QUANTUM MATERIALS
PHYS-C0250

Erkki Thuneberg

Department of Applied Physics
Aalto University
2024

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=42770

The web page contains the lecture material (these notes), advance assignments, exercises and other material.

Time table 2024

Lectures on Mondays and Wednesdays 12.15-14.00 (21.10.-27.11.).
Exercise sessions on Thursdays 8.15-10.00 and Fridays 10.15-12.00 (24.10.-29.11.).
All lectures and exercises in nanotalo room 228.

Teaching assistants: Tomas Capiten and Qingxin Yang

Advance assignments are small task that are aimed to prepare you for the lecture. They will be placed on the web page some days prior each lecture, and they should be completed before the lecture. The plan is to put each week's exercises in web page on Mondays soon after the lecture. There are two exercise sessions (Thursdays and Fridays). You can go to one or the other, or to both. Try to solve the problems before the exercise session, so that during the session you can pose questions to the assistant. You still have time until next Monday to show completed exercises.

Exam Thursday 5.12. at 13.00–16.00

The evaluation criteria are approximately as follows: advance assignments 10%, exercises 20%, examination 70%. Doing exercises is essential for learning, and thus important for your performance in the examination. The direct effect of the exercises to the evaluation should be seen as a small bonus on top of that.

Relation to other courses

This course is part of quantum technology bachelor program. Some related courses are listed below.

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.

PHYS-C0252 Quantum Mechanics (QMec). This course has introduced you to the formalism and some applications of quantum mechanics.

The present course, Quantum Materials. The aim is to learn how quantum mechanics can be used to understand the properties of matter, from atoms to molecules and condensed matter. Note that this course has different content than the course with the same name that was lectured in 2021 and 2022.

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

The title of this course "quantum materials" is used to express that all materials (gases, liquids and solids) have basic structure dictated by quantum mechanics. Just as a warning, the same name is sometimes used to refer to a more limited class of materials (for example, materials with 2 dimensional or topological structure).

Content

- Reminder of Schrödinger equation in 1 dimension and generalization to 3 dimensions
- Schrödinger equation applied to hydrogen atom: energy levels, angular momentum, electron spin
- Many-electron systems, justification for the periodic table of elements
- Atomic bonding, transitions in molecules
- Crystal structure of solids
- Lattice vibrations in solids, classical and quantum models and heat capacity.
- Electronic structure of solids, free electron model, electrons in a periodic potential, energy bands, semiclassical electron motion
- Metals: electrical conduction, superconductivity: phenomenology and pairing theory
- Semiconductors: intrinsic, n and p types, junctions and devices
- Electronic transitions

Learning outcomes (from Sisu):
After the course, the student has understanding on the atomic and solid state structure and can solve basic problems on the topics listed in the content.

Prerequisites

Basic knowledge of classical mechanics (Newton's equation of motion, work and energy, harmonic oscillator). Basic quantum mechanics [Schrödinger equation applied to one dimensional (1 D) problems].

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

Books

- Young and Freedman, University Physics (14th Edition), Pearson (YF). Library link. This is a vast book with 1600 pages. What is useful for the present course are chapters 41 “Quantum mechanics II: atomic structure” and 42 “Molecules and condensed matter”. The presentation of solid state physics, in particular, in this book is shorter than we will have in this course.
- S.H. Simon, The Oxford Solid State Basics (S). Library link. Year 2013, 290 pages. This is very nice introductory book of solid state physics, which goes much more in details than YF. No discussion of superconductivity.

The lecture notes below contain references to sections of both books (YF and S) that contain similar discussion. Not all topics presented in the lectures can be found in the two books. The way theory of superconductivity is presented here will not be found in any book or publication, as far as I know, although the basic theory is the same as explained in practically all books of superconductivity, also in my lecture notes.

Other good books

- N. Ashcroft ja D. Mermin, Solid state physics (AM). Year 1976, 826 pages. Very clear book in theoretical basics but lacks modern developments. Solid state physics seen from the point of a theoretical physicist, but still with minimal formalism. (warning: cgs units)
- M. Tinkham, Introduction to Superconductivity (1975, 1996). Very widely used book of superconductivity.

Notation, constants and formulas

$a/bc = \frac{a}{bc}$ 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. Quantum mechanics of a single particle

1.1 Schrödinger equation

(YF 40.1) The *Schrödinger equation* is

$$\boxed{i\hbar\frac{\partial\Psi}{\partial t}(x,t)=-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}(x,t)+U(x)\Psi(x,t).} \tag{1}$$

This equation was deduced starting from the formal theory of quantum mechanics in the course QMec. In the present course, we take the Schrödinger equation as our starting point. In this section we try to clarify the meaning of this equation.

Let us recall the quantities contained in the Schrödinger equation (1). $\Psi(x,t)$ is the *wave function*. It is also called *state*. It depends on the spatial coordinate x and on time t. It describes what we know of a particle that is allowed to move on the x axis. The Schrödinger equation contains partial derivatives of $\Psi(x,t)$, the first partial derivative with respect to t, and the second partial derivative with respect to x,

$$\frac{\partial\Psi}{\partial t}(x,t),\quad \frac{\partial^2\Psi}{\partial x^2}(x,t). \tag{2}$$

The partial derivative of a function of several variables is like the usual derivative while considering the other variables as constant. In addition (1) contains the Planck constant $\hbar=h/2\pi$, the mass m of the particle and the potential energy $U(x)$. On the left hand side is i, the imaginary unit. The wave function $\Psi(x,t)$ is complex valued. That is, it can be represented with two real functions $\Re\Psi(x,t)$ and $\Im\Psi(x,t)$ [called real and imaginary parts of $\Psi(x,t)$] so that $\Psi(x,t)=\Re\Psi(x,t)+i\Im\Psi(x,t)$. The absolute square of the wave function is

$$|\Psi(x,t)|^2=\Psi(x,t)\Psi^*(x,t)=[\Re\Psi(x,t)]^2+[\Im\Psi(x,t)]^2, \tag{3}$$

where $\Psi^*=\Re\Psi-i\Im\Psi$ is the complex conjugate of Ψ.

The physical interpretation of the wave function is that the probability to find the particle at location x is proportional $|\Psi(x,t)|^2$. In order to express this more precisely we define the *normalization integral*

$$I=\int|\Psi(x,t)|^2dx. \tag{4}$$

In many cases, I has a finite value ($\neq 0$, $\neq\infty$). When this holds, we can "normalize" the wave function by dividing by $\sqrt{I}$. As a result, we have a new $\Psi(x,t)$ that satisfies the *normalization condition*

$$\int|\Psi(x,t)|^2dx=1. \tag{5}$$

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

$$P(x,t)=|\Psi(x,t)|^2 \tag{6}$$

so that $P(x,t)dx$ gives the probability to find the particle within a small range dx around x. For example, in the interval $(x-\frac{1}{2}dx, x+\frac{1}{2}dx)$.

Mathematically, the Schrödinger equation (1) is a *partial differential equation* because it expresses a condition for a function of two variables and its partial derivatives. The idea is to find the wave function so that equation (1) is satisfied at all values of x and t in some appropriate domain. For example, the domain could be $-\infty<x<\infty$ and $t>t_0$. Often there is an *initial condition* that $\Psi(x,t_0)$ equals a given function of x at some initial time t_0 and/or there are *boundary conditions* that fix the behavior of $\Psi(x,t)$ at certain values of x.

The Schrödinger equation (1) should be compared with Newton's equation of motion. The latter for a particle moving along x is

$$m\frac{d^2x}{dt^2}(t)=F(x),\quad F(x)=-\frac{dU}{dx}(x), \tag{7}$$

where F is the force acting on the particle. This is an *ordinary differential equation* (not partial) for the function $x(t)$, the location of the particle as a function of time. Newton's equation of motion (7) is valid in *classical mechanics*. In quantum mechanics it is replaced by the Schrödinger equation (1).

An important property of the Schrödinger equation (1) 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.*

Of particular interest are states that have the form

$$\Psi(x,t)=\psi(x)e^{-i\omega t}, \tag{8}$$

where the time dependence is complex sinusoidal with angular frequency ω. Such states are called *stationary* since the probability density (6) is independent of time. Substituting (8) into the Schrödinger equation allows to cancel the factors $e^{-i\omega t}$. Defining $E=\hbar\omega$ we are left with *time-independent Schrödinger equation*

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}(x)+U(x)\psi(x)=E\psi(x). \tag{9}$$

Since ψ is a function of a single variable, the partial derivative has been replaced by ordinary derivative.

In oder to get better understanding of the Schrödinger equations (1) and (9), we define the *momentum operator*

$$p_{\mathrm{op}}=\frac{\hbar}{i}\frac{\partial}{\partial x}. \tag{10}$$

As an example, consider the sinusoidal wave

$$\Psi(x,t)=Ae^{i(kx-\omega t)}\quad\Leftrightarrow\quad\psi(x)=Ae^{ikx}. \tag{11}$$

Applying the momentum operator (10) gives

$$p_{\rm op}\Psi(x,t)=\frac{\hbar}{i}\frac{\partial}{\partial x}Ae^{i(kx-\omega t)}=\hbar k\,Ae^{i(kx-\omega t)}. \tag{12}$$

We express the result

$$p_{\rm op}\Psi(x,t)=\hbar k\,\Psi(x,t). \tag{13}$$

It is called that Ψ (11) 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 (11), 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 (13) is valid, the momentum of the state Ψ is precisely defined.

We see that the first term on left hand side of the Schrödinger equation (9) is operator $p^2_{\rm op}/2m$ acting on the wave function. This can be interpreted as the kinetic energy operator, as a particle with momentum p has kinetic energy $p^2/2m$. The second term on the left hand side can be seen as potential energy operator acting on ψ (where the operation is simply multiplication). Thus the operator on the left hand side acting on ψ, known as *Hamilton operator*

$$H=-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+U(x) \tag{14}$$

is the total energy operator. This allows to interpret E on the right hand side of (9) as the total energy. Thus the stationary states [solutions of (9)] are also known as *energy eigenstates*, as they have definite value of energy. Also names *energy levels* or *levels* are used.

Above we deduced the relation $E=\hbar\omega$. Since the frequency $\nu=\omega/2\pi$, and $h=2\pi\hbar$, this can also be expressed by

$$E=h\nu. \tag{15}$$

This relation is the same as suggested for quanta of electromagnetic radiation (photons) by Max Planck in 1900. Also note that k in (11) is the wave number. It is related to the wave length λ by the standard relation $k=2\pi/\lambda$, and thus

$$p=\frac{h}{\lambda}. \tag{16}$$

This relation was suggested for electron waves by de Broglie in 1924. Both relations (15) and (16) were suggested before the full theory of quantum mechanics was developed. We see that both are incorporated in the Schrödinger equation (1).

The momentum operator (10) is an example that in quantum mechanics physical quantities are represented by operators. Let O be an operator (which are hermitian in quantum mechanics). When

$$O\Psi_\lambda(x,t)=\lambda\Psi_\lambda(x,t), \tag{17}$$

where λ is a number, we say that $\Psi_\lambda(x,t)$ is an eigenstate of O with eigenvalue λ. More general wave functions can be presented as sum over different eigenstates,

$$\Psi(x,t)=\sum_\lambda c_\lambda\Psi_\lambda(x,t). \tag{18}$$

Measuring the quantity O in this state gives the result λ with probability $|c_\lambda|^2$ (assuming normalized states). The expectation value is

$$\langle O\rangle=\int\Psi^*(x,t)O\Psi(x,t)\,dx. \tag{19}$$

For example, the momentum

$$\langle p\rangle=\int\Psi^*(x,t)\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi(x,t)\,dx, \tag{20}$$

the x coordinate

$$\langle x\rangle=\int\Psi^*(x,t)x\Psi(x,t)\,dx=\int|\Psi(x,t)|^2x\,dx, \tag{21}$$

and the energy

$$\begin{aligned}\langle E\rangle &= \int\Psi^*(x,t)H\Psi(x,t)\,dx \qquad (22)\\ &= \int\left[\frac{\hbar^2}{2m}\left|\frac{\partial\Psi(x,t)}{\partial x}\right|^2+|\Psi(x,t)|^2U(x)\right]dx.\end{aligned}$$

Partial integration has been used to get the last form of (22).

For the sinusoidal wave (11) the momentum $p=\hbar k$ is precisely known but the x coordinate of the particle is completely unknown (since $|\psi(x)|^2$ is a constant). By making a superposition of sinusoidal waves with different k's, it is possible to form a localized *wave packet*, but then the momentum is no more well defined (exercise). We skip the precise definitions and the proof, but state that the uncertainties of momentum Δp and location Δx satisfy the inequality

$$\Delta x\,\Delta p\geq\frac{\hbar}{2} \tag{23}$$

known as the *uncertainty relation*.

Above we compared the Schrödinger equation (1) with Newton's equation of motion (7). The relation between the two is clarified by *Ehrenfest theorem*. It states that the expectation values of the location x and the force $F=-dU/dx$ obey Newton's equation of motion

$$m\frac{d^2\langle x\rangle}{dt^2}=-\langle\frac{dU}{dx}\rangle. \tag{24}$$

The proof of this relation is an exercise in integration by parts using the Schrödinger equation and its complex conjugate.

In the following we refresh our memory of some applications of the Schrödinger equation.

1.2 Free particle

A particle is called free if no force acts on it, $F = 0$. This is the case when the potential $U(x)$ is constant [see (7)]. For simplicity we assume U to vanish ($U(x) \equiv 0$). A solution of the Schrödinger equation (9) is the sinusoidal wave (11). The momentum and energy eigenvalues are

$$p = \hbar k, \quad E = \frac{\hbar^2 k^2}{2m}. \tag{25}$$

Here p can have an arbitrary value (positive or negative) and E can have an arbitrary positive value.

1.3 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

$$\psi(0) = \psi(L) = 0. \tag{26}$$

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

$$\begin{aligned} \psi(x) &= Ce^{ikx} + De^{-ikx} \\ &= A\cos(kx) + B\sin(kx). \end{aligned} \tag{27}$$

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.

