) + i\sin(kx - \omega t)$, but also the amplitude $A = \Re A + i\Im A$ can be complex. We substitute (91) into the Schrödinger equation (90). We find that the equation is satisfied when k and ω satisfy the dispersion relation (84). Thus it seems that (91) describes a free particle with momentum $p = \hbar k$ and energy $E = \hbar\omega$.

In oder to be more precise on the momentum, we define a *momentum operator*

$$p_{\text{op}} = \frac{\hbar}{i}\frac{\partial}{\partial x}. \tag{92}$$

We apply it to (91),

$$p_{\text{op}}\Psi(x,t) = \frac{\hbar}{i}\frac{\partial}{\partial x}Ae^{i(kx-\omega t)} = \hbar k\, Ae^{i(kx-\omega t)}. \tag{93}$$

We express the result

$$p_{\text{op}}\Psi(x,t) = \hbar k\, \Psi(x,t). \tag{94}$$

It is called that Ψ (91) is an *eigenfunction* of the momentum operator. The quantity of $\hbar k$ is called the *eigenvalue*. We see that operating with momentum operator on the wave state (91), gives the same state multiplied by the eigenvalue. The eigenvalue is the momentum of the state, $p = \hbar k$. This is the way we can know the momentum of a state. When (94) is valid, the momentum of the state Ψ is precisely defined.

We consider a superposition of two sinusoidal waves

$$\Psi = A_1 e^{i(k_1 x - \omega_1 t)} + A_2 e^{i(k_2 x - \omega_2 t)} \tag{95}$$

Applying with the momentum operator gives

$$p_{\text{op}}\Psi = \hbar k_1 A_1 e^{i(k_1 x - \omega_1 t)} + \hbar k_2 A_2 e^{i(k_2 x - \omega_2 t)} \tag{96}$$

We cannot write the result in the form (94) except in the special case $k_1 = k_2$. This means that the superposition is not an eigenstate. The momentum of the superposition state is not precisely known. It can be either $\hbar k_1$ or $\hbar k_2$.

How do we interpret the superposition wave function (95)? This is not a trivial problem, but the consensus seems to be the following. When we make a measurement of the momentum, the result is either $\hbar k_1$ or $\hbar k_2$. The measurement changes the state of the particle, as we discuss later more. In order to repeat the measurement under the same conditions, one should prepare the initial state (95) again, in the same way as it was done at the first time. Repeating the measurement seems to give randomly either $\hbar k_1$ or $\hbar k_2$. This is similar as we were measuring where the photons arrived in the two-slit experiment. One can count how often one gets the results $\hbar k_1$ and $\hbar k_2$. The counts indicate that the probabilities of the two momentum values are proportional to the amplitudes squared, $|A_1|^2$ and $|A_2|^2$, respectively. Again, this analogous to the two-slit experiment, where the probability was proportional to the intensity $I \propto A^2$. Since A_1 and A_2 can be complex valued, so can A_1^2 and A_2^2. Therefore, taking the absolute value is necessary to get real valued probabilities.

We claim that no physically measurable quantity is changed if one multiplies the wave function with a constant C. For example, the momentum eigenvalue in (94) is not dependent on A in (91). We can use this ambiguity to *normalize* the wave function. For example, in the case of the superposition (95), we can multiply the wave function by $C = 1/\sqrt{|A_1|^2 + |A_2|^2}$. The wave after multiplication has the same form (95), but the new coefficients satisfy $|A_1|^2 + |A_2|^2 = 1$. This allows to interpret the amplitude squares as probabilities $P_i = |A_i|^2$ so that they are normalized to one, $P_1 + P_2 = 1$. For example, if the initial amplitudes were real and equal, the normalized amplitudes $A_1 = A_2 = 1/\sqrt{2}$.

Besides the momentum operator, we can define the energy operator

$$E_{\text{op}} = i\hbar\frac{\partial}{\partial t}. \tag{97}$$

Applying it to the sinusoidal wave (91) gives

$$E_{\text{op}}\Psi(x,t) = \hbar\omega\, \Psi(x,t), \tag{98}$$

and we can interpret the eigenvalue $\hbar\omega$ as energy, in accordance to de Broglie formulas (81). It is interesting to note that the Schrödinger equation (90) can be written with these operators as

$$E_{\text{op}}\Psi(x,t) = \frac{p_{\text{op}}^2}{2m}\Psi(x,t) + U(x)\Psi(x,t), \tag{99}$$

where the momentum operator is applied twice to get the second derivative. Equation (99) looks like the energy in (83) as an operator applied to $\Psi(x,t)$.

We now look at the sinusoidal wave (91) as a function of x. Both the real and imaginary parts of this wave oscillate sinusoidally with x. For the absolute value squared we find

$$|\Psi(x,t)|^2 = |A|^2, \tag{100}$$

which is constant. What will we get if we try to measure where the particle is. In analogy with photons and particle momentum, we will find the particle somewhere, which appears random. The probability to find it at location x is proportional $|\Psi(x,t)|^2$. For the sinusoidal wave this is constant (100). This means the particle can be anywhere on the x axis with equal probability. This probability cannot be normalized. Later we will study cases where the wave function is more localized, so that the *normalization integral*

$$I = \int |\Psi(x,t)|^2 dx \tag{101}$$

has a finite value. In such cases, the wave function can be normalized by dividing by $\sqrt{I}$. As a result, we have a new $\Psi(x,t)$ that satisfies the *normalization condition*

$$\int |\Psi(x,t)|^2 dx = 1. \tag{102}$$

When this is satisfied, we define the *probability density*

$$P(x,t) = |\Psi(x,t)|^2 \tag{103}$$

so that $P(x,t)dx$ gives the probability to find the particle within a small range dx around x.

One important special case of the Schrödinger equation is to study *energy eigenstates* (98). They are also known as *energy levels*, *levels* or *stationary states*. In an energy eigenstate, the energy has a fixed value $E = \hbar\omega$. The time dependence of the wave function has a simple form,

$$\Psi(x,t) = \psi(x)e^{-i\omega t}. \tag{104}$$

Substituting this into the Schrödinger equation allows to cancel the factors $e^{-i\omega t}$, and we are left with *time-independent Schrödinger equation*

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}(x) + U(x)\psi(x) = E\psi(x) \tag{105}$$

Since ψ is a function of a single variable, the partial derivative has been replaced by ordinary derivative. In the following we study this in a number of cases.

4.3 Deep potential well

(YF 40.2) We study a particle in a deep *potential well.* We take $U(x) = 0$ in the range $0 < x < L$. Outside this range we take the limit $U \to \infty$. In order to get finite results, $\psi \to 0$ in the outside region. Thus the effective boundary conditions are

$$\psi(0) = \psi(L) = 0. \tag{106}$$

Instead of a traveling wave e^{ikx}, we now expect superposition of waves traveling in the positive and negative x directions,

$$\begin{aligned} \psi(x) &= Ce^{ikx} + De^{-ikx} \\ &= A\cos(kx) + B\sin(kx) \end{aligned} \tag{107}$$

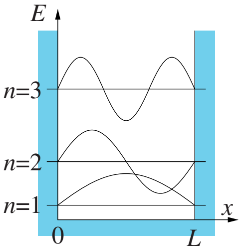
The latter form is obtained from the former using $e^{i\phi} = \cos\phi + i\sin\phi$. In both forms the coefficients A, B, C and D are complex in the general case. For the present problem the latter form is more convenient.

The boundary condition $\psi(0) = 0$ gives $A = 0$. The boundary condition $\psi(L) = 0$ gives $B\sin(kL) = 0$. Assuming $B \neq 0$ we should have $kL = \pi n$ with integer n. The nontrivial solutions have

$$k = \frac{\pi n}{L}, \quad n = 1, 2, 3, \ldots \tag{108}$$

The corresponding energies are obtained from (105), giving $E = \hbar^2 k^2/2m$, and thus

$$E = \frac{\pi^2\hbar^2 n^2}{2mL^2}, \quad n = 1, 2, 3, \ldots \tag{109}$$



The figure illustrates the three lowest energies and their wave functions. Note that wave numbers (108) are the same as for stretched string (35). In contrast to the string, the energies/frequencies (109) are not equidistant. This is caused by the dispersive character of the Schrödinger equation. Another noteworthy point is *zero point energy.* That is, the energy of the lowest energy state $n = 1$ is not zero.

In order to normalize the solutions we calculate I (101). We find $I = B^2 L/2$. Thus the normalized eigenstates are

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\frac{\pi n x}{L}, \quad n = 1, 2, 3, \ldots \tag{110}$$

The probability density is given by $|\psi(x)|^2$. (YF fig 40.12) Compare to the classical probability density, which is constant $1/L$.

4.4 Potential step etc.

(YF 40.3-4) We consider the case that $U(x) = 0$ for $x < 0$ and $U(x) = U_0 > 0$ for $x > 0$ and a particle arriving from the left. (figure) Consider first what happens according to classical mechanics. If the particle kinetic energy is less than U_0, the particle is reflected back. If the kinetic energy is larger than U_0, the particle is transmitted but with reduced speed.

In a quantum mechanical solution, we have to find the general wave functions in the regions $x < 0$ and $x > 0$. Then these waves have to be matched using boundary conditions. The boundary conditions are 1) the wave function has to be continuous and 2) the derivative of the wave function has to be continuous. These are understandable because the Schrödinger equation (105) is second order differential equation. If there were discontinuity in the wave function or its first derivative, the second derivative would be infinite at such a point. Considering potentials that are finite, this is not allowed.

Consider $E < U_0$. On the left ($x < 0$) the solution is of the standing wave form (107). On the right the energy of the solution is less than the potential energy, $E - U_0 < 0$,

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}(x) - (U_0 - E)\psi(x) = 0 \tag{111}$$

We try to solve this with ansatz $\psi = Fe^{\alpha x}$. We get $\alpha^2 = 2m(U_0 - E)/\hbar^2$, and $\alpha = \pm\sqrt{2m(U_0 - E)}/\hbar$. Only the negative sign leads to finite solution on the right. Thus

$$\psi(x) = F \exp\left(-\frac{\sqrt{2m(U_0 - E)}x}{\hbar}\right) \tag{112}$$

We see that this wave decays exponentially towards the right.

We will not go into applying the boundary conditions quantitatively, since it leads to complicated equations. Instead we sketch the results. (figure of step potential) We see that the particle can penetrate some distance into the potential. Since the wave functions vanishes exponentially, the particle must finally be reflected back. Thus the solution on the left looks similar as for the deep potential well, but shifted towards the positive x.

At energies $E > U_0$, there is both transmission and reflection.

We can apply these results to potential well of finite potential U_0 on both sides. The energy eigenvalues with $E < U_0$ are *discrete* and correspond to *bound states*, states confined to the potential well. Because these solutions penetrate under the potential, the corresponding k's and E's are lower than for deep potential well (at the same L). (YF fig 40.15) At energies the $E > U_0$ there are states at all energies. This is called *continuum*, in contrast to the discrete energies of the bound states. The continuum states are not bound to the well, but are partly reflected and partly transmitted while traveling through the well.

Let us consider a *potential barrier*. For simplicity, we assume a rectangular barrier, were $U(x) = 0$ everywhere except the range $0 < x < L$, where $U(x) = U_0 > 0$. (YF fig 40.20) Consider a particle at energy $E < U_0$. The situation resembles that for a potential step: incoming and reflected wave on the left and exponential decay (112) of the wave function in the barrier. Because the barrier has finite length L, the wave function is small but nonzero at $x = L$. In the region $x > L$, the wave can continue as a traveling wave. Thus, in quantum mechanics, a particle can *tunnel* through a potential barrier. Since part of the wave is transmitted, the probability for reflection is less than 1.