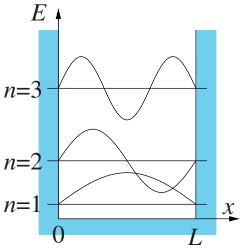
The boundary condition $\psi(0) = 0$ gives $A = 0$. The boundary condition $\psi(L) = 0$ gives $B\sin(kL) = 0$. Assuming $B \neq 0$ we should have $kL = \pi n$ with integer n. The nontrivial solutions have

$$k = \frac{\pi n}{L}, \quad n = 1, 2, 3, \ldots \tag{28}$$

The corresponding energies are obtained from (25), giving

$$E = \frac{\pi^2 \hbar^2 n^2}{2mL^2}, \quad n = 1, 2, 3, \ldots \tag{29}$$

We see that energy values are *discrete*, labelled by integer n.



The figure illustrates the three lowest energies and their wave functions. Although wave numbers (28) are constant times integer, the energies/frequencies (29) are not equidistant. Another noteworthy point is *zero point energy.* That is, the energy of the lowest energy state $n = 1$ is not zero.

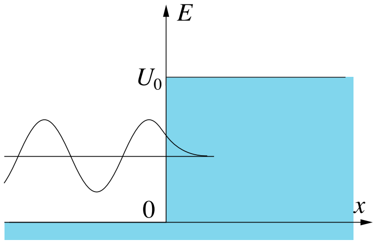
In order to normalize the solutions we calculate I (4). We find $I = B^2 L/2$. Thus the normalized eigenstates are

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\frac{\pi n x}{L}, \quad n = 1, 2, 3, \ldots \tag{30}$$

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

1.4 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.



Consider first what happens according to classical mechanics. If the particle's 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 (9) 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 (27). On the right the energy of the solution is less than the potential energy, $E - U_0 < 0$,

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}(x) - (U_0 - E)\psi(x) = 0. \tag{31}$$

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

$$\psi(x) = F \exp\left(-\frac{\sqrt{2m(U_0 - E)}x}{\hbar}\right). \tag{32}$$

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, see the figure above. 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 $E > U_0$ 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 (32) 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.

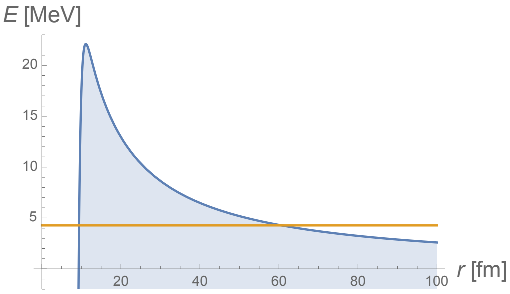
Let us estimate the tunneling probability. The amplitude of the transmitted wave relative to the incoming one is given by the ratio of (32) at $x = L$ and $x = 0$. Since the probability is proportional to the amplitude squared, we get the estimate of tunneling probability

$$T \approx \exp\left(-\frac{2\sqrt{2m(U_0 - E)}L}{\hbar}\right). \tag{33}$$

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 (33) 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 ^{234}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 to be drawn in the lecture)

1.5 Harmonic oscillator

(YF 40.5) Harmonic potential $U(x) = \frac{1}{2}Kx^2$, where K is the force constant. The classical oscillation angular frequency is $\omega_0 = \sqrt{K/m}$. Strangely enough, the quadratic dependence of kinetic energy on k and the form of the potential cancel each effects so that the energy levels are equally spaced,

$$E_n = \hbar\omega_0 \left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \ldots \tag{34}$$

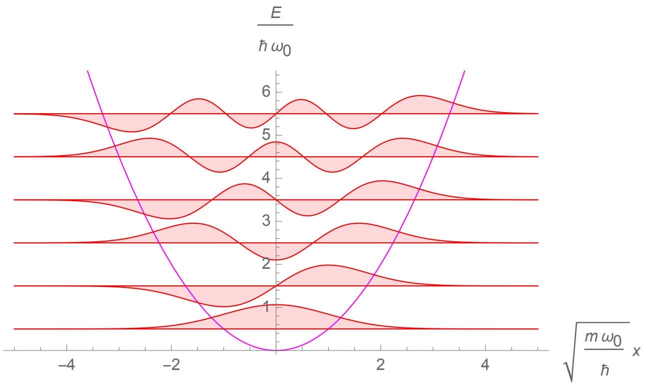


Figure: the parabolic potential energy, the 6 lowest energy eigenvalues (horizontal lines) and the corresponding wave functions.

We only give the ground state solution

$$\psi(x) = \left(\frac{m\omega_0}{\pi\hbar}\right)^{\frac{1}{4}} \exp\left(-\frac{m\omega_0}{2\hbar}x^2\right), \tag{35}$$

which corresponds to the minimal uncertainty in (23).

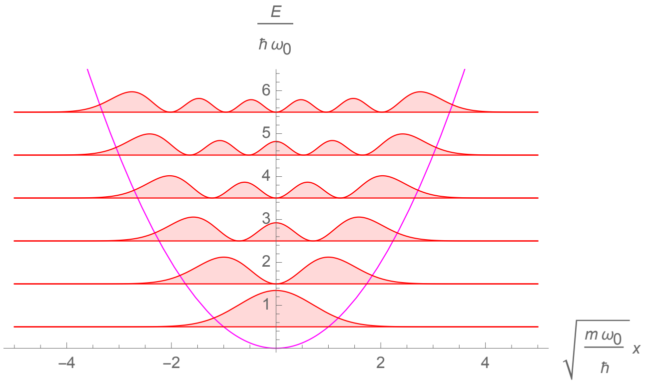


Figure: The probability densities $|\psi(x)|^2$ for the same states in the previous figure.

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.

1.6 Transitions

We have seen above that quantum mechanics in many cases leads to discrete energy levels. This is in contrast to classical mechanics, where the energy spectrum is continuous, i.e. all energy values above a minimum are allowed. The discreteness of energy has important effect on the emission and absorption of electromagnetic radiation. Let us consider an electron in some potential well, although the same could be applied to other particles as well.

Electromagnetic radiation consist of photons. They are quanta of energy $E = h\nu$, where ν is the frequency of the radiation. The *emission* and *absorption* of radiation is possible only at discrete frequencies, as follows. Suppose the electron is initially in an excited state n with energy E_n. The transition from level n to n' of lower energy ($E_{n'} < E_n$) takes place by emitting a photon of frequency ν so that the energy is conserved,

$$h\nu = E_n - E_{n'}. \tag{36}$$

A system in its ground state (lowest energy state) cannot emit radiation since there are no lower levels for the electron. Correspondingly, absorption can take place by an electron transiting from an energy level to a higher level so that the energy difference between them matches the photon energy.

As an example, consider hydrogen, the lightest of all substances. For hot hydrogen gas several spectral lines were found experimentally in the late 1800's. The wave lengths of these could be fitted by the expression

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \tag{37}$$

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). Our plan in the following is to understand these based on quantum mechanics.

1.7 Schrödinger equation in 3 dimension

(YF 41.1) We have written the Schrödinger equation (1) for one spatial dimension (1D). We now consider its generalization to 3 dimensions (3D). The wave function depends on the location vector $\boldsymbol{r}$ and on time, $\Psi(\boldsymbol{r}, t)$. The Schrödinger equation takes the form

$$i\hbar\frac{\partial\Psi}{\partial t}(\boldsymbol{r}, t) = -\frac{\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r}, t) + U(\boldsymbol{r})\Psi(\boldsymbol{r}, t). \tag{38}$$

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

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}, \tag{39}$$

$$\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{40}$$

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

$$\begin{aligned} i\hbar\frac{\partial\Psi}{\partial t} &= -\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2}\right) \\ &+ U(x, y, z)\Psi. \end{aligned} \tag{41}$$

The normalization condition (5) in 3D is

$$\int |\Psi(\boldsymbol{r},t)|^2 dV = 1, \tag{42}$$

where the integration is over 3D space.

The sinusoidal wave (11) has a straightforward generalization to 3D,

$$\Psi(\boldsymbol{r},t) = Ae^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}. \tag{43}$$

This is called *plane wave*, and $\boldsymbol{k}$ is the the *wave vector*

$$\boldsymbol{k} = k\hat{\boldsymbol{n}} = \frac{2\pi}{\lambda}\hat{\boldsymbol{n}}, \tag{44}$$

where $\hat{\boldsymbol{n}}$ is the unit vector in the propagation direction. 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

$$\Psi(\boldsymbol{r},t) = Ae^{ik_x x}e^{ik_y y}e^{ik_z z}e^{-i\omega t}. \tag{45}$$

[Exercise: verify that (43) is a solution of (38) 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 $\boldsymbol{r}$.

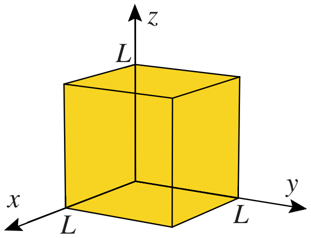
Similarly as in 1D, we can look at an energy eigenstate $\Psi(\boldsymbol{r},t) = e^{-i\omega t}\psi(\boldsymbol{r})$, and write the time-independent Schrödinger equation. In cartesian coordinates it is

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) + U(x,y,z)\psi = E\psi, \tag{46}$$

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

1.8 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 (46) is called *separation of variables*. We assume a solution that is a product of functions each depending on a single variable,

$$\psi(x,y,z) = X(x)Y(y)Z(z). \tag{47}$$

We substitute this in (46). Dividing the equation by ψ we get

$$\begin{aligned}\left(-\frac{\hbar^2}{2m}\frac{1}{X(x)}\frac{d^2X(x)}{dx^2}\right) + \left(-\frac{\hbar^2}{2m}\frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2}\right)& \\ + \left(-\frac{\hbar^2}{2m}\frac{1}{Z(z)}\frac{d^2Z(z)}{dz^2}\right) &= E.\end{aligned} \tag{48}$$

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 equal to a constant. We denote it by E_x. Similarly for y and z. Thus the equation separates into three ordinary differential equations

$$\begin{aligned}-\frac{\hbar^2}{2m}\frac{d^2X(x)}{dx^2} &= E_X X(x)\\ -\frac{\hbar^2}{2m}\frac{d^2Y(y)}{dy^2} &= E_Y Y(y)\\ -\frac{\hbar^2}{2m}\frac{d^2Z(z)}{dz^2} &= E_Z Z(z)\end{aligned} \tag{49}$$

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 (26). We can read the wave function from (30) and energy from (29). 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

$$\psi(x,y,z) = \left(\frac{2}{L}\right)^{3/2}\sin\frac{\pi n_1 x}{L}\sin\frac{\pi n_2 y}{L}\sin\frac{\pi n_3 z}{L}, \tag{50}$$

and the energy

$$E = \frac{\pi^2\hbar^2}{2mL^2}(n_1^2 + n_2^2 + n_3^2). \tag{51}$$

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 $E = 12E_0$.)

1.9 Hydrogen atom

(YF 41.3) The hydrogen atom consists of a single electron around a nucleus. The nucleus with proton charge $+e$ attracts the electron of charge $-e$ by the Coulomb force $F = (1/4\pi\epsilon_0)e^2/r^2$, where r is the distance of the electron from the nucleus. The corresponding potential energy is

$$U(\boldsymbol{r}) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}. \tag{52}$$

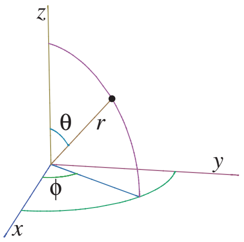
[Recall that the force is the gradient of U, $\boldsymbol{F} = -\boldsymbol{\nabla}U$.]

Our program is now to find the electron states in the hydrogen atom using the Schrödinger equation (38). For the stationary states, we use the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\boldsymbol{r}) + U(\boldsymbol{r})\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}). \tag{53}$$

The Coulomb potential (52) depends only on the radial distance $r = |\boldsymbol{r}|$. Because of spherical symmetry, the solutions of the Schrödinger equation (53) are most conveniently obtained using spherical coordinates. Spherical coordinates (r, θ, ϕ) are defined by

$$\boldsymbol{r} = r(\hat{\boldsymbol{x}}\cos\phi + \hat{\boldsymbol{y}}\sin\phi)\sin\theta + \hat{\boldsymbol{z}}r\cos\theta. \tag{54}$$



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 (53) contains the Laplace operator ∇^2 (40). In spherical coordinates it is given by

$$\begin{aligned}\nabla^2\psi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) \\ &\quad + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}.\end{aligned} \tag{55}$$

This formula is justified in Appendix A.

Let us first demonstrate one solution of the problem. Substitute the ansatz $\psi(r) = e^{\alpha r}$ into (53). 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) given in (69), and α is simply related to the *Bohr radius*

$$a = \frac{4\pi\epsilon_0\hbar^2}{e^2 m} = 0.529 \times 10^{-10}\ \text{m}. \tag{56}$$

(Exercise)

Historically, the "atomic radius" (56) was found by Niels Bohr in 1913. He considered a classical model of circular electron orbits around the nucleus with the additional assumption that the angular momentum $L = \hbar n$ and $n = 1, 2, \ldots$. This led to the correct energy spectrum (69), but his assumption about L can be seen to be in disagreement with (72) and (70), which are obtained by solving the Schrödinger equation.