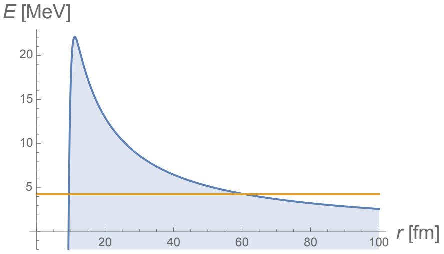
Let us estimate the tunneling probability. The amplitude of the transmitted wave relative to the incoming one is given by the ratio of (112) at $x = L$ and $x = 0$. Since the probability is proportional to the amplitude squared, we get the estimate of tunneling probability

$$T \approx \exp\left(-\frac{2\sqrt{2m(U_0 - E)}L}{\hbar}\right) \tag{113}$$

We see that this depends exponentially on L and on the energy relative to the top of the barrier, $U_0 - E$.

In classical mechanics, there is no tunneling, $T = 0$ at $E < U_0$ and complete transmission $T = 1$ for $E > U_0$. In fact, we can see that quantum mechanics corrects some problems that are present in the classical case. As $T(E)$ is discontinuous at $E = U_0$, one can ask what happens at this energy. Another problem is the discontinuity in $T(L)$: for $L = 0$ we expect $T = 1$ but for any nonzero L we expect $T = 0$. These problems disappear in quantum mechanics since in both limits, $E \to U_0$ and $L \to 0$, the transmission (113) grows continuously towards 1.

An important, and early, application of tunneling is to nuclear reactions, in particular to *alpha decay*. For example, ^{238}U can emit an alpha particle (^{4}He nucleus) to decay to ^{224}Th. Consider the alpha particle at different distances from the rest of the original nucleus. As both have positive charge, there is Coulomb repulsion between the two. Another force is the nuclear force, that is attractive and strong, but only of short range. Thus there is a potential barrier between the alpha particle being in the nucleus and being far away (on the nuclear scale).



The figure gives the potential $U(r)$ (blue line) and the energy E of an α particle (horizontal line) in ^{238}U, 1 fm$= 10^{-15}$ m.

If the energy of the alpha particle in the nucleus is lower than the potential far away, the original nucleus is stable. In the opposite case alpha emission is possible, but is delayed by the potential barrier. The fact that tunneling probability is exponential in energy, allows to understand the enormous variety in the half lives of alfa-decaying nuclei, from microseconds to 10^{15} years. ^{238}U half life is 4.468×10^9 years.

Examples where tunneling is essential: nuclear fusion, tunnel diode, Josephson junction, scanning tunneling microscope, biomolecules

Resonant tunneling. Consider two barriers in a row. We can consider the region between the barriers as finite potential well. Suppose it has a bound state. A solution of the Schrödinger equation gives that the transmission trough the double barrier depends strongly on the energy. At particle energy equal to the bound state energy, complete transmission can take place, (figure)

4.5 Harmonic oscillator

(YF 40.5) Harmonic potential $U(x) = \frac{1}{2}kx^2$. The classical oscillation angular frequency is $\omega_0 = \sqrt{k/m}$. Strangely enough, the dispersion of the Schrödinger equation and the form of the potential cancel each effects so that the energy levels are equally spaced,

$$E_n = \hbar\omega_0 \left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \dots \tag{114}$$

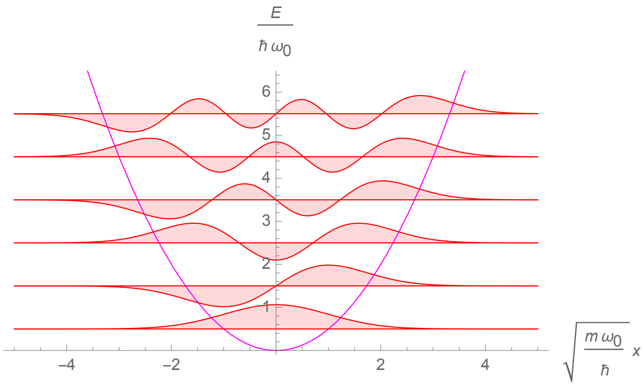


Figure: the parabolic potential energy, the 6 lowest energy eigenvalues (horizontal lines) and the corresponding wave functions.

The ground state solution

$$\psi(x) = \left(\frac{m\omega_0}{\pi\hbar}\right)^{\frac{1}{4}} \exp\left(-\frac{m\omega_0}{2\hbar}x^2\right). \tag{115}$$

corresponds to the minimal uncertainty in (67).

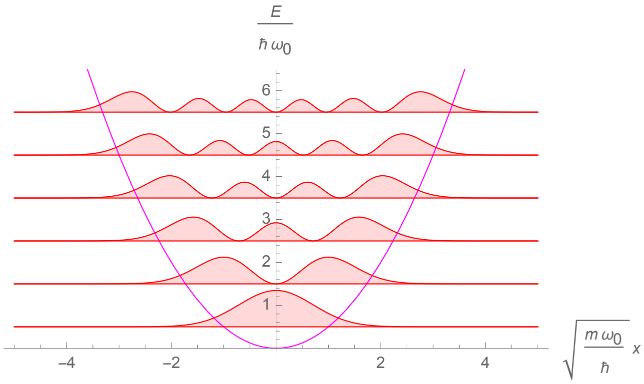


Figure: The probability densities $|\psi(x)|^2$ for the same states in the previous figure.

The harmonic oscillator forms a good approximation in several cases, although the potential is not strictly harmonic. Atomic vibrations in molecules and solids form an example.

4.6 Measurement

(YF 40.6) Let us return to the discussion of measurement. As an example, we considered the superposition of two sinusoidal waves (95). These could be the two counterpropagating waves in a bound state of a potential well. Detailed discussion of the measurement should take into account how the momentum is measured. For example, one can use a photon to scatter from the electron. Depending on what is the direction of the scattered photon, one could deduce the momentum of the electron. In our example, we measure one or the other value of the momentum (with probabilities proportional to the amplitudes squared, as discussed above). This means that we get new information in the measurement. The most common way to interpret this is called *Copenhagen interpretation*. This says that the new information should be used in constructing a new wave function, where the electron has precisely the momentum that has been measured. That is, out of the initial superposition, the part not having the measured value is simply wiped out. In the case of (95), if we measure $p = \hbar k_1$, the wave function after the measurement is

$$\Psi = Ae^{i(k_1 x - \omega_1 t)} \tag{116}$$

This is called *collapse of the wave function*. The collapse itself is not in the Schrödinger equation, but after the measurement the new wave function develops according to the Schrödinger equation.

Several physicists consider the collapse of the wave function problematic. In particular, when should the collapse take place, what is the time point when the measurement is done? The answer seems to be that it really does not matter. We may say that the collapse takes place when output of the measuring device shows one or the other value. But if we say the collapse takes one second after that, there seems to be no difference. This leads one to think, does the collapse ever take place? The answer ”never” is known as *many-world interpretation*. It says that in the measurement of the momentum, the world branches into two separate worlds so that one value is measured in one world and the other in the other world. We see only one measurement outcome since we live in one world only, but there would be the other world as well. As measurements take place continuously, one world would branches rapidly into a huge number of parallel worlds, which is difficult to imagine.

In addition to the collapse, there is additional effect of the measurement: the photon scattered from the electron changes the electron momentum. When speaking about measurements of microscopic objects, the recoil of the electron can be essential. Therefore, the electron wave function after the measurement has to constructed taken this recoil into account.

4.7 Schrödinger equation in 3 dimension

(YF 41.1) We have written the Schrödinger equation (90) for one spatial dimension. The generalization to 3 dimensions is similar as for the wave equation (40). The wave function depends on the location vector $\boldsymbol{r}$ and on time, $\Psi(\boldsymbol{r},t)$. The Schrödinger equation takes the form

$$i\hbar\frac{\partial\Psi}{\partial t}(\boldsymbol{r},t) = -\frac{\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r},t) + U(\boldsymbol{r})\Psi(\boldsymbol{r},t), \tag{117}$$

where ∇^2 is given by (42). Written explicitly using cartesian coordinates $\boldsymbol{r} = x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}} + z\hat{\boldsymbol{z}}$, we have $\Psi(x,y,z,t)$ and the equation

$$\begin{aligned} i\hbar\frac{\partial\Psi}{\partial t} &= -\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2}\right) \\ &+ U(x,y,z)\Psi. \end{aligned} \tag{118}$$

The normalization condition (102) in 3D is

$$\int |\Psi(\boldsymbol{r},t)|^2 dV = 1, \tag{119}$$

The sinusoidal wave (91) has a straightforward generalization to 3 D

$$\Psi(\boldsymbol{r},t) = Ae^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \tag{120}$$

where $\boldsymbol{k}$ is the wave vector (37). Noting that $\boldsymbol{k}\cdot\boldsymbol{r} = k_x x + k_y y + k_z z$, the plane wave can be written in product form

$$\Psi(\boldsymbol{r},t) = Ae^{ik_x x}e^{ik_y y}e^{ik_z z}e^{-i\omega t}. \tag{121}$$

[Exercise: verify that (120) is a solution of (117) for $U \equiv 0$ when $\omega = \hbar k^2/2m$.] The absolute value squared $|\Psi|^2 = |A|^2$ is a constant. That is, the probability to find the particle is the same at any location $\boldsymbol{r}$.

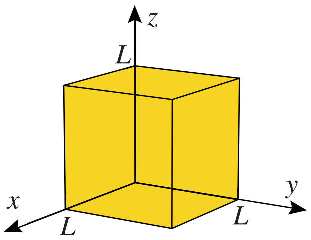
Similarly as in 1D, we can look at an energy eigenstate $\Psi(\boldsymbol{r},t) = e^{-i\omega t}\psi(\boldsymbol{r})$, and write the time-independent Schrödinger equation. In cartesian coordinates it is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) + U(x,y,z)\psi = E\psi, \tag{122}$$

where $\psi(x,y,z)$.

4.8 Deep potential well in 3 dimension

(YF 41.2) Let us now consider the generalization of the deep potential well to 3D. We assume $U \to \infty$ everywhere except a cube of side L: $0 < x < L$, $0 < y < L$ and $0 < z < L$, where $U = 0$.



The boundary condition for ψ is that it has to vanish at the sides of the cube and outside.

The general method to find solutions of partial differential equations, like (122) is called *separation of variables*. We assume a solution that is a product of functions each depending on a single variable,

$$\psi(x,y,z) = X(x)Y(y)Z(z). \tag{123}$$

We substitute this in (122). Dividing the equation by ψ we get

$$\begin{aligned} &\left(-\frac{\hbar^2}{2m}\frac{1}{X(x)}\frac{d^2X(x)}{dx^2}\right) + \left(-\frac{\hbar^2}{2m}\frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2}\right) \\ &\qquad + \left(-\frac{\hbar^2}{2m}\frac{1}{Z(z)}\frac{d^2Z(z)}{dz^2}\right) = E \end{aligned} \tag{124}$$

We see that the first term depends only on x, the second on y and the third on z, and the right hand side is a constant. Looking from to point of x, the second and third terms are constants as well. Thus the first term must be equalt to a constant. We denote it by E_x. Similarly for y and z. Thus the equation separates into three ordinary differential equations

$$\begin{aligned} -\frac{\hbar^2}{2m}\frac{d^2X(x)}{dx^2} &= E_X X(x) \\ -\frac{\hbar^2}{2m}\frac{d^2Y(y)}{dy^2} &= E_Y Y(y) \\ -\frac{\hbar^2}{2m}\frac{d^2Z(z)}{dz^2} &= E_Z Z(z) \end{aligned} \tag{125}$$

and $E_X + E_Y + E_Z = E$. We look at the first one. The boundary condition for $X(x)$ is $X(0) = 0$ and $X(L) = 0$. Thus the problem of solving $X(x)$ is the same as we had for $\psi(x)$ in the case of deep one-dimensional potential well in (106). We can read the wave function from (110) and energy from (109). Renaming some symbols they are $X(x) = \sqrt{2/L}\sin(\pi n_1 x/L)$ and $E_X = \pi^2\hbar^2 n_1^2/2mL^2$. Similarly we can find the solutions for $Y(y)$ and $Z(z)$. Putting the results together gives the wave function

$$\psi(x,y,z) = \left(\frac{2}{L}\right)^{3/2}\sin\frac{\pi n_1 x}{L}\sin\frac{\pi n_2 y}{L}\sin\frac{\pi n_3 z}{L}, \tag{126}$$

and the energy

$$E = \frac{\pi^2\hbar^2}{2mL^2}(n_1^2 + n_2^2 + n_3^2). \tag{127}$$

The numbers n_1, n_2 and n_3 are called *quantum numbers*. They are positive integers $1, 2, 3, \ldots$.