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

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi). \tag{57}$$

We substitute this in (53). Dividing the equation by ψ we get

$$\begin{aligned}&\frac{1}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)\right] \\ &+ \frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right] \\ &+ \frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r} = E.\end{aligned} \tag{58}$$

We would like to argue, as in the case of (48), 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

$$\begin{aligned}&\frac{r^2}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R - ER\right] \\ &+ \frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right] \\ &+ \frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] = 0.\end{aligned} \tag{59}$$

We define a new quantity L^2 so that $L^2/2m$ is equal to the two last terms in (59),

$$\begin{aligned}&\frac{1}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right] \\ &+ \frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin^2\theta}\frac{d^2\Phi}{d\phi^2}\right] = \frac{L^2}{2m}.\end{aligned} \tag{60}$$

We see that L^2 cannot depend on r since there is no r dependence in (60). Using (60) we can write (59) as

$$\frac{r^2}{R}\left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R - ER\right] = -\frac{L^2}{2m}. \tag{61}$$

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

$$\begin{aligned}&\frac{\sin^2\theta}{\Theta}\left[-\frac{\hbar^2}{2m}\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) - \frac{L^2}{2m}\Theta\right] \\ &+ \frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{d^2\Phi}{d\phi^2}\right] = 0.\end{aligned} \tag{62}$$

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 $L_z^2/2m$,

$$\frac{1}{\Phi}\left[-\frac{\hbar^2}{2m}\frac{d^2\Phi}{d\phi^2}\right]=\frac{L_z^2}{2m}. \tag{63}$$

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

$$\begin{aligned}-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{L^2}{2mr^2}R-\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}R=ER\\ -\frac{\hbar^2}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{L_z^2}{\sin^2\theta}\Theta=L^2\Theta\\ -\hbar^2\frac{d^2\Phi}{d\phi^2}=L_z^2\Phi.\end{aligned} \tag{64}$$

In solving (64) 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 (64) as an exponential $e^{ik\phi}$. The boundary condition limits $2\pi k$ to be an integer multiple of 2π. The quantum number associated with ϕ is called the *orbital magnetic quantum number* and marked by m_l. Thus

$$\Phi(\phi)=\frac{1}{\sqrt{2\pi}}e^{im_l\phi},\quad m_l=0,\pm1,\pm2,\ldots, \tag{65}$$

where the prefactor comes from normalization. Substituting this in the last one in (64) shows that it is satisfied when

$$L_z=\hbar m_l. \tag{66}$$

We can identify L_z as the z component of the angular momentum.

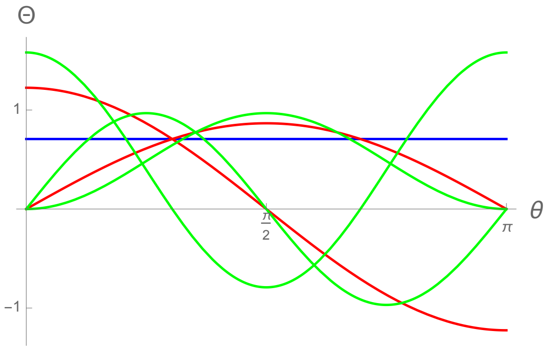
The next step is to solve the second equation in (64). The boundary condition for $\Theta(\theta)$ is that it should be finite in the whole interval $0\le\theta\le\pi$. We will give the general result without proof. The quantity L is the angular momentum. Its values are given by the *orbital angular momentum quantum number* l so that

$$L=\hbar\sqrt{l(l+1)},\quad l=0,1,2,\ldots. \tag{67}$$

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

$$\begin{aligned}\Theta&=\frac{1}{\sqrt{2}},\quad l=0,m_l=0\\ \Theta&=\sqrt{\frac{3}{2}}\cos\theta,\quad l=1,m_l=0\\ \Theta&=\sqrt{\frac{3}{4}}\sin\theta,\quad l=1,m_l=\pm1\\ \Theta&=\sqrt{\frac{5}{8}}(3\cos^2\theta-1),\quad l=2,m_l=0\\ \Theta&=\sqrt{\frac{15}{4}}\cos\theta\sin\theta,\quad l=2,m_l=\pm1\\ \Theta&=\sqrt{\frac{15}{16}}\sin^2\theta,\quad l=2,m_l=\pm2.\end{aligned} \tag{68}$$

Your are encouraged to check that these indeed satisfy the second of (64) with the eigenvalue (67).



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 (64). The boundary condition is that the solution should vanish with $r\to\infty$. Again, we give the result without proof. The energy eigenvalues are

$$E=-\frac{me^4}{2(4\pi\epsilon_0)^2\hbar^2n^2}=-\frac{13.60\text{ eV}}{n^2},\quad n=1,2,3,\ldots \tag{69}$$

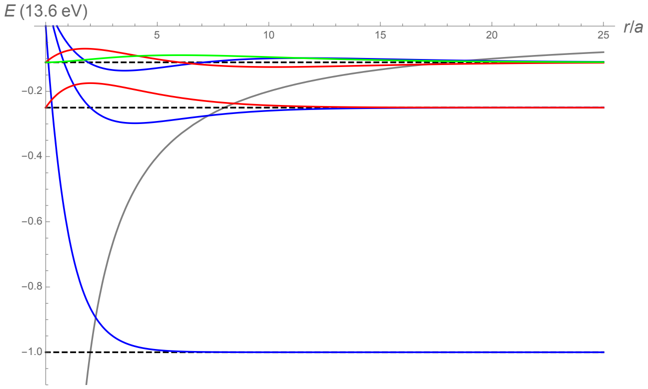
The integer n is called the *principal quantum number*. For a given value of n, the orbital angular quantum number l can take values from 0 to $n-1$,

$$l=0,1,\ldots,n-1. \tag{70}$$

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

$$\begin{aligned}R&=\frac{2}{a^{3/2}}e^{-r/a},\quad n=1,l=0\\ R&=\frac{1}{\sqrt{2}a^{3/2}}\left(1-\frac{r}{2a}\right)e^{-r/2a},\quad n=2,l=0\\ R&=\frac{1}{2\sqrt{6}a^{3/2}}\frac{r}{a}e^{-r/2a},\quad n=2,l=1,\end{aligned} \tag{71}$$

where a is the Bohr radius (56).



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.

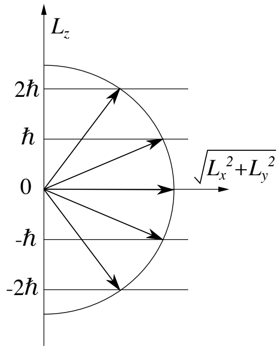
1.10 Angular momentum

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

$$\begin{aligned} L &= \hbar\sqrt{l(l+1)}, && l = 0, 1, 2, \ldots \\ L_z &= \hbar m_l, && m_l = 0, \pm 1, \ldots, \pm l. \end{aligned} \tag{72}$$

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.



This figure illustrates the angular momentum in case of $l = 2$

We now can ask whether we also could 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 (65) and (68), but with the angle ϕ denoting the rotation angle around the x axis and θ denoting the angle measured from 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.

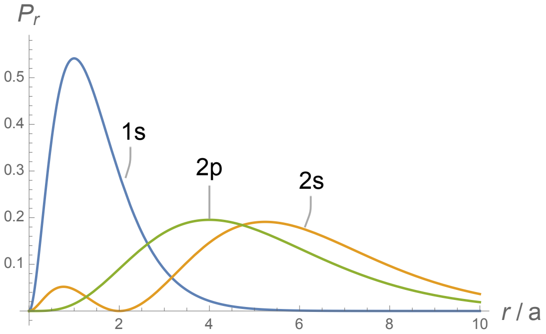
In the figure we discussed, we can only specify the L, L_z and $\sqrt{L_x^2 + L_y^2} = \sqrt{L^2 - L_z^2}$, but not L_x and L_y. However, this problem is alleviated when l is large. In this case the states with different m_l are dense, in comparison to l. Then making a linear combination of the states of different m_l, one can form a state with a 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 note that this discussion of angular momentum uses only the spherical symmetry, and is valid for any potential function $U(r)$, not just (52).

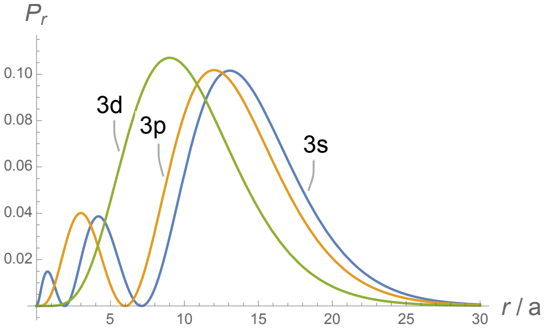
The states having $l = 0, 1, 2$ and 3 are commonly called s, p, d and f states.

1.11 Hydrogen probability distributions

The probability density of electron is given by $P = |\psi|^2$ (6). In 3 D these are not trivial to show because dependence on more than one variable. The θ dependence can be obtained by squaring (68). 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 $n = 1$ (1s) and $n = 2$ (2s and 2p).



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 a semiclassical Bohr model for states $n = 1, 2$ and 3 would be the Bohr radius a (56), $4a$ and $9a$. Compare these to the distributions in the figures. In the ground state (1s), 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 (65) as eigenstates of the operator L_z. It is defined by

$$(L_z)_{\text{op}} = \frac{\hbar}{i}\frac{\partial}{\partial\phi} \tag{73}$$

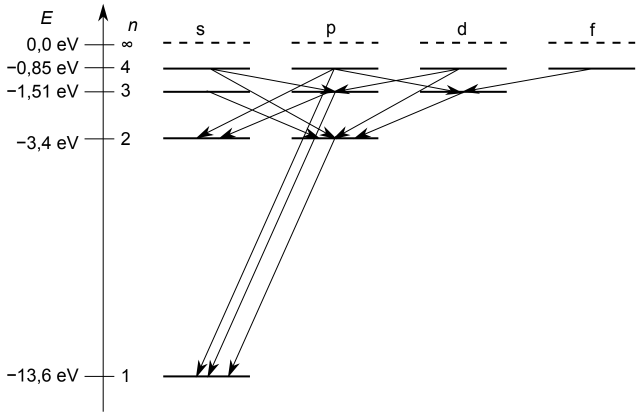
in close resemblance to the momentum operator (10). The L_z eigenstates (65) have constant $|\Phi|^2 = 1/2\pi$. For some purposes, in particular for application in chemistry, it is more convenient to choose linear combinations of them. For example, in the case of $l = 1$ we define

$$\begin{aligned}\Phi_{p_x}(\phi) &= \frac{1}{\sqrt{2}}\Big(\frac{e^{i\phi}}{\sqrt{2\pi}} + \frac{e^{-i\phi}}{\sqrt{2\pi}}\Big) = \frac{1}{\sqrt{\pi}}\cos\phi,\\ \Phi_{p_y}(\phi) &= \frac{1}{\sqrt{2}i}\Big(\frac{e^{i\phi}}{\sqrt{2\pi}} - \frac{e^{-i\phi}}{\sqrt{2\pi}}\Big) = \frac{1}{\sqrt{\pi}}\sin\phi.\end{aligned} \tag{74}$$

These are largest in the x and y directions, respectively, and are called p_x and p_y orbitals. Except for different orientation, they are identical to the p_z orbital, which corresponds to $m_l = 0$, see (68).

1.12 Hydrogen energy levels

(YF 41.4) We see that the energy (69) 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 (69) 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 (36). This just gives explanation to the observed wave lengths (37) of emission and absorption of hydrogen. 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 the 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

$$E_m = -\boldsymbol{B}\cdot\boldsymbol{\mu}. \tag{75}$$

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 $I = -e\nu$, where the frequency $\nu = v/2\pi r$. For the magnetic moment we get

$$\mu = -\frac{1}{2}evr = -\frac{e}{2m}L, \tag{76}$$

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

$$\boldsymbol{\mu} = -\frac{e}{2m}\boldsymbol{L}. \tag{77}$$

This is a simple result since $\boldsymbol{\mu}$ and $\boldsymbol{L}$ are related by natural constants, the elementary charge and electron mass. We claim that the result (77) 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 (77) and (72), we write the magnetic energy (75)

$$E_m = m_l \mu_{\rm B} B, \tag{78}$$

where we have defined *Bohr magneton* $\mu_{\rm B} = e\hbar/2m = 9.274 \times 10^{-24}$ J/T $= 5.8 \times 10^{-5}$ eV/T. The magnetic energy (78) should be added to the hydrogen energy levels energies (69). 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 that 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 (78). As a result the, the slitting of the (n, l) levels into sublevels could be observed.

1.13 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 (72):

$$\begin{aligned} S &= \hbar\sqrt{s(s+1)} = \hbar\sqrt{\tfrac{3}{4}} \\ S_z &= \hbar m_s, \qquad m_s = \pm\tfrac{1}{2}, \end{aligned} \tag{79}$$

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.

$$\mu_z = -g\frac{e}{2m}S_z, \tag{80}$$

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, $E_{m,s} = m_s g \mu_{\rm B} B$. 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 (80). The energy is of the form

$$E_{SO} = a\boldsymbol{L} \cdot \boldsymbol{S} \tag{81}$$

with some constant a.

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

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}. \tag{82}$$

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

$$J = \hbar\sqrt{j(j+1)}, \quad j = |l \pm \tfrac{1}{2}|. \tag{83}$$

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 (81) is thus different for the two states. As a consequence, the degeneracy of the levels of different l but same n in (69) is lifted.

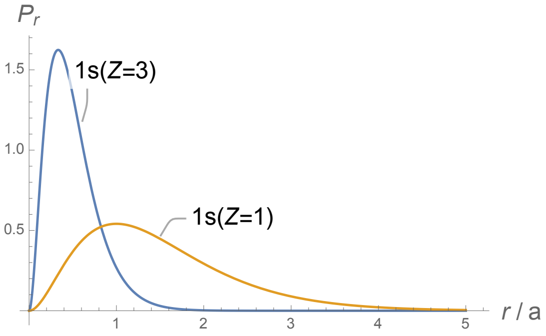
We stop here by noting some effects that still were neglected. The mass in (69) is the *reduced mass*,

$$\frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2} \quad \Leftrightarrow \quad m = \frac{m_1 m_2}{m_1 + m_2}, \tag{84}$$

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 (Appendix B). 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.

1.14 Hydrogen-like atoms

Helium nucleus has two protons, the atomic number $Z = 2$, and lithium three 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 (52) becomes multiplied by Z, the energy (69) multiplied by Z^2 and the radial scale, characterized by the Bohr radius (56), becomes multiplied by $1/Z$. We see that with increasing Z the binding energy grows rapidly and the electron is bound closer to the nucleus.



The figure illustrates that the probability density of 1s state is compressed towards the nucleus by factor 1/3 in Li^{2+} compared to H.