The lowest energy level has quantum numbers $(n_1, n_2, n_3) = (1,1,1)$ and energy $E = 3E_0$, where $E_0 = \pi^2\hbar^2/2mL^2$. The next lowest levels has $(2,1,1)$ and energy $E = 6E_0$. The same energy is obtained with quantum numbers $(1,2,1)$ and $(1,1,2)$ as well. When there are two or more states that have the same energy, the energy level is called *degenerate*. This is typically caused by some symmetry in the system. In the present case we have chosen the box of same length L in x, y and z directions. In a rectangular box with nonequal lengths in x, y and z directions, the three levels would not be degenerate. (Exercise: analyze all energy levels up to $E = 12E_0$.)

4.9 Hydrogen atom

(YF 41.3) Above we discussed hydrogen atom using the Bohr model. Now we are ready to study it using the Schrödinger equation (117). In looking for the stationary states, we use the time-independent Schrödinger equation

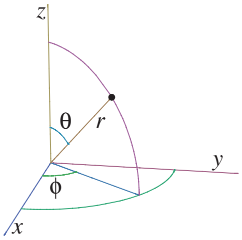
$$-\frac{\hbar^2}{2m}\nabla^2\psi(\boldsymbol{r}) + U(\boldsymbol{r})\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \tag{128}$$

The Coulomb potential between the electron and the nucleus is

$$U(\boldsymbol{r}) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}. \tag{129}$$

This depends only on the radial distance $r = |\boldsymbol{r}|$. Because of spherical symmetry, the solutions of the Schrödinger equation (128) are most conveniently obtained using spherical coordinates. Spherical coordinates (r, θ, ϕ) are defined by

$$\boldsymbol{r} = r(\hat{\boldsymbol{x}}\cos\phi + \hat{\boldsymbol{y}}\sin\phi)\sin\theta + \hat{\boldsymbol{z}}r\cos\theta. \tag{130}$$



All $\boldsymbol{r}$ can be described by the coordinates in the ranges $0 \le r$, $0 \le \theta \le \pi$ and $0 \le \phi < 2\pi$.

The Schrödinger equation (128) contains the Laplace operator ∇^2 (42). In spherical coordinates it is given by

$$\begin{aligned}\nabla^2\psi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)\\ &\quad + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}.\end{aligned} \tag{131}$$

This formula is justified in the appendix.

Let us first demonstrate one solution of the problem. Substitute the ansatz $\psi(r) = e^{\alpha r}$ into (128). Based on that determine the constant α and the energy eigenvalue E. Check that the E you get is equal to the ground state energy (the lowest energy) of the Bohr model, and α is simply related to the Bohr radius. (Advance assignment)

In order to find other solutions, we again use separation of variables. We write

$$\psi(r,\theta,\phi) = R(r)\Theta(\theta)\Phi(\phi). \tag{132}$$

We substitute this in (128). Dividing the equation by ψ we get

$$\begin{aligned}&\frac{1}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)\right]\\ &+\frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]\\ &+\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r} = E.\end{aligned} \tag{133}$$

We would like to argue, as in the case of (124), that each term is a constant. Now it does not work out immediately but can be achieved with some manipulations. We rearrange the terms and multiply by r^2

$$\begin{aligned}&\frac{r^2}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R - ER\right]\\ &+\frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]\\ &+\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] = 0.\end{aligned} \tag{134}$$

We define a new quantity L^2 so that $L^2/2m$ is equal to the two last terms in (134),

$$\begin{aligned}&\frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]\\ &+\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] = \frac{L^2}{2m}.\end{aligned} \tag{135}$$

We see that L^2 cannot depend on r since there is no r dependence in (135). Using (135) we can write (134) as

$$\frac{r^2}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R - ER\right] = -\frac{L^2}{2m}. \tag{136}$$

We see from this that L^2 cannot depend on θ and ϕ since there is neither θ nor ϕ in this equation. Thus we conclude L^2 is a constant (independent of the spherical coordinates).

We manipulate (135) rearranging the terms and multiplying by $\sin^2\theta$. We get

$$\begin{aligned}&\frac{\sin^2\theta}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) - \frac{L^2}{2m}\Theta\right]\\ &+\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{d^2\Phi}{d\phi^2}\right] = 0.\end{aligned} \tag{137}$$

We notice that the first term depends only on θ and the second only on ϕ. Thus they both have to be constants. We define L_z^2 so that the latter term is equal to $L_z^2/2m$,

$$\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{d^2\Phi}{d\phi^2}\right] = \frac{L_z^2}{2m}. \tag{138}$$

Based on these, we can now write independent equations for all the three functions

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{L^2}{2mr^2}R-\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R=ER$$
$$-\frac{\hbar^2}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{L_z^2}{\sin^2\theta}\Theta=L^2\Theta$$
$$-\hbar^2\frac{d^2\Phi}{d\phi^2}=L_z^2\Phi \quad (139)$$

In solving (139) it is best to start from the simplest, the one for $\Phi(\phi)$. Recall that ϕ is the angle around the z axis. In order to have unique solution $\Phi(\phi)$, it has to remain the same when ϕ changes by 2π. That is, we require *periodic boundary condition* $\Phi(\phi+2\pi)=\Phi(\phi)$. We look the solution of the last (139) as an exponential $e^{ik\phi}$. The boundary condition limits $2\pi k$ to be an integral multiple of 2π. The quantum number associated with ϕ is called the *orbital magnetic quantum number* and marked by m_l. Thus

$$\Phi(\phi)=\frac{1}{\sqrt{2\pi}}e^{im_l\phi},\quad m_l=0,\pm1,\pm2,\ldots, \quad (140)$$

where the prefactor comes from normalization. Substituting this in the last one in (139) shows that it is satisfied when

$$L_z=\hbar m_l \quad (141)$$

This resembles what we had in Bohr model (77). We can identify L_z as the z component of the angular momentum.

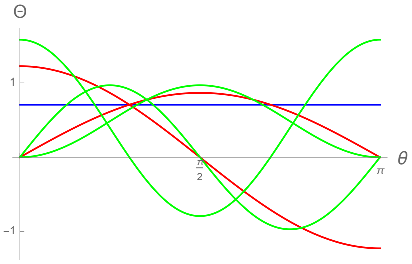
The next step is to solve the second equation in (139). The boundary condition for $\Theta(\theta)$ is that it should be finite in the whole interval $0\le\theta\le\pi$. We will give the general result without proof. The quantity L is the angular momentum. Its values are given by the *orbital angular momentum quantum number* l so that

$$L=\hbar\sqrt{l(l+1)},\quad l=0,1,2,\ldots. \quad (142)$$

In addition $|m_l|\le l$. The wave functions $\Theta(\theta)$ are polynomials of $\cos\theta$ and $\sin\theta$. We give the normalized functions corresponding to the lowest quantum numbers

$$\Theta=\frac{1}{\sqrt{2}},\quad l=0,m_l=0$$
$$\Theta=\sqrt{\frac{3}{2}}\cos\theta,\quad l=1,m_l=0$$
$$\Theta=\sqrt{\frac{3}{4}}\sin\theta,\quad l=1,m_l=\pm1$$
$$\Theta=\sqrt{\frac{5}{8}}(3\cos^2\theta-1),\quad l=2,m_l=0$$
$$\Theta=\sqrt{\frac{15}{4}}\cos\theta\sin\theta,\quad l=2,m_l=\pm1$$
$$\Theta=\sqrt{\frac{15}{16}}\sin^2\theta,\quad l=2,m_l=\pm2. \quad (143)$$

Your are encouraged to check that these indeed satisfy the second of (139) with the eigenvalue (142).



The figure gives the functions $\Theta(\theta)$ for $l=0$ (blue), $l=1$ (red) and $l=2$ (green).

Finally we are in the position to solve the radial equation, the first one in (139). The boundary condition is that the solution should vanish with $r\to\infty$. Again, we give the result without proof. The energy eigenvalues are

$$E=-\frac{me^4}{2(4\pi\epsilon_0)^2\hbar^2n^2}=-\frac{13.60\text{ eV}}{n^2},\quad n=1,2,3,\ldots \quad (144)$$

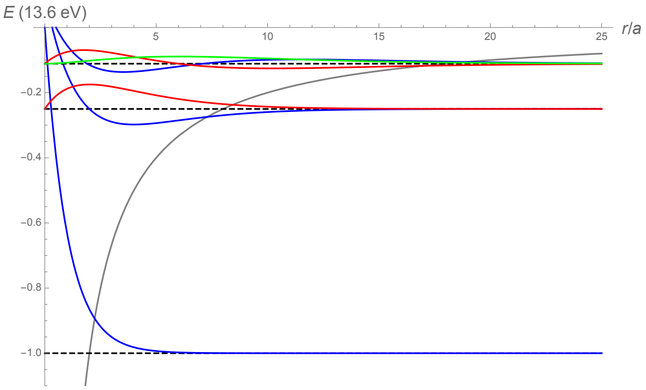
These values are the same as in the Bohr model [(75) and (77)]. The integer n is called the *principal quantum number*. For a given value of n, the orbital angular quantum number l can take values from 0 to $n-1$,

$$l=0,1,\ldots,n-1. \quad (145)$$

Thus for $n=1$ we have only $l=0$. For $n=2$ the values $l=0$ and $l=1$ are allowed. The radial functions are polynomials multiplying exponential function. Here we give lowest order ones

$$R=\frac{2}{a^{3/2}}e^{-r/a},\quad n=1,l=0$$
$$R=\frac{1}{\sqrt{2}a^{3/2}}\left(1-\frac{r}{2a}\right)e^{-r/2a},\quad n=2,l=0$$
$$R=\frac{1}{2\sqrt{6}a^{3/2}}\frac{r}{a}e^{-r/2a},\quad n=2,l=1. \quad (146)$$

where a is the Bohr radius (78).



The figure gives the hydrogen potential, the three lowest energy levels and the radial wave functions. The color coding corresponds to the figure of Θ functions above.

We summarize the quantum numbers

- n, the principal quantum number $n = 1, 2, 3, \ldots$.
- l, the orbital angular momentum quantum number, $l = 0, 1, \ldots, n-1$.
- m_l, the orbital magnetic quantum number, $m_l = 0, \pm 1, \ldots, \pm l$.

We have now solved the bound state energies and wave functions for the hydrogen atom. We now discuss some of the implications.

4.10 Angular momentum

We found above the angular momentum and its z component are described by quantum numbers l and m_l,

$$\begin{aligned} L &= \hbar\sqrt{l(l+1)}, & l &= 0, 1, 2, \ldots \\ L_z &= \hbar m_l, & m_l &= 0, \pm 1, \ldots, \pm l. \end{aligned} \qquad (147)$$

For example, if $l = 1$, we have $L = \sqrt{2}\hbar$ and L_z can have values 0 and $\pm\hbar$. We see that that L_z is always less than L, except the case $l = 0$, where $L = L_z = 0$. We can understand the impossibility of $L_z = L$ as follows. For that to hold, the particle has to circulate in the x-y plane, i.e. at $z = 0$. But the uncertainty relation then says that the momentum p_z cannot be zero. Thus the particle necessarily has to move out of the x-y plane, and the total angular momentum is thus larger than the z component.

The possible values of L and L_z can be illustrated by drawing the equidistant $L_z = \hbar m_l$ values on the L_z axis and then figure out a vectors $\boldsymbol{L} = L_x\hat{\boldsymbol{x}} + L_y\hat{\boldsymbol{y}} + L_z\hat{\boldsymbol{z}}$ of length $L = |\boldsymbol{L}|$. For each m_l we get a definite value $\sqrt{L_x^2 + L_y^2} = \sqrt{L^2 - L_z^2}$ of angular momentum perpendicular to z axis. (YF fig 41.6)

We now ask can we also know L_x and L_y. The answer is no. Consider some state with L and L_z determined by quantum numbers l and m_l. If one now measures L_x, this will collapse the original state to a state with the same l but to an eigenstate of L_x. These are similar as in (140), but with the angle ϕ denoting the rotation angle around the x axis. In this state, the information on L_z in the original state has disappeared. Thus, only the magnitude of the angular momentum and its component along one axis can be precisely known in quantum mechanics.

In the figure we discussed, we can only specify the L, L_z and $\sqrt{L_x^2 + L_y^2} = \sqrt{L^2 - L_z^2}$, but not L_x and L_y. However, this problem is alleviated when l is large. In this case the states with different m_l are dense, in comparison to l. Then forming a linear combination of the states of different m_l, one can from a state with rather accurately determined vector $\boldsymbol{L}$. This then becomes similar as in classical mechanics, where $\boldsymbol{L}$ can be known accurately. Thus for large quantum numbers, the prediction of quantum mechanics become closer to the ones obtained in classical mechanics. This is known as *correspondence principle*.

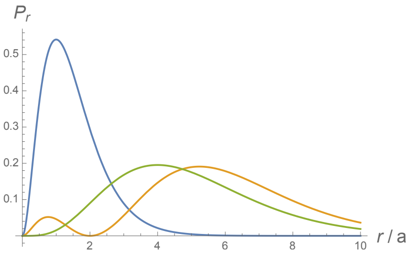
We make comparison to the Bohr model. He assumed $L = \hbar n$ (77). In quantum mechanics this is replaced by formulas (147) and (145), which are not the same. One clear difference is that the ground state $n = 1$ has $L = 1$ in Bohr model but $L = 0$ in quantum mechanics. Thus we conclude that the Bohr model is a nice introductory step to quantum mechanics, but contains incorrect physics.

We note that this discussion of angular momentum uses only the spherical symmetry, and is valid for any potential function $U(r)$, not just (129).

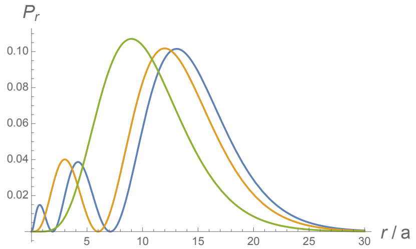
The states having $l = 0, 1, 2$ and 3 are commonly called s, p, d and f states.

4.11 Hydrogen probability distributions

The probability density of electron is given by $P = |\psi|^2$ (103). In 3 D these are not trivial to show because dependence on more than one variable. The θ dependence can be obtained by squaring (143). The higher l, the more there are zeros. In studying the radial dependence, note that $P = |\psi|^2$ gives the probability per volume dV. One can also consider probabilities P_r per radial distance dr. This is different since the volume $dV = 4\pi r^2 dr$, where $4\pi r^2$ is the surface area of a sphere of radius r. The radial functions $R(r)$ are normalized so that $P_r = r^2 R^2$.



The figure shows radial probability density $P_r(r)$ for states with $n = 1$ (blue) and $n = 2$.



The figure shows radial probability density $P_r(r)$ for states with $n = 3$. (Note different radial scales in the two figures.) The rule is that the larger n, the larger the average distance of the electron from the nucleus. Note that radii of the orbits in Bohr model for states $n = 1, 2$ and 3 would be the Bohr radius a (78), $4a$ and $9a$. Compare these to the distributions in the figures. In the ground state, the probability to find the electron within radius a is 0.32. You can identify values of l in the figures knowing that the larger l, the less there are zeros in the radial direction. This is compensated by there being more zeros in the angular dependence given by Θ functions. That is, for a given energy (n), the electron momenta (and the zeros) are mainly directed in the angular direction for large l, and in the radial direction for small l.

Above we have selected the Φ states (140) as eigenstates of the operator L_z. It is defined by

$$(L_z)_{\rm op} = \frac{\hbar}{i}\frac{\partial}{\partial\phi} \tag{148}$$

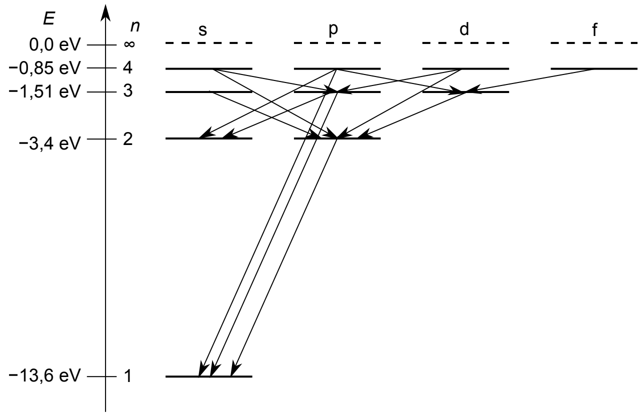
in close resemblance to the momentum operator (92). The L_z eigenstates (140) have constant $|\Phi|^2 = 1/2\pi$. For some purposes, in particular for application in chemistry, it is more convenient to choose linear combinations of them. For example, in the case of $l = 1$ we define

$$\begin{aligned}\Phi_{p_x}(x) &= \frac{1}{\sqrt{2}}\left(\frac{e^{i\phi}}{\sqrt{2\pi}} + \frac{e^{-i\phi}}{\sqrt{2\pi}}\right) = \frac{1}{\sqrt{\pi}}\cos\phi,\\ \Phi_{p_y}(x) &= \frac{1}{\sqrt{2}i}\left(\frac{e^{i\phi}}{\sqrt{2\pi}} - \frac{e^{-i\phi}}{\sqrt{2\pi}}\right) = \frac{1}{\sqrt{\pi}}\sin\phi.\end{aligned} \tag{149}$$

These are largest in the x and y directions, respectively, and are called p_x and p_y orbitals. Expect of different orientation, they are identical to the p_z orbital, which corresponds to $m_l = 0$, see (143).

4.12 Hydrogen energy levels

(YF 41.4) We see that the energy (144) depends on n only. This means that all states with different values of l and m_l are degenerate. The energy levels can be visualized by a figure, with E on the vertical axis and the energies (144) are drawn as short horizontal lines at locations $l = 0, 1, 2, \ldots$.



(Figure: wikipedia Grotrian diagram)

Electromagnetic radiation causes transitions between the levels, or an excited state can spontaneously decay by emission. In these processes the photon energy corresponds to the energy difference of the levels (76). Not all transitions are equally likely. Based on the wave functions, one can calculate the rates of different transitions. We leave such nice calculations for your next course of quantum mechanics. Here we only state that the strongest transitions correspond to change of l by unity, $\Delta l = \pm 1$. Such a rule is called *selection rule*. These transitions are drawn in the figure above.

We mentioned above that degeneracy of energy levels is commonly associated with a symmetry. The degeneracy with respect to m_l comes from the spherical symmetry. That can be broken by placing the atom in an external magnetic field, as we will discuss next.

We start by defining the *magnetic moment* $\boldsymbol{\mu}$ on the basis of classical electromagnetism. It is defined as the electric current I flowing in a planar circuit times the area of the circuit A: $\mu = IA$. The direction of the magnetic moment is perpendicular to the plane of the circuit, using the right hand rule. In external magnetic field $\boldsymbol{B}$, the magnetic moment is associated with energy

$$E_m = -\boldsymbol{B}\cdot\boldsymbol{\mu}. \tag{150}$$

Consider now the motion of an electron in an atom according to classical physics. We also assume circular orbit of radius r and velocity v. We can think the electron motion to form an electric current $I = -ef$, where the frequency $f = v/2\pi r$. For the magnetic moment we get

$$\mu = -\frac{1}{2}evr = -\frac{e}{2m}L, \tag{151}$$

where $L = mrv$ is the angular momentum. We write the result in vector form

$$\boldsymbol{\mu} = -\frac{e}{2m}\boldsymbol{L}. \tag{152}$$

This is simple result since $\boldsymbol{\mu}$ and $\boldsymbol{L}$ are related by natural constants, the elementary charge and electron mass. We claim that the result (152) is generally valid, independently of the simple assumptions about the electron orbit.

We return to quantum mechanics. We assume the magnetic field in the z direction, $\boldsymbol{B} = B\hat{\boldsymbol{z}}$. Using (152) and (147), we write the magnetic energy (150)

$$E_m = m_l\mu_{\rm B}B, \tag{153}$$

where we have defined *Bohr magneton* $\mu_{\rm B} = e\hbar/2m = 9.274\times10^{-24}$ J/T $= 5.8\times10^{-5}$ eV/T. The magnetic energy (153) should be added to the hydrogen energy levels energies (144). Its magnitude is much smaller, but still measurable. Since there are $2l+1$ different values of m_l in a level with l, the the energies of the l levels split into $2l+1$ equally spaced levels. That is, there are 1, 3 and 5 sublevels for s, p and d states. (YF fig 41.13). This is know as *Zeeman effect*. Again, not all transitions are equally likely. In addition to the $\Delta l = \pm 1$ rule, there is $\Delta m_l = 0$ or ± 1 selection rule. This affects how the levels appear in a spectroscopy experiment.

Besides spectroscopy experiments, the level splitting was observed in experiment made by Otto Stern and Walter Gerlach in 1922. They had a beam of atoms is moving through a highly inhomogeneous magnetic field. The gradient of the field makes the beam deflect by amount that is proportional to the magnetic energy E_m (153). As a result the, the slitting of the (n, l) levels into sublevels could be observed.

4.13 Electron spin

(YF 41.5) It was found experimentally that there were splitting of the energy levels that could not be understood by the theory presented above. In order to understand these, Samuel Goudsmidt and George Uhlenbeck suggested in 1925 that there is additional angular momentum that is intrinsic to the electron. This is known as *spin* of the electron. The simple minded picture is that the electron is spinning around itself. However, no internal structure of the electron, other the spin itself, has been observed. Therefore we have to say that the electron behaves like it would be rotating around itself.

The electron spin differs from the orbital angular momentum in that the *spin quantum number* $s = \frac{1}{2}$. Otherwise it is like the orbital angular momentum in (147):

$$\begin{aligned} S &= \hbar\sqrt{s(s+1)} = \hbar\sqrt{\tfrac{3}{4}} \\ S_z &= \hbar m_s, \qquad m_s = \pm\tfrac{1}{2}, \end{aligned} \tag{154}$$

where S is the magnitude of the spin angular momentum and S_z the component in the z direction. m_s is the *spin magnetic quantum number*. Its two values are often called ”spin up” and ”spin down”. Associated with the spin, the electron also has magnetic moment.

$$\mu_z = -g\frac{e}{2m}S_z, \tag{155}$$

where the *g factor* $g = 2.002319304364 \approx 2$. This is one of the best known constants in nature, which can be both measured and explained theoretically using quantum electro dynamic theory, which is outside the scope of these lectures.

The electron spin causes additional splitting of the hydrogen levels in external magnetic field because of magnetic energy associated with spin, $E_{m,s} = m_s g \mu_{\mathrm{B}} B$. But there is an effect without external magnetic field as well. This is called *spin-orbit coupling*. The electrons orbital motion generates a magnetic field whose direction can be characterized by $\boldsymbol{L}$. This field couples to the spin magnetic moment (155). The energy is of the form

$$E_{SO} = a\boldsymbol{L}\cdot\boldsymbol{S}. \tag{156}$$

with some constant a.

Now the total angular momentum $\boldsymbol{J}$ of the electron is the sum of orbital and spin parts,

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S} \tag{157}$$

We claim that also $\boldsymbol{J}$ is described by a quantum number, *the total angular momentum quantum number*, so that

$$J = \hbar\sqrt{j(j+1)}, \quad j = |l \pm \tfrac{1}{2}|. \tag{158}$$

The two values of j (in the case of $l \neq 0$) correspond to different relative orientations of $\boldsymbol{L}$ and $\boldsymbol{S}$. The spin-orbit energy (156) is thus different for the two states. As a consequence, the degeneracy of the levels of different l but same n in (144) is lifted.