2. Many-particle systems

2.1 Many-body problem

(YF 41.6, S 5.1) 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 (38) 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,

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t). \tag{85}$$

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,

$$-\frac{\hbar^2}{2m}\nabla_1^2\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t) - \frac{\hbar^2}{2m}\nabla_2^2\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, t), \tag{86}$$

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,

$$U_{\text{central}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\frac{Ze^2}{4\pi\epsilon_0}\frac{1}{r_1} - \frac{Ze^2}{4\pi\epsilon_0}\frac{1}{r_2}, \tag{87}$$

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

$$U_{\text{el-el}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{e^2}{4\pi\epsilon_0}\frac{1}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|}. \tag{88}$$

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

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \psi_1(\boldsymbol{r}_1)\psi_2(\boldsymbol{r}_2). \tag{89}$$

This would work fine for other terms, but there is a problem in electron-electron interaction (88): where is electron 1, affects the potential of electron 2. Thus their motion is correlated, and cannot be represented in the form (89). 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 (38), but with an effective potential $U = U_{\text{eff}}$, that includes the averaged effect of all other particles.

Let us emphasize the important step we have to make here. Most of the problems to be studied in the following are so complicated that solving them from first principles is impossible. Instead, we aim to make simple models to learn about the physical system we are studying. If the model is insufficient, we can try a more accurate model. This hopefully leads to better understanding, but simultaneously the work we have to do increases as well.

2.2 Atomic structure

(YF 41.6, S 5.2-3) 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, $U_{\text{eff}}(\boldsymbol{r}) = U_{\text{eff}}(r)$.

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 (72). What still remains to be done is solve the radial wave functions from the radial equation

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{\hbar^2 l(l+1)}{2mr^2}R+U_{\text{eff}}(r)R = ER. \quad (90)$$

As long as $U_{\text{eff}}(r)$ is not known, 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 (90). This degeneracy can be traced to the $1/r$ form of potential (52). In many-electron system we do not expect the effective potential $U_{\text{eff}}(r)$ 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 (72). The radial wave functions from (90) give the principal quantum number n. In addition we have the magnetic spin quantum number m_s (79). 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*.

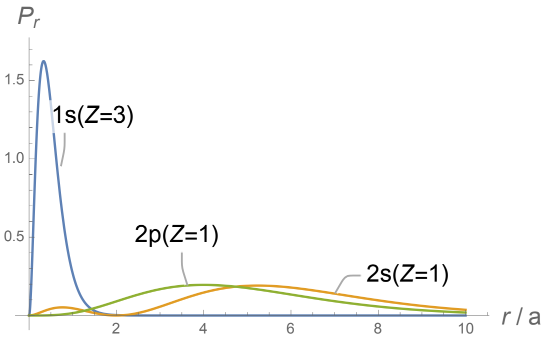
The ground state of hydrogen has one electron in state $n = 1$. It is called 1s *configuration*. In helium the additional electron can also go to $n = 1$ orbital supposing the m_s numbers are different. That is, one electron has $m_s = +\frac{1}{2}$ and the other $m_s = -\frac{1}{2}$. He configuration 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 $[\text{Ne}]3s^2 3p^6$. 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 $[\text{Ar}]4s^2$. Only after that are the 3d orbitals filled, with maximum 10 electrons, up to Zn, $Z = 30$, $[\text{Ar}]3d^{10}4s^2$.

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.



The figure gives the probability densities of Li^{2+} and H wave functions that approximate the ones in Li. Two electrons occupy the 1s state close to the nucleus so that the effective nuclear charge for the third electron is $1e$. The 2s state has lower energy than the 2p state because it extends more in the region close to the nucleus ($r \sim a$) where the screening of the nuclear charge $3e$ is weaker.

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 is consistent with the exclusion principle. This means that in $2\mathrm{p}^2$ and $2\mathrm{p}^3$ configurations the total spin is $S = \hbar$ and $S = \frac{3}{2}\hbar$, respectively. In $2\mathrm{p}^4$ the Pauli principle does not allow more than $S = \hbar$, and in $2\mathrm{p}^5$ only $S = \frac{1}{2}\hbar$. (Exercise: why?)

(YF 41.7) The deep levels can be seen as characteristic X-ray radiation, see figure below. 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.

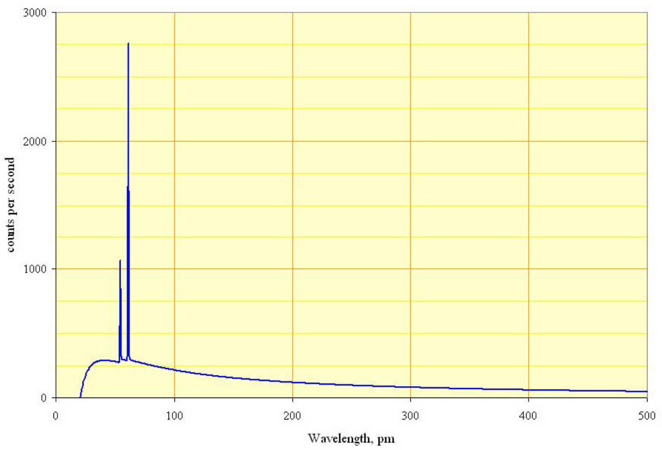


Figure: Spectrum of the X-rays emitted by an X-ray tube with a rhodium target, operated at 60 kV. The smooth, continuous curve is due to bremsstrahlung, and the spikes are characteristic K lines for rhodium atoms. (source Wikipedia)

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)

2.3 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 (85).

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

$$\psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2). \tag{91}$$

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

$$\boxed{\psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2) = \pm\psi(\boldsymbol{r}_2, \sigma_2, \boldsymbol{r}_1, \sigma_1).} \tag{92}$$

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 (92) 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 (92), the two particle wave functions should be

$$\psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2) = \frac{1}{\sqrt{2}} \left[\phi_a(\boldsymbol{r}_1, \sigma_1)\phi_b(\boldsymbol{r}_2, \sigma_2) \pm \phi_b(\boldsymbol{r}_1, \sigma_1)\phi_a(\boldsymbol{r}_2, \sigma_2)\right]. \tag{93}$$

[Exercise: check that (93) satisfies the condition (92).] Now test the case that the quantum number sets a and b are the same, $a = b$. We get from (93)

$$\psi(\boldsymbol{r}_1, \sigma_1, \boldsymbol{r}_2, \sigma_2) = \begin{cases} \sqrt{2}\phi_a(\boldsymbol{r}_1, \sigma_1)\phi_a(\boldsymbol{r}_2, \sigma_2) & \text{for bosons,} \\ 0 & \text{for fermions.} \end{cases} \tag{94}$$

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

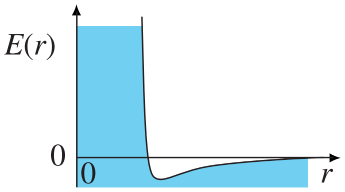
We mentions a general *spin-statistics law.* It states that all particles with half-integer spin are fermions and all particles with integer spin are bosons. Protons and neutrons are fermions, photons are bosons. (The polarization of the electromagnetic field of the photon can be interpreted as spin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

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.

2.4 Chemical bond

(YF 42.1, S 6.1-5) An important application of quantum mechanics is to understand bonding between atoms. Unfortunately, this is also very difficult topic. Here we briefly mention the main types of bonding.

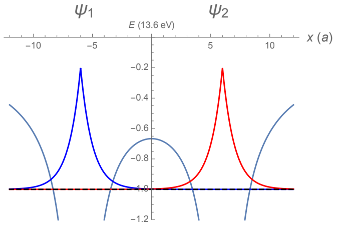
At short distances the force between two atoms is repulsive. To large extend this is caused by the exclusion principle, that prevents two electrons to be in the same state. Also the Coulomb repulsion between electrons is essential. At larger distances the force is often attractive.



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 charged ions.

Covalent bond. There is attractive force because the atoms share part of their electrons. As a simple case, consider two protons at some distance and a single electron. The figure illustrates the ground state wave functions localized to the two nuclei, ψ_1 and ψ_2.



When the wave functions start to overlap, appropriate wave functions to describe stationary states in the molecule are

$$\psi_{\text{even}} \approx \psi_1 + \psi_2, \quad \psi_{\text{odd}} \approx \psi_1 - \psi_2. \tag{95}$$

The probability density $|\psi|^2$ in these two states differs especially in the region between the atoms: it has a zero in ψ_{odd} but stays finite in ψ_{even}.

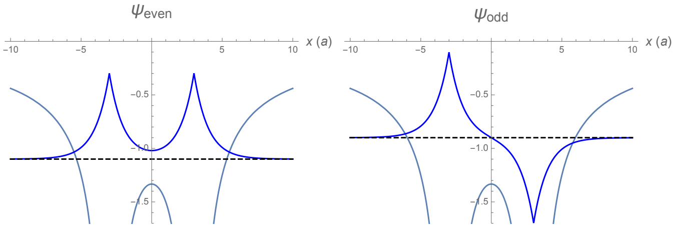


Figure: The energies are only qualitatively correct. The atoms are still far apart compared to the bond length 0.74 nm $= 1.4a$ of H_2.

The expectation values (22) of both the kinetic and the potential energy are lower in ψ_{even} than in ψ_{odd}. The kinetic energy is lower because the gradient of the wave function is lower. The potential energy is lower because the probability density is larger in the region between the nuclei, where the potential is lower (compared to other directions from a nucleus at the same distance).

Let us now consider adding a second electron, that is, we consider two hydrogen atoms. Now both electrons can go to the lower energy molecular orbital ψ_{even} with opposite spins. Because the energy is lower than in two separated hydrogen atoms, this leads to boding of the atoms. For a more complete study of the problem, one should also include the electrostatic repulsion between the nuclei, and the electrostatic repulsion between the electrons. While the former is easy, the latter is difficult to calculate (as in the case of He atom above).

The reduction of kinetic energy is similar as in potential well, (29) or (51), 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 other bond 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. A bond between non-equal atoms has partly covalent and partly ionic nature.

Metallic bond. A large amount of atoms share some of their electrons, which can move almost like free particles through the material. 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.

2.5 Transitions in molecules

(YF 42.2) Similarly as in an atom, there can be transitions between electronic levels in a molecule. An additional feature in a molecule is that the nuclei can have relative motion. This results in transitions that we will study now. For simplicity we first consider only a diatomic molecule, consisting of two nuclei (and of electrons).

The relative motion of the nuclei in a diatomic molecule can be studies similarly as for two particles (a nucleus and an electron) in a hydrogen atom. There are two essential differences. 1) Whereas the electron is much lighter than the proton, the atoms in a molecule can be comparable in mass. Therefore it is essential to use the reduced mass (84) that takes into account that both particles are moving relative to their center of mass. 2) The potential is of the form depicted in the first figure of Sec. 2.4. Let r_0 denote the distance corresponding to the minimum of the potential. For small oscillations around this minimum, we approximate the potential by harmonic oscillator with force constant K. The change of r compared to r_0 is assumed to be small. This implies that we can treat the *rotation* of the molecule separately from the *vibration*, the oscillation of r.

Consider first rotation. The angular momentum is given by the quantum numbers l and m_l as given in (72). The rotational kinetic energy can be read from the first of equations (64) by comparing the terms containing L^2 and E. Because $r \approx r_0$, we get

$$E_{\rm rot} = \frac{\hbar^2}{2I} l(l+1), \quad l = 0, 1, 2, \ldots \tag{96}$$

where the moment of inertia $I = mr_0^2$. The energies are thus 1, 2, 6, 12 … times $\hbar^2/2I$. For carbon monoxide (CO) $\hbar^2/2I = 0.48$ meV. The rule (36) gives that the transition between the lowest states corresponds to electromagnetic radiation of wave length 2.6 mm, in the microwave region. Similarly to transitions in atoms, there is a selection rule $\Delta l = \pm 1$.

As the radial motion is modeled by a force constant K, the energy levels are as in harmonic oscillator (34),

$$E_{\rm vib} = \hbar\omega_0 (n + \frac{1}{2}), \quad n = 0, 1, 2, \ldots \tag{97}$$

where $\omega_0 = \sqrt{K/m}$. For CO the energy separation of the levels is 0.27 eV. The transition between the levels correspond to electromagnetic radiation with wave length 4.6 μm, which is in the infrared region. Harmonic oscillator has selection rule $\Delta n = \pm 1$, meaning that this is the only wave length in the harmonic approximation.

For comparison, the lowest electronic transition in CO apparently has wave length 154 nm, in the ultraviolet region. Thus we see that the three types of transitions are well separated from each other. There are no transitions in the visible region, which makes CO transparent and colorless.

In addition to the "pure" transitions discussed above there are mixed ones. For example, a transition between vibrational states is associated with simultaneous change of the rotational state. Therefore molecular spectra consist of bands of several closely spaced lines. The same applies to combined electronic and vibrational transitions.

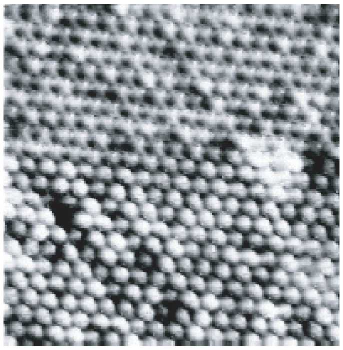
In order for the rotations and vibrations to interact with electromagnetic radiation, the molecule should have a permanent electric dipole moment. Only in this case the electric field drives the two (or more) atoms in a different way. Therefore we used CO as an example above, since in homonuclear molecules like N_2 and O_2, the electromagnetic radiation does not couple to rotation and vibration.

In molecules with more than two atoms (H_2O, CO_2, CH_4, etc.) there are several vibrational modes. As an example, consider CO_2. It is a linear molecule in equilibrium, but its bending couples to electromagnetic radiation at the wave length of 15 μm (in the infrared region). This vibrational transition is crucial to cause of global warming, when the concentration of CO_2 in the atmosphere increases.

3. Solid state: lattice properites

3.1 Crystal structure

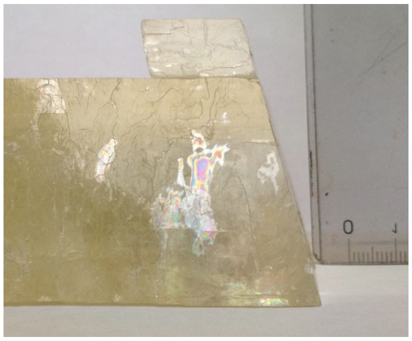
(S 12) Substances are commonly classified as gasses, liquids and solids. Solid materials are formed by boding of a large number of atoms together. The atoms in a solid are often arranged so that they have a periodically repeating pattern. Such solids are called *crystalline*.



A picture obtained by scanning tunneling microscope from the surface of $NbSe_2$. The atoms are arranged in a regular pattern with nearest neighbor distance of 0.35 nm. (http://www.pma.caltech.edu/GSR/condmat.html)

Two quartz crystals (SiO_2).

Many crystalline materials show up in a form where flat surfaces appear at definite angles. Such angles can be understood based on the atomic ordering.



Two calcite crystals ($CaCO_3$) that have the same angles between their faces, Solid bodies are most commonly multicrystalline. This means that they consist of several crystal pieces that have different orientations, joined together. For example, a single crystal could consists of $\sim 10^{18}$ atoms, while the macroscopic body may consist of $\sim 10^{23}$ atoms. These numbers illustrates that a single crystal can be huge compared to atomic scale although it is tiny on a macroscopic scale.

Samples with different sizes of crystals of pyrite (FeS_2).

Generally, the structure of a solid can be complicated. Even though it consists of units having the same atoms, the structure may not have periodic order. One example is glass, which consists of SiO_2 units. Such materials are called *amorphous.*

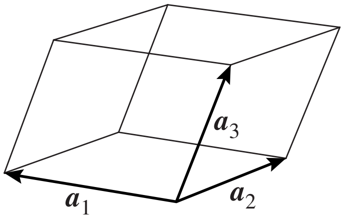
We study the ideal case, where all deviations from periodicity of the crystal can be neglected. The description of the crystal can be divided in two steps.

1) a group of atoms (or any pictures), which forms the periodically repeating object. This is called *basis.*

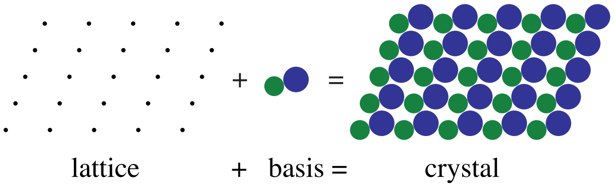
2) a set of points in space, where the basis should be placed (only by moving, not turning) so that the crystal is formed. Such a set of points is represented in form

$$\boldsymbol{r} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3. \tag{98}$$

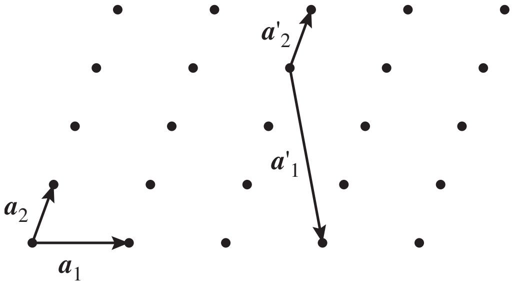
Here n_1, n_2 and n_3 integers $(0, \pm 1, \pm 2, \ldots)$. The vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are called *primitive vectors.* (They have to be linearly independent.) The set of points (98) is called *Bravais lattice*, and its vectors *lattice points.*



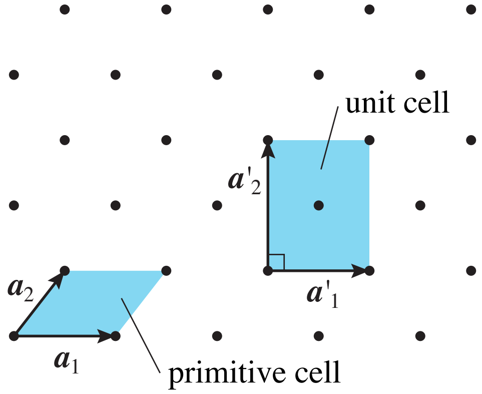
The cell formed by the basis vectors in the figure is called *primitive cell.*



Here an example in two dimensions.



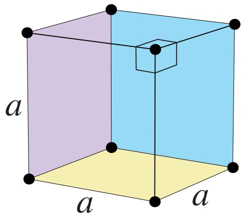
The selection of the primitive vectors is not unique. In the figure we can use $\boldsymbol{a}_1'$ and $\boldsymbol{a}_2'$ as primitive vectors. Using them we get the same lattice points as with $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.



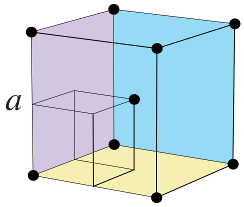
In some symmetric lattices it is convenient to use rectangular basis (although they do not give all lattice points). The cell formed by such vectors is called conventional unit cell. The lengths of the sides of the conventional unit cell are called *lattice constants.*

3.2 Examples of lattices

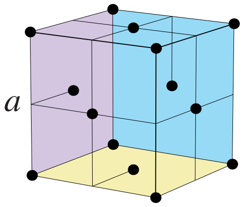
In the figures below, the dots denote the locations of atomic nuclei. The whole crystal is obtained by repeating the unit shown.



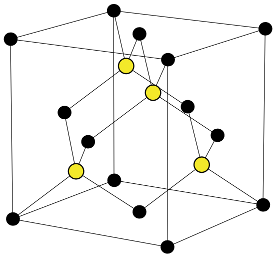
Simple cubic (sc) structure has all sides with equal length and at right angles.



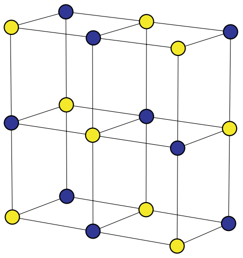
Body centered cubic (bcc) structure has an additional point in the middle of the cube.



Face centered cubic (fcc) has additional points at the centers of the faces compared to simple cubic structure. The structures sc, bcc and fcc are Bravais lattices. (exercise)

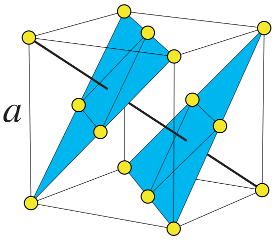


Diamond structure has additional points so that they are in the middle of the tetrahedra formed by four neighbor lattice points of the face centered cubic lattice. This structure appears in carbon C, and in silicon Si, where atoms are of the same type. A related structure appears in zincblende ZnS, where every second atom is S (for example, the yellow dots in the figure above). The structure can be seen as fcc with a basis consisting of two atoms.

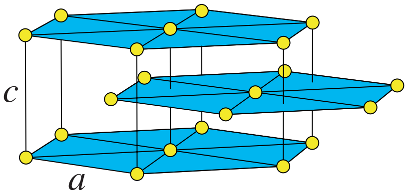


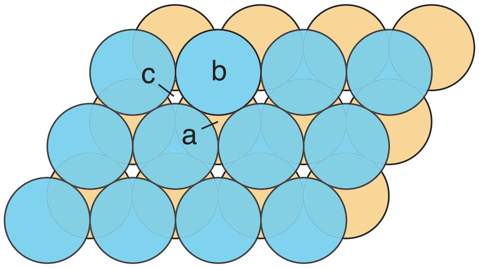
Sodium chloride structure is cubic, where Na and Cl atoms alternate. This can be seen as fcc lattice where the basis consist of a NaCl pair.

Face centered cubic structure is *close packed*: if one packs rigid balls, the maximal density is obtained in this structure. The figure below tries to illustrate how the lattice consist of planes with maximal packing density.



An alternative close packed structure is hexagonal close packing (hcp). It is formed of planes as in fcc, but they are placed differently. In ideal (rigid ball) case $c = \sqrt{8/3}\,a$.



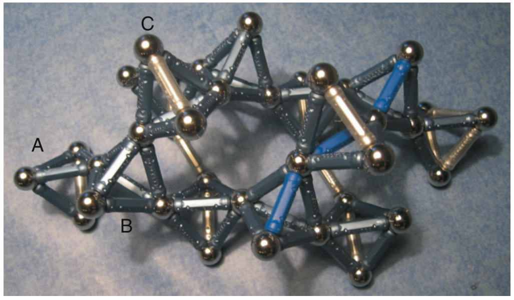


The difference between fcc and hcp structures is presented in the figure, which shows two layers of balls. When the third layer is added on top, it comes in fcc at the locations (c) which were left open in the two lower layers. In hcp they come on the same locations (a) that were used in the lowest layer.

Both close packed structures are common to elemental metals. If atoms were like rigid balls, they would not distinguish between the two structures. Instead, atoms prefer to choose one or the other. For example Al, Ni and Cu prefer fcc and Mg, Zn and Co prefer hcp. In hcp the ratio c/a deviates from the rigid ball value 1.63. A substantial deviation appears in zinc (Zn), which has $c/a = 1.86$.

In addition to close-packed structure, the body centered cubic is common in elements, for example K, Cr, Mn, Fe. The simple cubic is essentially nonexistent in pure elements. The sodium-chloride structure can be understood because of the different size of the atoms: the small sodium ions just fit in the space left between large chloride ions. The diamond structure corresponds to low density packing. It appears in substances which tend to form bonds so that they all are in the same angle relative to each other (109.5°).

Many materials have crystal structure that is more complicated than presented above. The figure below presents the structure of quartz.



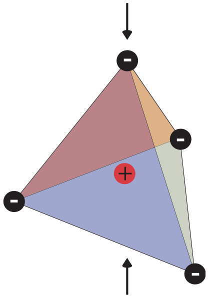
Model of quartz (SiO_2) crystal structure. The balls represent oxygen atoms, the silicon atoms (not shown) are at the center of each O_4 tetrahedron. The basis of the Bravais lattice consists of three tetrahedra (for example A, B and C).

SiO_2 often appears in amorphous form (glass). Looking at the complicated structure of quartz, one maybe can understand that cooling rapidly a molten SiO_2, the atoms have no time to arrange to this structure but remain disordered.

Crystals can be classified based on their symmetry. A crystal is always symmetric in translation by a lattice vector. That is, when the crystal is moved by a lattice vector (98), there is no change. In addition to translation, the crystal can have symmetry in rotation or reflection, for example. It has been show that by using different criteria, the crystals can be classified into 7 crystal systems, 14 Bravais lattices, 32 crystal classes and 230 space groups.

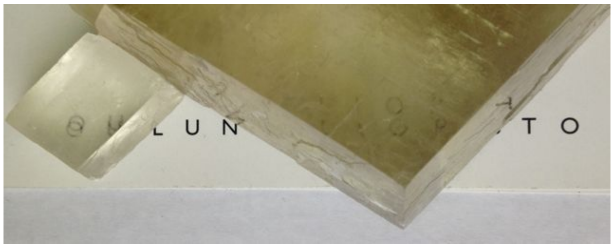
Many properties of solids depend on the crystal symmetry. For example, in an ionic crystal of cubic symmetry the center of positive and negative charge are the same. This implies vanishing electric dipole moment (for example, NaCl). If the symmetry is less (such as rectangular prism) the negative and positive charge can have different centers. Such a material, which has non-zero electric dipole moment, is called *ferroelectric*.

Another case is that the dipole moment vanishes in equilibrium. This is the case in the model figure below, which has negative charges at the corners of a tetrahedron and positive in the center.



When such a material is pressed in the direction of the arrows, the charge centers are no more equal. Such a material is called *piezoelectric*. The effect works also the other way round: when a piezoelectric material is placed in electric field, it changes its shape. Quartz, for example, is piezoelectric. Detailed explanation of the phenomenon is difficult because of the complicated crystal structure.

A phenomenon related to crystal structure is also *double refraction*. In such a material, the speed of light depends on the polarization of light (direction of the electric field in the wave). This divides a light ray in two beams.

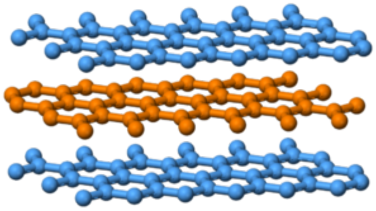


The figure shows double refraction in two calcite ($CaCO_3$) crystals of different thickness.

Let us recall the five types of chemical bonding discussed above. The following table gives some examples of each type (separated by a horizontal line). The second column gives the melting temperature, which reflects the strength of the bonding. The third column gives the crystal structure (when it is simple enough to express).

substance	melting temperature (K)	crystal structure
CsCl	918	sc, CsCl basis
NaCl	1075	fcc, NaCl basis
Si	1683	diamond
C	(4300)	diamond
GaAs	1511	zincblende
SiO_2	1670	
Al_2O_3	2044	
Hg	234.3	
Na	371	bcc
Al	933	fcc
Cu	1356	fcc
Fe	1808	bcc
W	3683	bcc
H_2O	273	
He	-	
Ne	24.5	fcc
Ar	83.9	fcc
H_2	14	
O_2	54.7	

Besides diamond, carbon has an allotropic form *graphite*. It has planes formed by carbon atoms. Within a plane, the atoms are bonded by covalent bonds. The forces between the planes are of van der Waals type, and thus weak. Thus the planes can easily slide past each other. A single plane detached from bulk graphite is called *graphene*.



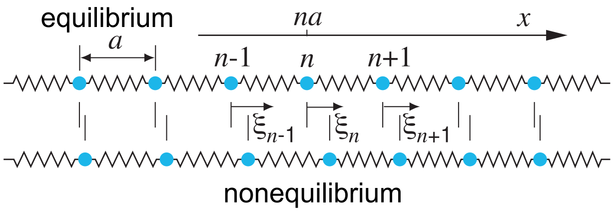
The figure shows model of graphite where the balls represent carbon atoms and the sticks represent covalent bonds. (figure from Wikipedia)

The covalent bond often forms in a particular direction. The covalent bonds are hard, but brittle. That is, they break from sudden strong force. Metallic bond does not depend essentially on the direction. Thus metal atoms can glide past each other without breaking the bond. Therefore metals are malleable.