We stop here by noting some effects that still were neglected. The mass in (144) is the reduced mass,

$$\frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2} \quad \Leftrightarrow \quad m = \frac{m_1 m_2}{m_1 + m_2}, \tag{159}$$

where m_1 and m_2 are the electron and proton masses. This is because both the electron and the proton are rotating around their center of mass. There is relativistic correction to electron kinetic energy. This together with the spin-orbit correction is called *fine structure*. There are still smaller corrections called *hyperfine structure*, caused by the structure of the nucleus and so-called radiative corrections,

4.14 Hydrogen-like atoms

Helium nucleus has two protons, the atomic number $Z = 2$, and lithium there protons, $Z = 3$. In uncharged states these have the same number of electrons. It is possible to ionize these by removing electrons. For He^+ and Li^{2+} and so on, only one electron is left. The calculations above for hydrogen can be generalized to these taking into account that the charge of the nucleus is now Ze instead of e. Thus the potential (129) becomes multiplied by Z, the energy (144) multiplied by Z^2 and the characteristic scale of the ground state, the Bohr radius (78) becomes multiplied by $1/Z$. We see that with increasing Z the binding energy grows rapidly and the electron is bound much closer to the nucleus.

5. Many-particle systems

5.1 Many-body prolem

(YF 41.6) Above we have considered the motion of a single particle in a given potential. In this case, accurate calculations can be done. The situation changes radically when we start considering, for example, the helium atom with two electrons, not to speak about heavier atoms, molecules or the solid state. Let us try to figure out how could we generalize the Schrödinger equation (117) to the presence of two electrons, the He atom.

Let us denote the coordinates of the two electrons by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Obviously, the wave function should depend on both of these,

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t). \tag{160}$$

Note that using coordinates (cartesian or spherical), there are 6 spatial coordinates. The kinetic energy part should contain the sum of the separate kinetic energies,

$$-\frac{\hbar^2}{2m}\nabla_1^2\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t) - \frac{\hbar^2}{2m}\nabla_2^2\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t), \tag{161}$$

where the Laplace operators ∇_n^2 operate on the coordinates $\boldsymbol{r}_n$ of the n'th particle. The potential energy should contain the interactions of both electrons with the nucleus,

$$U_{\mathrm{central}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\frac{Ze^2}{4\pi\epsilon_0}\frac{1}{r_1} - \frac{Ze^2}{4\pi\epsilon_0}\frac{1}{r_2}, \tag{162}$$

but in addition there is the Coulomb repulsion between the two electrons

$$U_{\mathrm{el-el}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{e^2}{4\pi\epsilon_0}\frac{1}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|} \tag{163}$$

We could try to solve the two-body Schrödinger equation by separation of variables

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \psi_1(\boldsymbol{r}_1)\psi_2(\boldsymbol{r}_2) \tag{164}$$

This would work fine for other terms, but there is a problem in electron-electron interaction (163): where is electron 1, affects the potential of electron 2. Thus their motion is correlated, and cannot be represented in the form (164). This is the essential difficulty in *many-body problem.*

It is not possible to find analytic solution for the helium atom. Precise numerical methods can still be used, but they also run in rapidly growing difficulties when the electron number grows towards 5. For all heavier atoms, molecules, and solid state, approximations are absolutely necessary.

Very common approximation is *mean field approximation* or *independent-electron approximation.* The idea is that the particle, instead of being correlated with others, only feels the average effect of the other particles. That is, we can still use the one-particle Schrödinger equation (117), but with an effective potential $U = U_{\mathrm{eff}}$, that includes the averaged effect of all other particles.

5.2 Atomic struture

We will now be using the independent-electron approximation. That is, the electron moves in a static potential caused by the nucleus and all other electrons. In addition, we make *central field approximation*, where the effective potential is assumed to have spherical symmetry, $U_{\mathrm{eff}}(\boldsymbol{r}) = U_{\mathrm{eff}}(r)$.

The spherical symmetry means that the angular part of the wave functions, which we derived for hydrogen, is valid for all electrons in all atoms. Thus, the electrons of an atom are labelled with the angular momentum quantum numbers given in (147). What still remains to be done is solve the radial wave functions from the radial equation

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \frac{\hbar^2 l(l+1)}{2mr^2}R + U_{\mathrm{eff}}(r)R = ER \tag{165}$$

As long as $U_{\mathrm{eff}}(r)$ is not know, we cannot do precise calculations. However, we still expect that we can label the solutions with a quantum number $n = 1, 2, \ldots$, as in the case of hydrogen.

In the case of hydrogen, we found that the energies E did not depend on l in (165). This degeneracy can be traced to the $1/r$ form of potential (129). In many-electron system we do not expect the effective potential $U_{\mathrm{eff}}(r)$ to have $1/r$ form. Therefore, we can expect that different l values are no more degenerate. That is, the energies depend both on n and l.

Let us summarize the quantum numbers. We have the angular quantum numbers l and m_l (147). The radial wave functions from (165) give the principal quantum number n. In addition we have the magnetic spin quantum number m_s (154). In summary,

- n, the principal quantum number $n = 1, 2, 3, \ldots$.
- l, the orbital angular momentum quantum number, $l = 0, 1, \ldots, n-1$.
- m_l, the orbital magnetic quantum number, $m_l = 0, \pm 1, \ldots, \pm l$.
- m_s, the spin magnetic quantum number, $m_s = \pm\frac{1}{2}$.

At this stage, it is good to take a sidestep to recall the *periodic table of elements.* It was found that elements formed groups that had similar properties, for example the alkali metals Li, Na, K, Rb, Cs. The atomic weights of several elements were measured. In 1869 Dmitri

Mendeleev published a list were elements were arranged according to their atomic weights so that members of the groups of similar properties repeated periodically. With more measurements and new elements discovered, it was found that the periodicity is more accurate using the atomic number Z than the atomic mass. The periodicity in Z seemed to obey numbers $2, 8, 18, 32, \ldots$.

In 1925 Wolfgang Pauli was thinking the Bohr model and the periodic table. He came to the conclusion that the principal features of the periodic table could be explained by assuming a rule, now known as *Pauli exclusion principle*. The principle says that in an atom, no more than one electron can occupy a state with all same quantum numbers. That is, for any pair of electrons, at least one of the numbers n, l, m_l and m_s has to be different. (In 1925 the spin was not yet known, but Pauli hypothesized the existence of a two-valued quantum number, in addition to n, l, m_l.)

Let us apply the Pauli principle to explain the atomic structure and the periodic table. In this connection, the states with $n = 1, 2, 3, \ldots$ are often called K, L, M,... *shells*.

The ground state of hydrogen has one electron in state $n = 1$. It is called 1s *configurartion*. In helium the additional electron can also go to $n = 1$ orbital supposing the m_s numbers are different. That is, one electron has $m_s = +\frac{1}{2}$ and the other $m_s = -\frac{1}{2}$. He configuration is $1s^2$, the superscript 2 meaning two electrons. With these two electrons the K shell is full.

In Li, the third electron has to go to the L shell. There $l = 0$ and 1 are possible. It is found experimentally that the s state has lower energy. Thus Li has configuration $1s^2 2s$. The next ones are Be with $1s^2 2s^2$ and B with $1s^2 2s^2 2p$. There are 6 p states corresponding to $m_l = 0, \pm 1$ and $m_s = \pm\frac{1}{2}$. The L shells is full at $Z = 10$ Ne with configuration $1s^2 2s^2 2p^6$.

In Na, $Z = 11$ the configuration is [Ne]3s, continuing to Ar, $Z = 18$ and [Ne]$3s^2 3p^6$. After that could 3d orbitals be filled. However, it is found that 4s has lower energy, and is therefore filled first. Thus K, $Z = 19$ is [Ar]4s, and Ca, $Z = 20$ is [Ar]$4s^2$. Only after that are the 3d orbitals filled, with maximum 10 electrons, up to Zn, $Z = 30$, [Ar]$3d^{10} 4s^2$.

Instead of continuing this, let us look closer at some points. Although we cannot use the radial wave functions calculated above for hydrogen, we can expect that the ones with smaller n are more close to the nucleus. Consider Li as example. The two 1s electrons feel the nuclear charge $3e$ and thus are at $\sim a/3$ distance from the nucleus. Since the 2s electron is further a way, the 1s electrons *screen* the charge of the nucleus, so that the 2s electron approximately sees effective nuclear charge $1e$. Thus the binding energy 5.4 eV is not much larger than the binding energy 3.4 eV of $n = 2$ states in hydrogen. Thus Li can easily be ionized to free one electron. The same applies to all alkali metals.

Similar arguments can be used to excited states, measured by optical transitions. Consider as example Na. The ground state $E = -5.138$ eV corresponding to the M shell electron in state 3s. The excited states have $E = -3.023$ eV for 3p, $E = -1.521$ eV for 3d and $E = -1.947$ eV for 4s. Note, in contrast to hydrogen, the energies depend on l. The 3d energy is quite close to the hydrogen energy -1.51 eV. The 3s and 3p energies are lower since the radial wave function has more weight close to the nucleus and thus feels stronger attraction. For the same reason the 4s state has lower energy than -0.85 eV in hydrogen.

There are more detailed rules how the different orbitals are filled. We only mention the first Hund's rule. It says that the different orbitals with same n and l are filled so that the total spin, which is the sum of the spins of the electrons, has the maximum value that consistent with the exclusion principle. This means that in $2p^2$ and $2p^3$ configurations the total spin is $S = \hbar$ and $S = \frac{3}{2}\hbar$, respectively. In $2p^4$ the Pauli principle does not allow more than $S = \hbar$, and in $2p^5$ only $S = \frac{1}{2}\hbar$. (Exercise: why?)

(YF 41.7) The deep levels can be seen as characteristic X-ray radiation, see figure earlier. An electron accelerated in an X-ray tube may kick out an inner electron from an atom in the target material. This electron hole (= missing electron) may by filled by transition from an occupied higher orbital. The emitted quantum has the energy of the difference of the two level energies. This can be used to study the atomic structure of the target material.

The deep levels can also be seen in absorption of X rays as *absorption edge*. Studied as a function of energy $E = hf$, the absorption grows abruptly when the photon energy is sufficient to cause an emission of inner electron from an atom. (YF fig 41.26)

5.3 Pauli principle as symmetry

(YF 41.8) Above we have stated the Pauli exclusion principle using the quantum numbers n, l, m_l and m_s. This statement works in the central field approximation. For other cases, a more fundamental statement is needed. This is possible by specifying a symmetry of the many-body wave function (160).

As a first step, we should generalize the one particle wave function $\psi(\boldsymbol{r})$ to include the spin. We write the wave function $\psi(\boldsymbol{r}, \sigma)$ with the spin magnetic quantum number $\sigma = m_s = \pm\frac{1}{2}$ as a second argument. That is, we have two wave functions, $\psi(\boldsymbol{r}, +\frac{1}{2})$ for up spin and $\psi(\boldsymbol{r}, -\frac{1}{2})$ for down spin. For two-body wave function we then have

$$\psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2). \tag{166}$$

We now claim that the wave function should have *definite* *symmetry in exchange of the two particles*,

$$\boxed{\psi(\boldsymbol{r}_1,\sigma_1,\boldsymbol{r}_2,\sigma_2)=\pm\psi(\boldsymbol{r}_2,\sigma_2,\boldsymbol{r}_1,\sigma_1).} \tag{167}$$

The particles that obey the plus sign are called *bosons*. The particles that obey the minus sign are called *fermions*.