(S 14) The crystal structure is often studied using x-ray diffraction. In order to save time for other topics, we skip this important topic here.

3.3 Lattice vibrations

(S 9) We studied above the vibrations of atoms in a molecule. Now we consider vibrations of in a crystal. For simplicity, we study first a one-dimensional atom chain. The oscillation amplitude is assumed to be small. Therefore we can use model where the forces are linearly proportional to the deviations in the nearest-neighbor distance. This is equivalent that the atoms were bound to each other by springs.



In equilibrium the atoms are at equal distance a from their neighbors and there is no forces on them. The location of atom n in equilibrium is denoted by $x_n = na$ with integer n. In nonequilibrium, the location of atom n is $x_n = na + \xi_n$, where ξ_n is the deviation of atom n from its equilibrium location. There is a force on atom n from the right-hand-side atom $n+1$ by

$$K(\xi_{n+1} - \xi_n), \tag{99}$$

where K is the force constant and $\xi_{n+1} - \xi_n$ is the stretch of the spiring between atoms n and $n+1$. On the left-hand-side atom $n-1$ exerts the force

$$-K(\xi_n - \xi_{n-1}). \tag{100}$$

The negative sign is because the force is to the negative direction if $\xi_n > \xi_{n-1}$. We sum the forces and form Newton's equation of motion

$$\begin{aligned} M\frac{d^2\xi_n}{dt^2} &= K(\xi_{n+1} - \xi_n) - K(\xi_n - \xi_{n-1}) \\ &= K(\xi_{n+1} - 2\xi_n + \xi_{n-1}), \end{aligned} \tag{101}$$

where M is the mass of the atom.

We try to solve (101) with the ansatz

$$\xi_n(t) = Ae^{i(kna-\omega t)}. \tag{102}$$

Let us study it a bit.

The ansatz (102) is complex because $e^{i\phi} = \cos\phi + i\sin\phi$. The true physical solution should be interpreted at the real part of this,

$$\begin{aligned} \xi_n^{\text{physical}} &= \text{Re}\xi_n \\ &= (\text{Re}A)\cos(kna - \omega t) - (\text{Im}A)\sin(kna - \omega t). \end{aligned} \tag{103}$$

The reason for using complex ansatz is that it simplifies the following calculations, compared to using real-valued sine and cosine functions.

Secondly, we note that $na \approx x_n$ is the x coordinate of atom n (because $\xi_n \ll a$). The ansatz (102) is thus $Ae^{i(kx-\omega t)}$, which is the same form as we used for traveling wave in the Schrödinger equation (11). Recall how the wave number k and the angular frequency ω are related to wave length λ, the frequency ν and to the velocity v of the wave:

$$k = \frac{2\pi}{\lambda}, \ \omega = 2\pi\nu, \ v = \frac{\omega}{k} = \lambda\nu. \tag{104}$$

The velocity of the wave is easiest to deduce by writing the ansatz (102) in the form $f(x - vt)$.

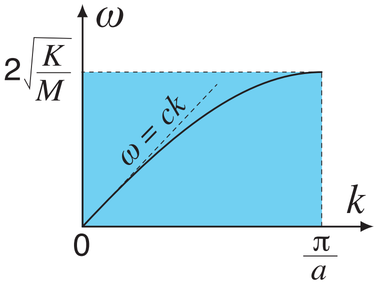
Substituting the ansatz (102) to the equation of motion (101) and cancelling common factors gives

$$-M\omega^2 = K(e^{ika} - 2 + e^{-ika}) = -4K\sin^2\frac{ka}{2}. \tag{105}$$

Here we have used formulas $e^{ix} = \cos x + i\sin x$ and $\frac{1}{2}(1 - \cos x) = \sin^2\frac{x}{2}$ [or, alternatively, we noticed $e^{ix} - 2 + e^{-ix} = (e^{ix/2} - e^{-ix/2})^2$]. From (105) we get the angular frequency

$$\omega = \pm 2\sqrt{\frac{K}{M}}\sin\frac{ka}{2}. \tag{106}$$

This so-called *dispersion relation* is presented in the figure below.



We study different ranges of k. (The case $\omega < 0$ does not bring anything new so we limit to $\omega \geq 0$).

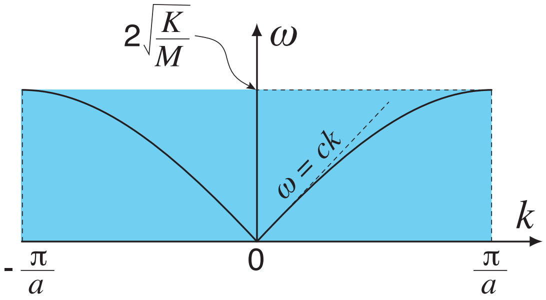
1) $0 < k \ll \pi/a$. This can be written $\lambda \gg a$, that is, the wave length is much larger that the lattice spacing. The wave means oscillation in density. Thus it can be identified as *sound wave* in a solid. In this region the dispersion relation (106) can be approximated by linear dependence

$$\omega = ck, \tag{107}$$

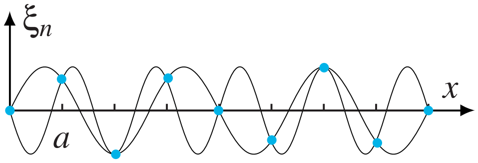
where $c = a\sqrt{K/M}$ is the speed of sound.

2) $k \sim \pi/a$. In this region, the wave length is of the same order of magnitude as the lattice constant a. The angular frequency is close to its maximum value $\omega_{\max} = 2\sqrt{K/M}$. In this region the speed $v = \omega/k$ (104) of the wave deviates from c. The corresponding property for light (dependence of speed on the wave length and thus on its color) is called dispersion. This is the reason for the name of the relation $\omega(k)$ (106).

3) $k < 0$. These solutions correspond to waves traveling to negative x direction. There are included in the following figure



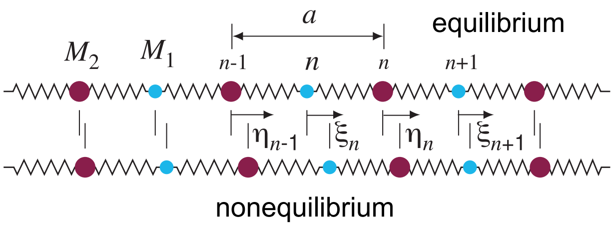
4) $|k| > \pi/a$. This region does not lead to any new solution. This can be seen writing the solution (102) for the wave number $k + (2\pi/a)$. We see that $\xi_n(t)$ is the same as for wave number k. The figure below aims to clarify that the same ξ_n's are obtained by different k's. Thus the wave number can be limited to range $|k| \leq \pi/a$.



The results above show that the lattice vibrations are not possible above the frequency $\nu_{\max} = \omega_{\max}/2\pi \sim 10^{13}$ Hz. This is very large compared to audible sound frequencies, which are below 20000 Hz. In the audible range the linear approximation (107) is very accuate.

Two-atom basis

(S 10.2) We study lattice vibrations in an atomic string with basis of two atoms of different mass.



As above, we can write equations of motion

$$M_1 \frac{d^2\xi_n}{dt^2} = K(\eta_n - 2\xi_n + \eta_{n-1})$$

$$M_2 \frac{d^2\eta_n}{dt^2} = K(\xi_{n+1} - 2\eta_n + \xi_n). \qquad (108)$$

We make the ansatz

$$\begin{aligned} \xi_n(t) &= A_1 e^{i(kna-\omega t)} \\ \eta_n(t) &= A_2 e^{i(kna-\omega t)}. \end{aligned} \qquad (109)$$

Substituting and cancelling common factors gives

$$\begin{aligned} -M_1\omega^2 A_1 &= -2KA_1 + K(1+e^{-ika})A_2 \\ -M_2\omega^2 A_2 &= K(1+e^{ika})A_1 - 2KA_2. \end{aligned} \qquad (110)$$

We rearrange this as

$$\begin{aligned} (M_1\omega^2 - 2K)A_1 + K(1+e^{-ika})A_2 = 0 \\ K(1+e^{ika})A_1 + (M_2\omega^2 - 2K)A_2 = 0. \end{aligned} \qquad (111)$$

In order to find a nonzero solution, the two equations have effectively to be a single equation only. The standard way to express this is that determinant of the coefficients have to vanish,

$$\det \begin{pmatrix} M_1\omega^2 - 2K & K(1+e^{-ika}) \\ K(1+e^{ika}) & M_2\omega^2 - 2K \end{pmatrix} = 0. \qquad (112)$$

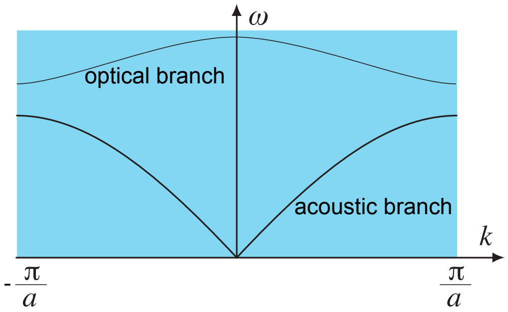
We get

$$M_1 M_2 \omega^4 - 2K(M_1+M_2)\omega^2 + 4K^2 \sin^2\frac{ka}{2} = 0, \qquad (113)$$

that can be solved

$$\omega^2 = \frac{K}{M_1 M_2}\left[M_1 + M_2 \pm \sqrt{M_1^2 + M_2^2 + 2M_1M_2\cos(ka)}\right]. \qquad (114)$$

Equation (114) gives two solutions, which are illustrated in the figure. One solution, called the *acoustic branch*, is of the same type as in a chain of identical atoms (106). The additional solution is called the *optical branch*.

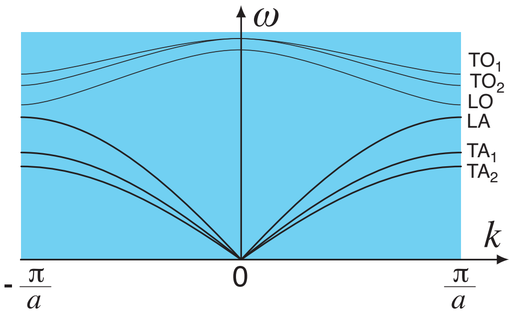


In the optical branch neighboring atoms vibrate against each other, whereas in the acoustic branch they mostly move in the same direction (most accurately at $|k| \ll \pi/a$).

The frequency ν in the optical branch has order of magnitude 10^{13} Hz. This corresponds to the infrared region of electromagnetic radiation. Especially in materials of ionic bonding the optical branch couples strongly with electromagnetic radiation because the electric field drives the oppositely charged basis atoms in opposite directions.

Lattice vibrations in 3 dimensions

We considered above longitudinal vibrations in a one-dimensional atomic chain. In 3 D we have to take account that the vibrations can take place in 3 perpendicular direction. This means that instead of one mode, there are three vibrational modes. In addition, all modes depend on the propagation direction of the wave, which is determined by the direction of the wave vector $\boldsymbol{k} = (2\pi/\lambda)\hat{\boldsymbol{k}}$. In special directions the modes can be described as one longitudinal mode and two transverse modes. The figure illustrates the obtained dispersion relations, where TA_1 = transverse acoustic mode 1 etc.



Quantum theory of lattice vibrations

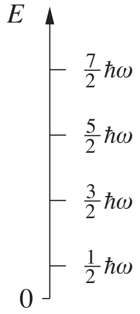
(S 9.3) Let us consider one *mode* of lattice vibration: we fix $\boldsymbol{k}$ and the oscillation direction (labelled by index s). Each oscillation mode $(\boldsymbol{k}, s)$ has a precise angular frequency ω. The vibration is harmonic because

$$\xi_n^{\text{physical}} = \mathrm{Re}(Ae^{ikan})\cos(\omega t) + \mathrm{Im}(Ae^{ikan})\sin(\omega t). \tag{115}$$

Thus a mode corresponds to a harmonic oscillator. Mode's energy $E = E_{\text{kin}} + E_{\text{pot}}$ is the sum of the kinetic energies of the atoms and the potential energies of the bonds (modeled by springs) associated with this mode.

The oscillations of the crystal can be seen as independent oscillations in each of the modes $(\boldsymbol{k}, s)$.

We now consider lattice vibrations using quantum mechanics. The dispersion relations (106) and (114) above remain unchanged but now each harmonic oscillator should obey quantum mechanics. This means, for each mode the possible energy eigenvalues are given by (34), where ω is the oscillation frequency of the mode.



The quantization of the energy eigenvalues has important effect on several physical phenomena. Here we consider *heat capacity*.

Heat capacity of a harmonic oscillator

(S 2.1) Our starting point is *Boltzmann distribution*, also known as *Gibbs distribution*. It will be derived in the course Thermodynamics and Statistical Physics. The Boltzmann distribution says that if we have states j with energies E_j, the probability of each state is

$$p_j = \frac{1}{Z} e^{-\beta E_j}. \tag{116}$$

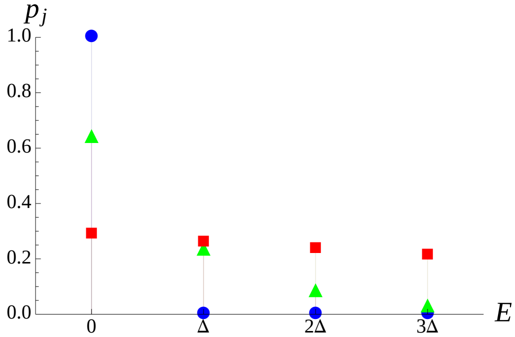
Here Z is a normalization factor, which is fixed by the condition that sum of probabilities of all states is equal to unity,

$$\sum_j p_j = 1 \Leftrightarrow Z = \sum_j e^{-\beta E_j}. \tag{117}$$

The factor multiplying energy in the exponent is

$$\beta = \frac{1}{k_{\text{B}} T}, \tag{118}$$

where T is the absolute temperature and $k_{\text{B}} = 1.380658 \times 10^{-23}$ J/K is the Boltzmann constant.