In order to see whether electrons are bosons or fermions, we apply the condition (167) to the wave functions we had above. We assume one particle is in wave function $\phi_a(\boldsymbol{r},\sigma)$ and the other in $\phi_b(\boldsymbol{r},\sigma)$. Here a and b are short hand notation for the set of quantum numbers $\{n,l,m_l,m_s\}$. In order to satisfy (167), the two particle wave functions should be

$$\psi(\boldsymbol{r}_1,\sigma_1,\boldsymbol{r}_2,\sigma_2)= \\ \frac{1}{\sqrt{2}}\left[\phi_a(\boldsymbol{r}_1,\sigma_1)\phi_b(\boldsymbol{r}_2,\sigma_2)\pm\phi_b(\boldsymbol{r}_1,\sigma_1)\phi_a(\boldsymbol{r}_2,\sigma_2)\right]. \tag{168}$$

[Exercise: check that (168) satisfies the condition (167).] Now test the case that the quantum number sets a and b are the same, $a=b$. We get from (168)

$$\psi(\boldsymbol{r}_1,\sigma_1,\boldsymbol{r}_2,\sigma_2)=\begin{cases}\sqrt{2}\phi_a(\boldsymbol{r}_1,\sigma_1)\phi_a(\boldsymbol{r}_2,\sigma_2) & \text{for bosons,}\\ 0 & \text{for fermions.}\end{cases} \tag{169}$$

The latter case is what describes electrons: having two electrons with the same quantum numbers, leads to vanishing of the wave function. That is, such a state does not exist. Thus we conclude *electrons are fermions.*

The symmetry condition (167) can be generalized to the presence of more particles. The requirement is that the symmetry is obeyed in any exchange of two particles coordinates.

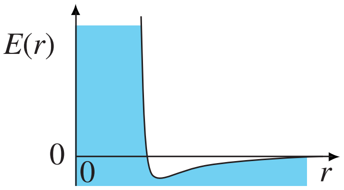
We mentions a general *spin-statistics law*. It states that all particles with half-integer spin are fermions and all particles with integer spin are bosons. Protons and neutrons are fermions, photons are bosons. (The polarization of the electromagnetic field of the photon can be interpreted as spin $s=1$.)

The symmetry of the many-body wave function implies that all electrons are identical. We cannot label them in any way, without affecting their behavior.

5.4 Chemical bond

(YF 42.1) An important application of quantum mechanics is to understand bonding between atoms. Unfortunately, this is also very difficult topic. Here we only list the main types of bonding.

At short distances the force between two atoms is repulsive. To large extend this is caused by the exclusion principle, that prevents two electrons to be in the same state. Also the Coulomb repulsion between electrons is essential. At larger distances the force is often attractive.



The figure depicts the potential energy of two atoms as a function of their distance r.

Ionic bond. Consider NaCl as an example. A simple description of the bonding is that Na gives its 3s electron to Cl, where it fills the last 3p state in the M shell. As a result, the Na^+ and Cl^- have the fully filled shells. The attraction then comes from the Coulomb attraction between the opposite charges.

Covalent bond. There is attractive force because the atoms share part of their electrons. Because the electrons can move in a larger region, the kinetic energy is reduced. This is similar as in potential well, (109) or (127), where the energy eigenvalues are reduced by increasing the well size L. There are two effects that can oppose the formation of covalent bond: the exclusion principle and the Coulomb repulsion between the electrons. (The former of these is studied as an exercise.)

Covalent bond prefers to form in definite directions. For example, carbon C tends to form four bonds, that are maximally separated form each other directions, in the directions of the corners of a tetrahedron (equivalently, every second corner of a cube). The four wave functions concentrated in these directions are obtained as linear combinations of the one s and three p wave functions.

The covalent bond appears in purest form between similar atoms, for example in O_2. The bonds between non-equal atoms has partly covalent and partly ionic nature.

Metallic bond. A great amount of atoms share some of their electrons, which can move almost like free particles through the crystal. The justification of the bonding is similar as for the covalent bond.

Hydrogen bond. Hydrogen has only one electron. When hydrogen bonds to oxygen, the main part of the electron wave function goes to the oxygen so that the hydrogen gets positive charge. This charge attracts a third atom, and this boding is called hydrogen bond. This is essential, for example, in water and ice.

Van der Waals bond. This gives a weak attraction also between uncharged atoms. The idea is that because of motion of electrons in an atom, the atom is an oscillating electric dipole. This causes an oscillating electric field, that can polarize a second atom close by. The electric interaction between the electric dipoles gives a weak attraction, where the interaction energy depends on the distance proportional to $1/r^6$. This bonding is always present, but becomes important only in case other bonding does not take place. Examples are boding in noble gases and between molecules like O_2, that cause them to liquify and solidify at low temperatures. ■

5.5 Molecules

(YF 42.2) Consider diatomic molecules. The rotational kinetic energy $E = \frac{1}{2}I\omega^2$, where ω is the angular velocity and I the moment of inertia. The angular momentum $L = I\omega$. Using this and the quantization of angular momentum (147), we get the energy levels

$$E = l(l+1)\frac{\hbar^2}{2I}. \tag{170}$$

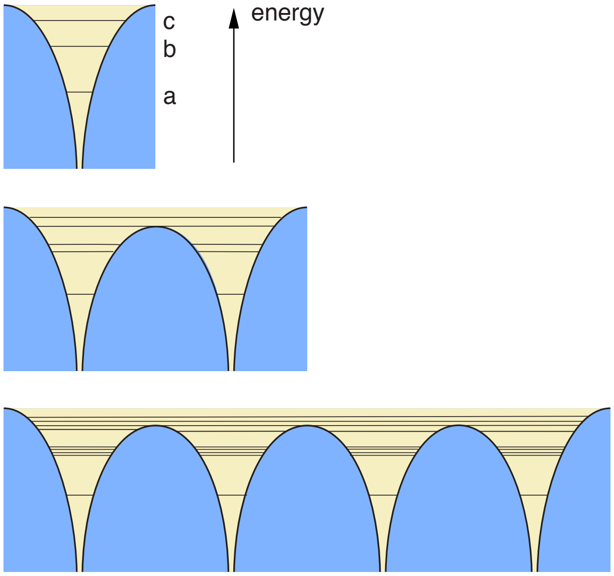
A vibration in diatomic molecule means oscillation in the distance of the two atoms. At not too high amplitudes, the potential is harmonic (the bottom of the $U(r)$ curve above). The energies are then given by the harmonic oscillator energies $E = \hbar\omega_0(n + \frac{1}{2})$ (114). Here $\omega_0 = \sqrt{k/m}$ is the oscillation frequency determined by spring constant k of the bond and the reduced mass m (159) of the two-atom system.

Transitions in molecules are of three types: 1) transitions between electronic configurations, typically in visible light and ultraviolet. 2) vibrations and 3) rotations. The two latter are in the infrared region. The transparency of gases in the visible range is a result of the transition from the ground state to the lowest excited configuration being in ultraviolet and the vibrations and rotations being in infrared region.

6. Solid state

(YF 42.4) We now start applying quantum mechanics to the solid state. It should be reminded that the following is based on the independent electron approximation, where the Coulomb interaction between the electrons is neglected. The approximation works surprisingly well for many properties, but not all.

Above we have considered electron states in an atom. Now we study what happens to these when atoms first make a molecule and then a solid. The figure below depicts three energy level in an atom, marked by a, b and c.



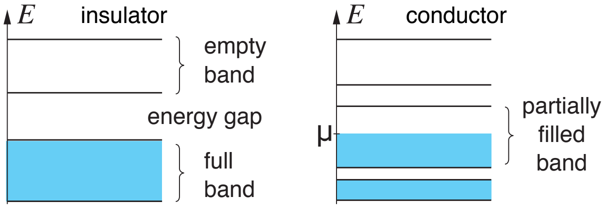
When two atoms are brought close to each other, the energy barrier between them gets lower. This means that the electrons from one atom can tunnel to the other one. However, for the state a the potential barrier is high, and nothing really happens. This is an important result: the inner electron states in an atom remain unchanged in the formation of a bond.

At level b the electron tunneling is important. At level c the electrons can move freely in the molecule.

When the atoms form a bond, the energy levels are restructured. In a molecule of two atoms, an atomic energy level froms two levels at different energies. The energy difference of these levels is the larger the stronger is the tunneling between the atoms. In a chain of 4 atoms, each atomic level is split into 4 levels of different energy. A solid has a great number of atoms, $N \sim 10^{23}$. This implies that instead of separate energy levels, there are *energy bands.* An energy in energy band is allowed for electrons. Between energy bands there are forbidden energies, that are called *energy gaps.* In some cases the energy bands arising from different atomic levels overlap, implying that there is no energy gap.

Similar as in atoms, the electrons in a solid fill the energy levels starting from the lowest energy. Solids can be divided in three groups based on the filling of the energy bands.

- Insulator: all bands are either filled or empty. The highest filled band is separated from the lowest empty band by an energy gap of several electron volts.
- Semiconductor: like insulator, but the band gap is smaller so that some electrons are excited from the valence band (the highest filled band) to the conduction band (the lowest empty band).
- Conductor: there is one partially filled band (or more). (”Metal ”and ”conductor” are used as synonyms.)



6.1 Metals

(YF 42.5) In metals the partly filled band is called *conduction band*. The ability of metals to conduct electricity is based on the existence of a conduction band. This is because the bands that are full do not conduct. (We will discuss this soon.) The electrons in the conduction band are called *conduction electrons*. These can move rather freely through the metal. We now study their behavior in more detail.

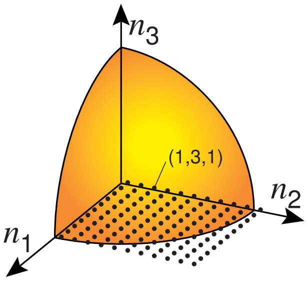
Free electron model

We assume electrons are moving in a potential $U(\boldsymbol{r})$ that is constant, $U \equiv U_0$. Since the constant does not affect the motion, we choose for simplicity $U_0 = 0$. The wave functions have been solved above. They are plane waves (120) with energy $E = \hbar^2 k^2/2m$.

The next step is the following. Given the density of conduction electrons in a metal, up to what energy are the states filled. For that we consider a lump of metal and count the number of levels below a given energy E. We use the deep 3D potential well as a model of a piece of metal. The energy levels are given by (127) with positive integers n_1, n_2 and n_3. The number of wave functions below E correspond to quantum numbers satisfying

$$n_1^2 + n_2^2 + n_3^2 < \frac{2mL^2E}{\pi^2\hbar^2} = \kappa^2. \tag{171}$$

We draw a space with axes n_1, n_2 and n_3. In this space the condition (171) means dots inside 1/8 of sphere of radius κ.



In the limit the box is large, the dots are very dense, $\kappa \gg 1$. Then the number of dots $N_d(E)$ can be estimated from the volume of the 1/8 sphere

$$\begin{aligned} N_d(E) &= \frac{1}{8}\left(\frac{4}{3}\pi\kappa^3\right) \\ &= \frac{\sqrt{2m^3}V}{3\pi^2\hbar^3}E^{3/2}, \end{aligned} \tag{172}$$

where the box volume $V = L^3$.

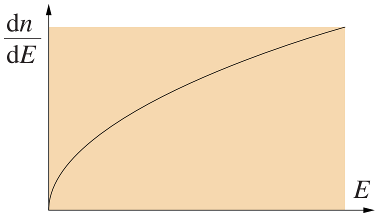
We define the density (number/volume) of electron states $n(E)$ below energy E. It is given $n(E) = 2N_d(E)/V$, where the factor 2 comes from two values of the spin quantum number m_s. We get

$$n(E) = \frac{2\sqrt{2m^3}}{3\pi^2\hbar^3}E^{3/2}. \tag{173}$$

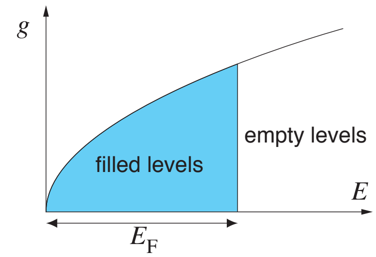
We also define *density of states*

$$g(E) = \frac{dn}{dE} = \frac{\sqrt{2m^3}}{\pi^2\hbar^3}\sqrt{E}. \tag{174}$$

Here “density” means the number of states per energy interval. The number of states in energy interval $(E, E + dE)$ is $g(E)dE$. This is proportional to the square root of energy.



Suppose the enectron density n_e. These fill the electron states from 0 to an energy E_{F} called *Fermi energy.*



The condition $n_e = n(E_{\mathrm{F}})$ gives, using (173),

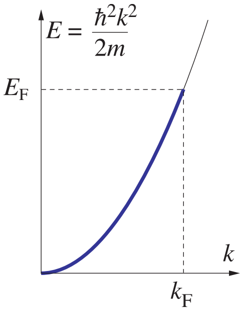
$$E_{\mathrm{F}} = \frac{\hbar^2}{2m}\left(3\pi^2 n_e\right)^{2/3}. \tag{175}$$

We apply the free electron model to the electrons in the conduction band. We suppose each atom contributes Z_c elecrrons to the conduction band. The following table shows estimation of Z_c for some elements. When we know the density of the element, one can calculate n_e, and the Fermi energy (175).

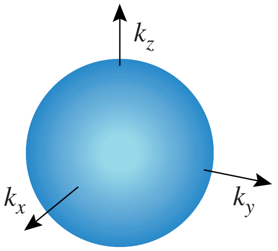
	Z_c	E_{F} (eV)
Na	1	3.2
Cu	1	7.0
Fe	2	11.1
Al	3	11.7
Pb	4	9.5

We derived that the energy levels are filled up to Fermi energy (175). According to $E = \hbar^2 k^2/2m$ there is corresponding *Fermi wave nunber*

$$k_{\mathrm{F}} = \left(3\pi^2 n_e\right)^{1/3}. \tag{176}$$



We can also think three dimensional $\boldsymbol{k}$ space with axes k_x, k_y and k_z. In this space, the filled states are inside a sphere of radius k_{F}. This is called the *Fermi sphere* and its surface the *Fermi surface*.

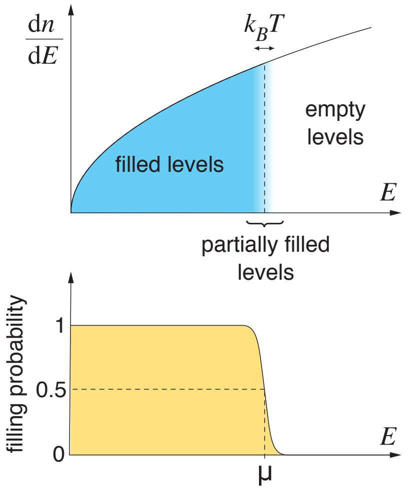


One also defines Fermi momentum $p_{\mathrm{F}} = \hbar k_{\mathrm{F}}$ and Fermi velocity $v_{\mathrm{F}} = p_{\mathrm{F}}/m$. The typical order of magnitude $v_{\mathrm{F}} \sim 10^6$ m/s.

More generally, at a given temperature, the levels are filled according to the Fermi-Dirac distribution,

$$f(E) = \frac{1}{e^{(E-\mu)/k_{\mathrm{B}}T} + 1}. \tag{177}$$

The change of the occupation from 1 to 0 takes place in energy range characterized by the thermal energy $k_{\mathrm{B}}T$. At room temperature $k_{\mathrm{B}}T = 0.025$ eV $\ll E_{\mathrm{F}}$.



Electric conduction in metals

The simplest theory for electron dynamics is the free electron model. This was first studied by Drude (1900). The Drude model considers the momentum $\boldsymbol{p}$, which is the averaged value over all conduction electrons (those ones in the partially filled band). Its equation of motion is assumed be

$$\frac{d\boldsymbol{p}}{dt} = -\frac{\boldsymbol{p}}{\tau} + \boldsymbol{F}. \tag{178}$$

This is Newton's equation of motion, with the additional term $-\boldsymbol{p}/\tau$. This term takes into account that the electrons loose their momentum in collisions on the average in a time τ, which is called the relaxation time. The force term can arise from external electric field $\boldsymbol{F} = -e\boldsymbol{\mathcal{E}}$, where $-e$ is the charge of an electron. The momentum is related to the electron velocity by the standard relation $\boldsymbol{p} = m\boldsymbol{v}$, where m is the mass. Drude originally assumed m to be the electron mass. The electric current density $\boldsymbol{j} = -en_e\boldsymbol{v}$, where n_e is the number density of the conduction electrons.

One particular case that can easily be solved is the response to a constant electric field $\boldsymbol{\mathcal{E}}$. Setting $d\boldsymbol{p}/dt = 0$, one gets the Ohm's law, stating that the current density is proportional to the electric field,

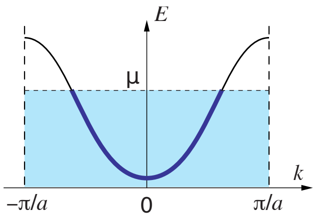
$$\boldsymbol{j} = \sigma\boldsymbol{\mathcal{E}}, \tag{179}$$

where the *conductivity* σ is given by

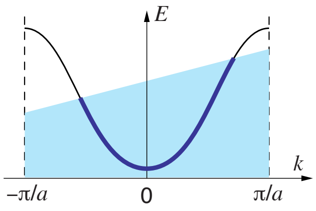
$$\sigma = \frac{e^2 n_e \tau}{m}. \tag{180}$$

The *resistivity* $\rho = 1/\sigma$. The *resistance* R of a piece of wire of length L and cross-sectional area A is $R = \rho L/A$.

A more realistic model than the free electron one is to take into account the dispersion $E(k)$ caused by the crystal lattice of the solid. In a solid with a period a, also known as the *lattice constant*, the conduction band could look like in the figure below.



electron flows to the right
and to the left cancel



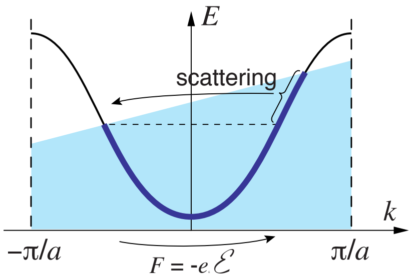
net electron flow to the right

In equilibrium the energy levels are filled up to the Fermi level μ. In equilibrium, corresponding to a filled level with $\boldsymbol{k}$ and velocity $\boldsymbol{v}$, also the level with $-\boldsymbol{k}$ and $-\boldsymbol{v}$ is filled. Thus, in spite of rapidly moving electrons, there is no net current, $\boldsymbol{j} = 0$. We also see that a full band cannot conduct for the same reason.

In order to have electric current, $j \neq 0$, there has to be asymmetry in the occupation of some levels with $\boldsymbol{k}$ and $-\boldsymbol{k}$. This is sketched in the right figure above, where the levels with velocity to the right are occupied to higher energy than the ones moving to the left. (Warning: the asymmetry is strongly exaggerated.) This conducting state could be achieved as follows. The force $F = -e\mathcal{E}$ accelerates the electrons so that their momentum $p = \hbar k$ grows. The imperfections of the lattice

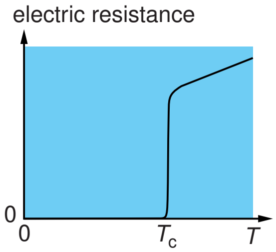
- lattice vibrations
- static imperfections of the lattice: atoms missing, extra atoms, wrong type of atoms, dislocations, grain boundaries. . .

cause scattering of the electrons. The electrons predominantly scatter to empty levels that are at the same or lower energy. A stationary current flow is reached when the driving electric field and the scattering balance each other.



Superconductivity

(YF 42..8) By superconductivity we mean a phenomenon where the electrical resistivity of a material disappears below some temperature T_c. In elemental metals T_c is typically a few kelvin (see table at the end of this section).



What causes superconductivity: the formation of *pairs* caused by attractive electron-electron interaction. The many-body wave function takes the form

$$\begin{aligned}\psi_0(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2, \boldsymbol{r}_3\sigma_3, \ldots) = \phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) \times \\ \phi(\boldsymbol{r}_3\sigma_3, \boldsymbol{r}_4\sigma_4)\phi(\boldsymbol{r}_5\sigma_5, \boldsymbol{r}_6\sigma_6)\ldots, \qquad (181)\end{aligned}$$

All *pair states are the same*!. This function does *not* vanish in antisymmetrization as long as the pair function is antisymmetric,

$$\phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) = -\phi(\boldsymbol{r}_2\sigma_2, \boldsymbol{r}_1\sigma_1). \qquad (182)$$

In exchanging pairs one gets the factor $(-1)^2 = 1$, similarly to bosons.

Consider a pair of particles. Assuming the pair is at rest, the wave function can depend only on the relative location $\boldsymbol{r}_1 - \boldsymbol{r}_2$, not on the center of mass coordinate $\boldsymbol{r} = \frac{1}{2}(\boldsymbol{r}_1 + \boldsymbol{r}_2)$. We also assume the spin state is a singlet, $\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow)$, which is believed to be the case in most superconductors. Thus

$$\phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) = \psi(\boldsymbol{r}_1 - \boldsymbol{r}_2)\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow). \qquad (183)$$

If we now look the orbital part ψ in k space, we see that it is a superposition of states, where, if one particle is in state $\boldsymbol{k}$, the other is in the opposite state $-\boldsymbol{k}$:

$$\psi(\boldsymbol{r}_1 - \boldsymbol{r}_2) = \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}_1} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_2} \qquad (184)$$

This is understandable because the pair is assumed to be at rest, and therefore the total momentum should vanish.

A special case of the pair state (184) is that $c_{\boldsymbol{k}}$ vanishes for all wave vectors except one, say $\boldsymbol{k} = \boldsymbol{k}_0$. Such a state, which describes two fermions with wave vectors $\boldsymbol{k}_0$ and $-\boldsymbol{k}_0$, is an energy eigenstate if there is no interaction between the particles. In the presence of interaction, the two particles will collide and scatter to some other wave vectors $\boldsymbol{k}_1$ and $-\boldsymbol{k}_1$. These then scatter to wave vectors $\boldsymbol{k}_2$ and $-\boldsymbol{k}_2$, and so on. Thus the energy eigenstate in the presence of interactions is a coherent superposition of opposite wave-vector states as in (184) with many non-vanishing $c_{\boldsymbol{k}}$'s.

Let us look the pair state from the point of a fixed single $\boldsymbol{k} = \boldsymbol{k}_0$. The states included in (184) are $|1, 1\rangle$ and $|0, 0\rangle$. Here the two numbers mean the occupation of levels with wave vectors $\boldsymbol{k}_0$ and $-\boldsymbol{k}_0$, respectively. The presence of

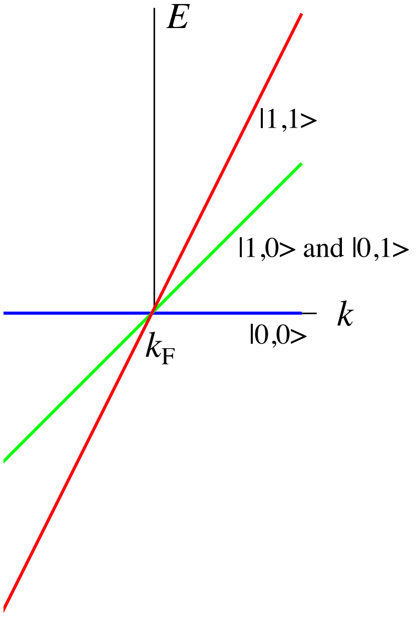
$|0,0\rangle$ state means that the particles are scattered into some other wave-vector states leaving the levels $\boldsymbol{k}_0$ and $-\boldsymbol{k}_0$ empty. In order to have complete basis states for wave vectors $\boldsymbol{k}_0$ and $-\boldsymbol{k}_0$, we also need states $|1,0\rangle$ and $|0,1\rangle$. These mean states where only the former or the latter level is occupied, repectively.

Let us study the energies of the states $|0,0\rangle$, $|1,0\rangle$, $|0,1\rangle$ and $|1,1\rangle$. These contain different number of particles. The standard method to deal with such a case is to connect the system with a particle reservoir. An ideal particle reservoir can store particles at a constant energy. The energy μ is called chemical potential. In the free electron model μ is equal to E_F at $T=0$.