The figure is an example of probabilities of states in a system consisting of four states at energies $E_j = n\Delta$ where $n = 0, 1, 2$ and 3. The circles give the probabilities at a low temperature and the squares at a high temperature compared to temperature Δ/k_B, at which the occupations are shown by triangles. At room temperature $k_{\text{B}}T = 0.025$ eV. Thus the triangles correspond to room temperature if $\Delta = 0.025$ eV. The transition between such levels correspond to frequency

$\nu = \Delta/h = 6 \times 10^{12}$ Hz of electromagnetic radiation, or wave length $\lambda = c/\nu = 50\ \mu$m, in the infrared region.

A simple rule is the following: the states whose energy is less than the “thermal energy” $k_\mathrm{B}T$ above the ground-state energy are possible. The higher energy states are unlikely.

We apply the distribution (116) to a harmonic oscillator. We get

$$\begin{aligned} Z &= \sum_j e^{-\beta E_j} = e^{-\beta\hbar\omega/2}\sum_{j=0}^{\infty} e^{-\beta\hbar\omega j} \\ &= e^{-\beta\hbar\omega/2}\sum_{j=0}^{\infty}\left(e^{-\beta\hbar\omega}\right)^j = e^{-\beta\hbar\omega/2}\frac{1}{1-e^{-\beta\hbar\omega}}, \end{aligned} \tag{119}$$

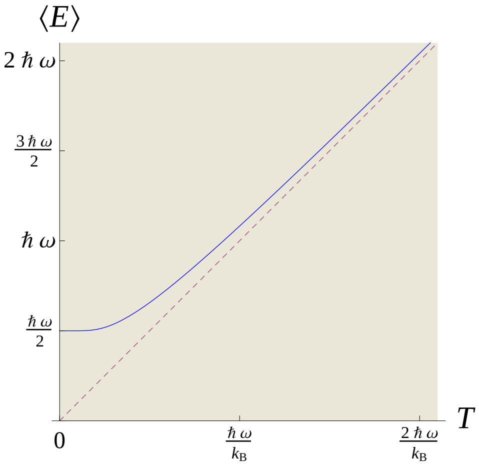
where we have used the expression of geometric series in the last step. In order to calculate the energy, it is convenient first to derive the general formula

$$\begin{aligned} \langle E\rangle &= \sum_j E_j p_j = \frac{1}{Z}\sum_j E_j e^{-\beta E_j} = -\frac{1}{Z}\frac{d}{d\beta}\sum_j e^{-\beta E_j} \\ &= -\frac{1}{Z}\frac{dZ}{d\beta} = -\frac{d\ln Z}{d\beta}. \end{aligned} \tag{120}$$

Applying this to harmonic oscillator (119), a short calculation gives

$$\langle E\rangle = \frac{\hbar\omega}{2} + \frac{\hbar\omega}{e^{\beta\hbar\omega}-1}. \tag{121}$$

This is shown by continuous line in the figure.



The result can be understood as follows. At low temperatures, $T \ll \hbar\omega/k_\mathrm{B}$, the harmonic oscillations is in ground state (the lowest energy state). In this case the expectation value of the energy equals the ground state energy, $\langle E\rangle = \frac{1}{2}\hbar\omega$. At high temperatures, $T \gg \hbar\omega/k_\mathrm{B}$, many states of the oscillator are possible, and $\langle E\rangle \approx k_\mathrm{B}T$, which is shown by dashed line.

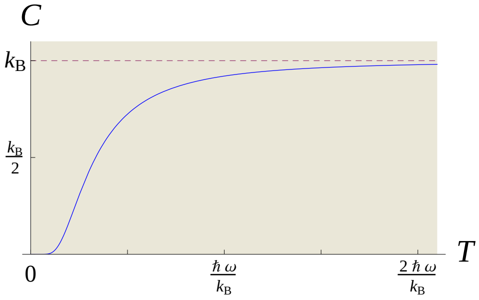
The heat capacity is defined as the derivative of average energy with respect to temperature,

$$C = \frac{d\langle E\rangle}{dT}. \tag{122}$$

Heat capacity tells how much energy has to be imported in order to raise the temperature by a given amount. Using formula (121) we get the heat capacity of a harmonic oscillator

$$\begin{aligned} C = \frac{d\langle E\rangle}{dT} &= \frac{\hbar^2\omega^2}{k_\mathrm{B}T^2}\frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega}-1)^2} \\ &= \frac{\hbar^2\omega^2}{4k_\mathrm{B}T^2\sinh^2\frac{\hbar\omega}{2k_\mathrm{B}T}}. \end{aligned} \tag{123}$$

This is shown by continuous line in the following figure.

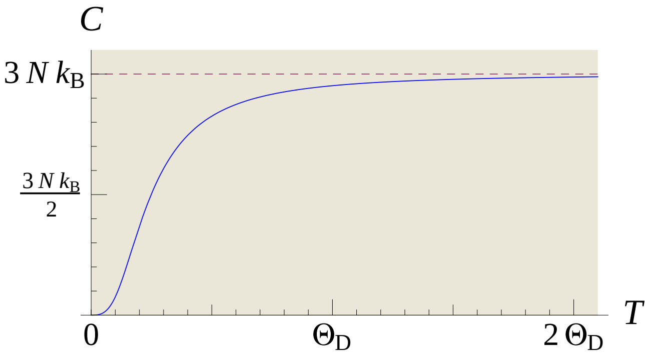


At high temperatures $k_\mathrm{B}T > \hbar\omega$, the heat capacity is nearly constant, $C \approx k_\mathrm{B}$. This is the classical result $C \equiv k_\mathrm{B}$ (dashed line), which is obtained in the limit $\hbar \to 0$. At low temperatures $k_\mathrm{B}T \ll \hbar\omega$, the heat capacity approaches zero. This can be understood so that when the thermal energy $k_\mathrm{B}T$ is small compared to the energy difference $\hbar\omega$ between the first excited state and the ground state, the probability of the excited state stays small, and thus the oscillator does not take energy in spite of a raise of temperature. Raising the temperature, the probability of the excited states is increased substantially when $k_\mathrm{B}T$ starts to be of the same order of magnitude as the excitation energy $\hbar\omega$.

We apply the result to lattice vibrations. The simplest model of a solid would be that all atoms oscillate at the same frequency ω. The number of such modes is $3N$, where N is the number of atoms in the crystal. The factor 3 comes because each atom can oscillate in three orthogonal directions. Such a model is called *Einstein model.* Einstein model has heat capacity that is the same as for a single harmonic oscillator but multiplied by $3N$. Thus at high temperatures, the oscillators obey classical result and the heat capacity $C \approx 3Nk_\mathrm{B}$, which is known as *Dulong-Petit law.* At low temperatures ($T < \hbar\omega/k_\mathrm{B}$) the heat capacity is reduced and approaches zero when $T \to 0$.

Historically, it was found experimentally that the heat capacity at high temperatures approaches the classical value $C \approx 3Nk_\mathrm{B}$, but is reduced at low temperatures. Einstein showed in 1907 with his model that this temperature dependence can be understood on the basis of quantum mechanics.

(S 2.2) A more advanced model is called *Debye model.* It takes in to account that instead of one frequency, there are many different frequencies in a crystal, as we derived above [Eq. (106)]. When the temperature is high so that $k_B T > \hbar\omega_{\max}$, all oscillators obey the classical law and the heat capacity $C \approx 3Nk_B$. At low temperatures $k_B T \ll \hbar\omega_{\max}$, part of the oscillators is "frozen" because their energy is so high that they are not excited. They do not contribute to the heat capacity. Only those modes with small frequency (small k) contribute to the heat capacity at a low temperature.



The figure shows heat capacity according to Debye model. It assumes linear dependence $\omega = ck$ for $k < k_D$ where the "cut-off" k_D is determined from condition that there are $3N$ modes. It describes rather well the heat capacity of many substances. The quantity $\Theta_D = \hbar c k_D / k_B \sim \hbar\omega_{\max}/k_B$ is called Debye temperature. It is typically a few hundred degrees in elemental substances.

substance	Debye temperature (K)
C (diamond)	1860
Na	150
Al	394
Si	625
Fe	420
Cu	315
Hg	100
NaCl	321

Phonons

(S 9.3) The concept of *phonon* is often used in connection with lattice vibrations. Let us consider the case that the mode $(\boldsymbol{k}, s)$ is in state j, i.e. the mode energy is $E = \hbar\omega(j + \frac{1}{2})$. An alternative way to express this is to say that the mode has j phonons. One can speak about phonons as they were particles, although they represent oscillation states. For example, one phonon energy is $\hbar\omega_{\boldsymbol{k},s}$. The whole lattice energy is

$$E = \sum_{\boldsymbol{k}} \sum_{s} \hbar\omega_{\boldsymbol{k},s}(j_{\boldsymbol{k},s} + \frac{1}{2}), \tag{124}$$

where $j_{\boldsymbol{k},s}$ the number of phonos in mode $(\boldsymbol{k}, s)$. The expectation value of energy (121) is interpreted that the average number phonons is

$$\langle j \rangle = \frac{1}{e^{\beta\hbar\omega} - 1}, \tag{125}$$

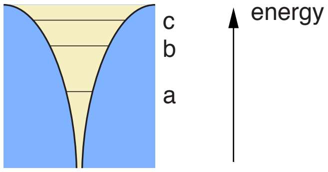
which is a special case of Bose-Einstein distribution.

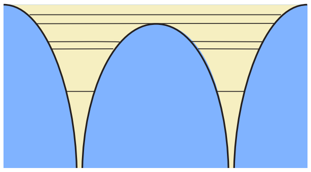
Phonons are quanta of lattice vibrations. This is analogous to photons, which are quanta of electromagnetic field. One phonon energy is $h\nu = \hbar\omega$. A photon has momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$. Continuing the analogy, the phonon also has momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$. The phonon concept is useful, for example, in study of heat conductivity. Phonons transport heat from region of high temperature to region of low temperature. What limits the transport is that phonons collide with each other and with defects of the crystal, and get scattered to other directions.

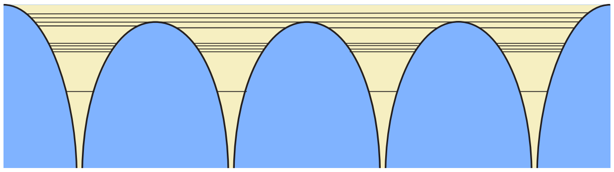
4. Solid state: electronic properites

(YF 42.4, S 11 and 16.1) We now start applying quantum mechanics to electrons in 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. We also considered the binding of two atoms to form a molecule. Now we study what happens the atoms form a solid. The figure below depicts three energy level in an atom, marked by a, b and c.







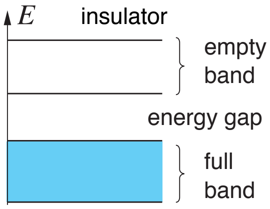
When two atoms are brought close to each other, the energy barrier between them gets lower. This means that the electrons from one atom can tunnel to the other one. However, for the state a the potential barrier is high, and nothing really happens. This is an important result: the inner electron states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

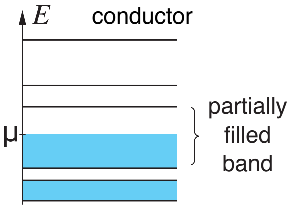
At level *b* the electron tunneling is important. At level *c* the electrons can move freely in the molecule.

When the atoms form a bond, the energy levels are restructured. In a molecule of two atoms, an atomic energy level forms 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*. The energies 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.)





We will consider each case separately below.

4.1 Free electron model of metals

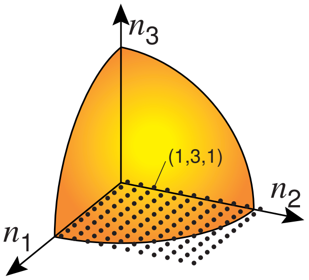
(YF 42.5, S 4.1-4.2) In metals the partly filled band is called *conduction band*. The ability of metals to conduct electricity is based on the existence of a conduction band. This is because the bands that are full do not conduct. (We will discuss this soon.) The electrons in the conduction band are called *conduction electrons*. These can move rather freely through the metal. We now study their behavior in more detail.

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 (43) 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 (51) with positive integers n_1, n_2 and n_3. The number of wave functions below E correspond to quantum numbers satisfying

$$n_1^2 + n_2^2 + n_3^2 < \frac{2mL^2E}{\pi^2\hbar^2} = \kappa^2. \qquad (126)$$

We draw a space with axes n_1, n_2 and n_3. In this space the condition (126) 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

$$N_d(E) = \frac{1}{8}\left(\frac{4}{3}\pi\kappa^3\right) = \frac{\sqrt{2m^3}V}{3\pi^2\hbar^3}E^{3/2}, \tag{127}$$

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

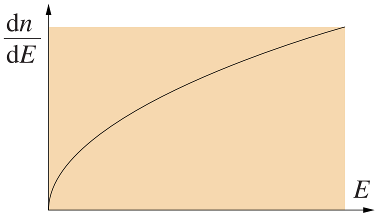
We define the density (number/volume) of electron states $n(E)$ below energy E. It is given $n(E) = 2N_d(E)/V$, where the factor 2 comes from two values of the spin quantum number m_s. We get

$$n(E) = \frac{2\sqrt{2m^3}}{3\pi^2\hbar^3}E^{3/2}. \tag{128}$$

We also define *density of states*

$$g(E) = \frac{dn}{dE} = \frac{\sqrt{2m^3}}{\pi^2\hbar^3}\sqrt{E}. \tag{129}$$

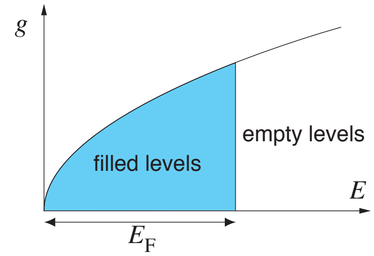
Here "density" means the number of states per energy interval. The number of states in energy interval $(E, E + dE)$ is $g(E)dE$. This is proportional to the square root of energy.