We can now calculate the energies of the four states in the absence of interactions. It is convenient to choose μ as zero of energy, and to express other energies relative to that. We get

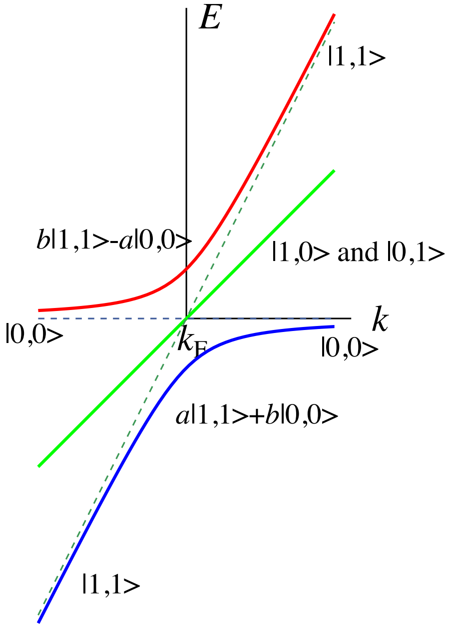
$$\begin{aligned} |0,0\rangle &: E=0 \\ |1,0\rangle &: E=\xi_k \\ |1,0\rangle &: E=\xi_k \\ |1,1\rangle &: E=2\xi_k \end{aligned} \tag{185}$$

where $\xi_k=\hbar v_F(k-k_F)$. The energies are illustrated in the figure as function of k.



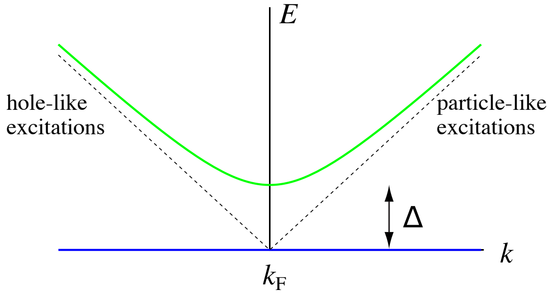
It is obvious that the system ground state consist of states $|1,1\rangle$ at $k<k_F$ and of states $|0,0\rangle$ at $k>k_F$. This is the Fermi sphere of noninteracting particles.

Let us then consider what happens when the attractive interaction is switched on. As discussed above the formation of pair state leads to coupling between different momentum states, and this leads to coupling of states $|1,1\rangle$ and $|0,0\rangle$. It is a standard problem in quantum mechanics that such a coupling leads strongest mixing of the states when the unperturbed energies are close to each other. In the present case this happens just at the Fermi surface ($k=k_F$), where the states $|1,1\rangle$ and $|0,0\rangle$ are degenerate in the absence of interaction. This leads to the *avoided crossing*, which is a common phenomenon in quantum mechanics (or in any wave theory).



The ground state (blue in the figure) is now a superposition of states $|1,1\rangle$ and $|0,0\rangle$. However, deep inside the Fermi sphere ($k\ll k_F$) it is to a good approximation $|1,1\rangle$, and far outside ($k\gg k_F$) it is to a good approximation $|0,0\rangle$, as in the normal state. Thus pairing only changes the electron state near the Fermi surface, $k\approx k_F$. But this is just the region that is essential for transport properties at low temperatures.

In addition to the superconducting ground state, the picture above gives also the elementary excited states. Namely, the lowest energy exited state is to break the pair a some $\boldsymbol{k}$ and replace it by state $|1,0\rangle$ or $|0,1\rangle$. The excitation energy is just the energy difference between the green and blue lines in the figure above. The excited pair states (red in the figure) correspond to double excitation at the same $\boldsymbol{k}$. The excitation energy of the elementary excitations is replotted in the figure below.



We see that there is minimum of the excitation energy. It is called the superconducting energy gap and denoted by Δ. This is in contrast to excitations in the normal state which have no gap (shown by dashed line).

	T_c (K)	Δ (eV)
Al	1.2	1.7×10^{-4}
Sn	3.75	5.6×10^{-4}
Pb	7.22	13×10^{-4}

6.2 Insulators

Insulators have a large band gap between the full valence band and the empty conduction band. The table lists band gaps of some insulators and semiconductors.

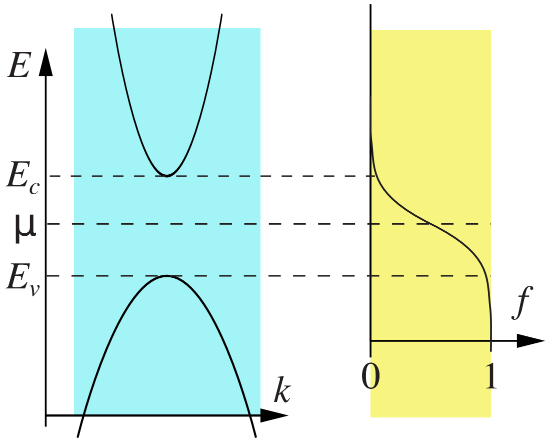
	E_g (eV)
C (diamond)	5.5
Si	1.12
Ge	0.67
InSb	0.17
GaAs	1.4
NaCl	9.0
Al_2O_3	9.9
SiO_2	11

A full band does not conduct electricity. Conduction is possible by excitation of electrons across the band gap, but this is unlikely because of the large band gap.

Pure insulators are transparent to visible light if the band gap is larger than $E = hf$ for visible light. Insulators absorb light in ultraviolet, corresponding to the frequency of the band gap. The strong dispersion of the refractive index for visible light in some materials is related to the band gap being near in the ultraviolet. Impurities in insulating materials make them absorb light in the visible range, for example in natural stones.

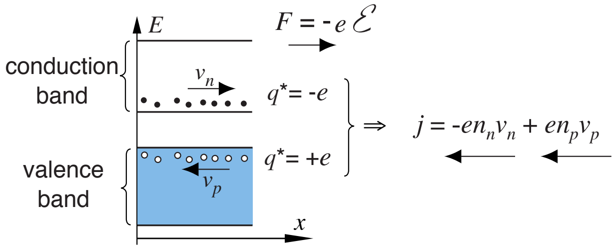
6.3 Semiconductors

(YF 42.6) Pure semiconductors in ground state ($T = 0$) have all bands either filled or empty. The highest filled band is called the *valence band*, and lowest empty band the *conduction band*. The band gap between them is smaller than in insulators, typically 2 eV or smaller.



The right hand side of the figure shows the Fermi function (177). The Fermi level energy μ (where $f = 1/2$) is in the band gap. Because the band gap is not too big compared to the thermal energy k_BT, f is small but not negligible at E_c (lowest energy in the conduction band). Thus there are some electrons in the conduction band. Similarly, there is a small difference of f from unity at E_v (the highest energy in the valence band). This means that there are holes (= missing electrons) in the valence band as some electrons are excited to the conduction band.

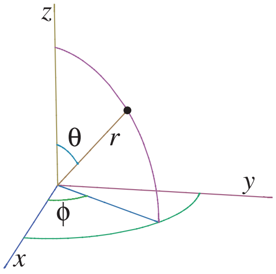
The eletric conduction in semiconductors takes place both by electrons in the conduction band and holes in the valence band. This is illustrated in the figure below (where q^* is the effective charge).



There would be much interesting to tell more, but we have to leave it for this course. Read more about semiconductors from the book.

Appendixes

A. Spherical coordinates



The spherical coordinates r, θ and ϕ are defined by writing

$$\boldsymbol{r} = r\sin\theta(\cos\phi\,\hat{\boldsymbol{x}} + \sin\phi\,\hat{\boldsymbol{y}}) + r\cos\theta\,\hat{\boldsymbol{z}}. \tag{186}$$

Alternatively

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta. \tag{187}$$

All $\boldsymbol{r}$ can be described by the coordinates in the ranges $0 \le r$, $0 \le \theta \le \pi$ and $0 \le \phi < 2\pi$.

At each point $\boldsymbol{r}$, one can define the basis vectors

$$\begin{aligned}
\hat{\boldsymbol{r}} &= \frac{\partial \boldsymbol{r}}{\partial r} = \sin\theta\cos\phi\,\hat{\boldsymbol{x}} + \sin\theta\sin\phi\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}} \\
\hat{\boldsymbol{\theta}} &= \frac{1}{r}\frac{\partial \boldsymbol{r}}{\partial \theta} = \cos\theta\cos\phi\,\hat{\boldsymbol{x}} + \cos\theta\sin\phi\,\hat{\boldsymbol{y}} - \sin\theta\,\hat{\boldsymbol{z}} \\
\hat{\boldsymbol{\phi}} &= \frac{1}{r\sin\theta}\frac{\partial \boldsymbol{r}}{\partial \phi} = -\sin\phi\,\hat{\boldsymbol{x}} + \cos\phi\,\hat{\boldsymbol{y}}.
\end{aligned} \tag{188}$$

These three vectors form an orthonormal vector basis. Orthonormality means that they are orthogonal to each other and have unit length. The latter property is achieved by the factors $1/r$ and $1/r\sin\theta$ in the definition of the unit vectors. The factor $r\partial\theta$ can be understood as the distance traveled when θ changes by $\partial\theta$. Similarly $r\sin\theta\partial\phi$ is the distance travelled when ϕ changes by $\partial\phi$. With change of r by ∂r, the distance travelled is ∂r, and no extra factor is needed.

In cartesian coordinates the operator $\boldsymbol{\nabla}$ is defined by

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}. \tag{189}$$

Using this we can define the gradient $\boldsymbol{\nabla}\Phi$ of a scalar function $\Phi(x,y,z)$ by

$$\boldsymbol{\nabla}\Phi = \hat{\boldsymbol{x}}\frac{\partial \Phi}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial \Phi}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial \Phi}{\partial z}. \tag{190}$$

We now claim that in spherical coordinates

$$\boldsymbol{\nabla} = \hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}. \tag{191}$$

We can recognize here the same factors as in the definition of the basis vectors (188). The gradient of a scalar function $\Phi(r,\theta,\phi)$ is

$$\boldsymbol{\nabla}\Phi = \hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}. \tag{192}$$

The Laplace operator is defined as $\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}$. In cartesian coordinate we get from (189)

$$\begin{aligned}
\nabla^2\Phi &= \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\Phi \\
&= \left(\hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}\right)\cdot\left(\hat{\boldsymbol{x}}\frac{\partial \Phi}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial \Phi}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial \Phi}{\partial z}\right) \\
&= \hat{\boldsymbol{x}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial x^2} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial x\partial y} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial x\partial z}\right) \\
&\quad + \hat{\boldsymbol{y}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial y\partial x} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial y^2} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial y\partial z}\right) \\
&\quad + \hat{\boldsymbol{z}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial z\partial x} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial z\partial y} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial z^2}\right) \\
&= \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}
\end{aligned} \tag{193}$$

The last form is obtained because of the orthonormality of the basis $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$.

In spherical coordinates we have

$$\begin{aligned}
\nabla^2\Phi &= \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\Phi \\
&= \left(\hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\right) \\
&\quad \cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&= \hat{\boldsymbol{r}}\frac{\partial}{\partial r}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&\quad + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&\quad + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right)
\end{aligned} \tag{194}$$

The added complication here compared to the cartesian case (193) is the extra factors $1/r$ and $1/r\sin\theta$ but also that the unit vectors (188) depend on θ and ϕ. The following derivatives will lead to nonvanishing terms in the Laplace operator:

$$\frac{\partial \hat{\boldsymbol{r}}}{\partial \theta} = \hat{\boldsymbol{\theta}}, \quad \frac{\partial \hat{\boldsymbol{r}}}{\partial \phi} = \sin\theta\hat{\boldsymbol{\phi}}, \quad \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \phi} = \cos\theta\hat{\boldsymbol{\phi}}. \tag{195}$$

In the same order as in (194), we get 6 nonvanishing terms

$$\begin{aligned}
\nabla^2\Phi &= \frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \theta^2} \\
&\quad + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial \Phi}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 \Phi}{\partial \phi^2}.
\end{aligned} \tag{196}$$

Verify that this is equivalent to

$$\begin{aligned}
\nabla^2\Phi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial \Phi}{\partial \theta}\right) \\
&\quad + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 \Phi}{\partial \phi^2}.
\end{aligned} \tag{197}$$