The number of levels (128) below energy E is recovered as integral over g,

$$n(E) = \int_0^E g(E)dE = \frac{2\sqrt{2m^3}}{3\pi^2\hbar^3}E^{3/2}. \tag{130}$$

Suppose the electron 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 (128),

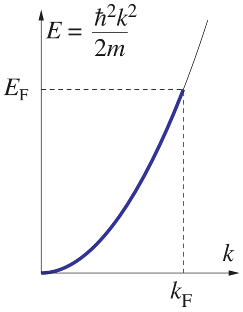
$$E_{\mathrm{F}} = \frac{\hbar^2}{2m}\left(3\pi^2 n_e\right)^{2/3}. \tag{131}$$

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 (131).

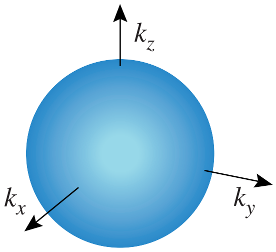
	Z_c	E_{F} (eV)
Na	1	3.2
Cu	1	7.0
Fe	2	11.1
Al	3	11.7
Pb	4	9.5

We derived that the energy levels are filled up to Fermi energy (131). According to $E = \hbar^2k^2/2m$ there is corresponding *Fermi wave nunber*

$$k_{\mathrm{F}} = \left(3\pi^2 n_e\right)^{1/3}. \tag{132}$$



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.*

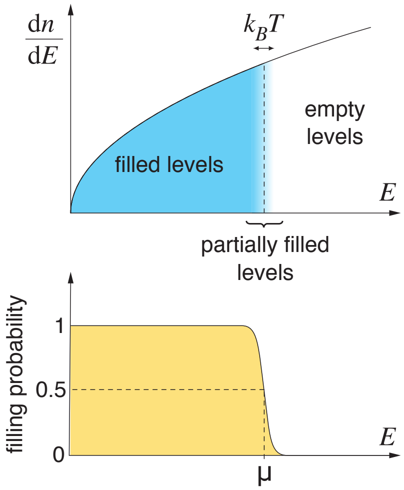


One also defines Fermi momentum $p_{\mathrm{F}} = \hbar k_{\mathrm{F}}$ and Fermi velocity $v_{\mathrm{F}} = p_{\mathrm{F}}/m$. The typical order of magnitude $v_{\mathrm{F}} \sim 10^6$ m/s.

More generally, at a given temperature, the levels are filled according to the Fermi-Dirac distribution (also known as Fermi function)

$$f(E) = \frac{1}{e^{(E-\mu)/k_{\mathrm{B}}T} + 1}. \tag{133}$$

This formula will be justified in the course Thermodynamics and Statistical Physics. 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. Because this is small compared to E_{F}, we conclude that the partially filled levels occur only in vicinity of Fermi energy



4.2 Electron in a periodic potential

(S 15) Above we studied free electrons, which was then applied to conduction electrons in metals. The assumption in free electron model is that the potential vanishes. In reality, the ions generate a nontrivial potential $U(x)$, which we study next.

In a crystalline material the effect of $U(x)$ is simplified because of symmetry. We consider one-dimensional case with lattice constant a. Then

$$U(x + a) \equiv U(x). \tag{134}$$

That is, $U(x)$ is a periodic function with period a. This means that we need to know the potential only in one unit cell. Because of periodicity (134), the potential is then known in the whole lattice.

Sums of sinusoidal waves can be used to represent more complicated functions. This is especially useful for periodic functions, where it is called *Fourier series*. We apply this to express $U(x)$,

$$U(x) = \sum_{n=-\infty}^{\infty} U_n e^{i2\pi nx/a}, \quad n = 0, \pm1, \pm2, \ldots. \tag{135}$$

We notice that the exponential factor $e^{i2\pi nx/a}$ is periodic in x with period a. Thus the sum of them with coefficients U_n satisfies the periodicity (134). Mathematicians have shown that the infinite sum (135) can represent all sufficiently well behaved functions. The fact that $U(x)$ is real leads to the constraint $U_{-n} = U_n^*$.

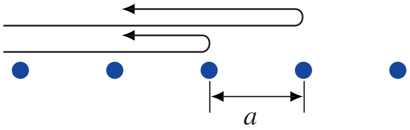
Our task is to solve the Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x). \tag{9}$$

Let us consider a traveling wave (11) $\psi = Ae^{ikx}$. This is an eigenstate of the kinetic energy operator $-(\hbar^2/2m)d^2/dx^2$ but operation by the potential energy gives

$$U(x)e^{ikx} = \sum_{n=-\infty}^{\infty} U_n e^{i(k+2\pi n/a)x}. \tag{136}$$

We see that the original wave at wave number k is coupled to waves at wave numbers $k + 2\pi n/a$ with $n = \pm1, \pm2, \ldots$. This is a complicated problem in the general case. Here we can make progress by assuming $U(x)$ is small. Then the main effect of the potential is to scatter electrons from one direction to another but nearly at the same energy. The figure below illustrates the scattering of a wave from two consecutive atoms in the crystal.



There is a constructive interference of the two scattered waves when the path length difference $2a$ equals integer number of wave lengths λ. This gives

$$k = \frac{l\pi}{a}, \quad l = \pm1, \pm2, \ldots \tag{137}$$

Let us concentrate on the case $l = 1$. We claim that for solving the Schrödinger equation (9) in the region $k \approx \pi/a$, a reasonable approximation is to consider the ansatz

$$\psi = Ae^{ikx} + Be^{i(k-2\pi/a)x}, \tag{138}$$

and to neglect the couplings to all other wave number states. This procedure is known as *degenerate perturbation theory.*

We substitute (138) into (9), neglect coupling to other states, and simplify by assuming $U_0 = 0$. Demanding a solution valid at all x gives the pair of equations

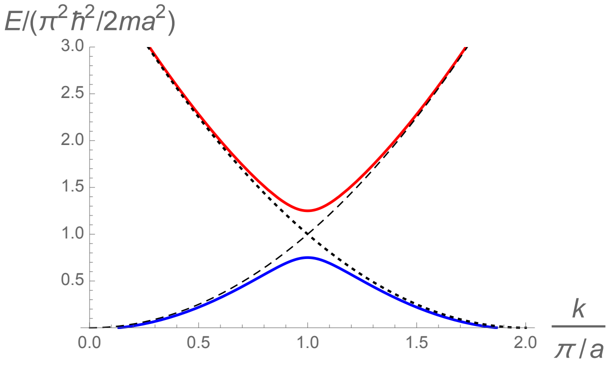
$$\begin{gathered}\frac{\hbar^2 k^2}{2m} A + U_1 B = EA \\ \frac{\hbar^2}{2m}\left(k - \frac{2\pi}{a}\right)^2 B + U_{-1} A = EB.\end{gathered} \tag{139}$$

For nontrivial solution the determinant of the coefficient matrix of A and B has to vanish,

$$\det\begin{pmatrix} \frac{\hbar^2k^2}{2m}-E & U_1 \\ U_{-1} & \frac{\hbar^2}{2m}\left(k-\frac{2\pi}{a}\right)^2-E \end{pmatrix}=0. \tag{140}$$

From this we find the energy eigenvalues

$$E=\frac{\hbar^2}{4m}\left[k^2+\left(k-\frac{2\pi}{a}\right)^2\right] \pm\sqrt{\frac{\hbar^4}{16m^2}\left[k^2-\left(k-\frac{2\pi}{a}\right)^2\right]^2+|U_1|^2}. \tag{141}$$



The result (141) is plotted in the figure. We discuss several features of the result below.

1) We see that the deviation from the free electron model (25) (dashed line) is small except close to $k=\pi/a$. Instead of the continuous spectrum of the free-electron model, a gap of magnitude $2|U_1|$ opens in the energy spectrum. Perviously, we argued about energy bands and gaps starting from weak overlap of atomic orbitals. Here we started from the opposite assumption of free electrons, and introduced weak periodic potential. Neither of the two approximations is quantitatively correct in real solids. In spite of this, the result that in both limits energy bands and gaps are obtained, makes us to believe that they do appear in real solids as well.

2) Besides results close to the dashed line, the figure shows branches close to the dotted line (the upper branch at $k<\pi/a$ and the lower branch at $k>\pi/a$). We can study the corresponding wave function by solving (139). This gives that the B term is the dominant one in the wave function (138). Thus the dominant wave number is $k-2\pi/a$ instead of k. This ambiguity is possible because the wave function (138) is not an eigenstate of the momentum operator (10).

More generally, one can prove *Bloch theorem*. It states that any solution of the Schrödinger equation (9) for a periodic potential (134) can be written in the form

$$\psi(x)=e^{ikx}u(x),\quad u(x+a)=u(x). \tag{142}$$

This *Bloch wave function* is a product of two factors. The exponential part is as for a free particle (11). The part $u(x)$ is a periodic function with period a. The wave number k in (142) is called *crystal momentum*, see explanation below.

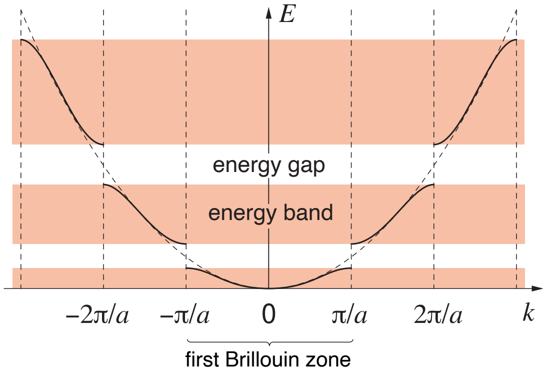
One consequence of the Bloch theorem is that $|\psi(x)|$ and thus the probability density is a periodic, although $\psi(x)$ is not.

Because of the periodicity of $u(x)$, we can write it using Fourier series as in (135),

$$u(x)=\sum_{n=-\infty}^{\infty}u_ne^{i2\pi nx/a},\quad n=0,\pm1,\pm2,\ldots. \tag{143}$$

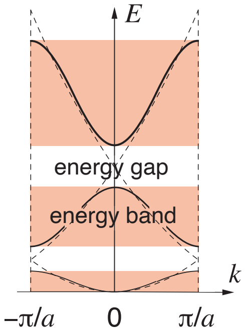
Using this we see that the Bloch wave function $\psi(x)$ (142) is a superposition of states of momenta $p=\hbar k+2\pi\hbar n/a$ with amplitudes u_n at integer n. Thus, neither k nor $\hbar k$ is the momentum of the state, but the crystal momentum k specifies the momenta that occur in the superposition state. This is a more general statement of the result we obtained in the calculation and figure above with the superposition state of two momentum values (138).

3) The calculation above addressed the effect of the periodic potential near $k=\pi/a$. More generally, a similar calculation should be done at all $k=\pi l/a$ with integer l. The result is that gaps will open at $l=\pm1,\pm2,\ldots$ (depending on the value of $|U_l|$).



The figure shows the electron energies in so-called extended-zone scheme. The *first Brillouin zone*, where $-\pi/a\le k\le\pi/a$, is indicated.

4) We saw (in point 2) that k specifies the momenta in the superposition state but it does not determine which of them appears with largest amplitude. We can use this freedom to select the values of k in a specified range.



The figure shows all our results moved to the first Brillouin zone. This compact representation is called reduced zone scheme. One has to remember that the states at the boundaries of the zone, $k = \pm\pi/a$, should be identified. Sometimes this looks too discontinuous, and it is advantageous to use other schemes, where the curves are continued outside of the first Brillouin zone (but at the expense that same physical state may be represented more than once).

5) Let us finally note that the result obtained in figure after eq. (141) is an example of *avoided crossing* of levels. It is common in quantum mechanics (or in any wave theory) that two uncoupled levels (here the dashed and dotted lines) cross in energy as a function of some parameter (here k). When a coupling between the two levels is added, the crossing is avoided. [Note that we encountered a somewhat similar situation for lattice vibrations, where a gap opened in the frequency spectrum in a chain of atoms of different mass, eq. (114).]

Electron dynamics in a lattice

(S 17.1) Above we considered electron waves, which are characterized by the wave number k and angular frequency $\omega = E/\hbar$. These waves are delocalized. That is, they extent in principle from $x = -\infty$ to $x = +\infty$. In order to have a localized wave, one has to form a wave packet that contains several wave numbers, for example in the range $(k - \Delta k, k + \Delta k)$. (earlier exercise) We claim that such a wave packet moves with *group velocity*

$$v = \frac{d\omega}{dk}. \tag{144}$$

This formula is explained in appendix C. (exericse)

Often the group velocity (144) depends on k. In this case the form of the wave packet deforms as it propagates, and the group velocity should be interpreted as some average velocity.

We apply the group velocity (144) to electrons. Using $\omega = E/\hbar$ we get the velocity of the electron

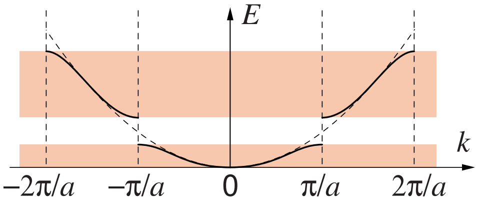
$$v = \frac{1}{\hbar}\frac{dE}{dk}. \tag{145}$$

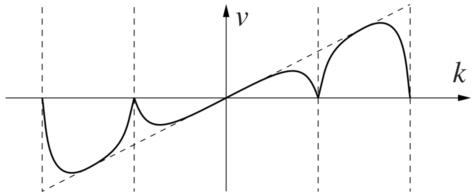
We suppose an external force F is applied to the electron. It makes the work $Fv\,dt$ which equals the energy change $dE = (dE/dk)dk$. Using formula (145) we get

$$\hbar\frac{dk}{dt} = F. \tag{146}$$

The force F could be caused by an electric field $\mathcal{E}$ so that $F = -e\mathcal{E}$, where $-e$ is the electron charge.

Equations (145) and (146) are familiar to us in the case of free electrons, $\hbar k = mv$ and $E = \hbar^2k^2/2m$. In a lattice their content is different because of the nontrivial $E(k)$.





The figure skethes v (145). In regions where the free electron model is good, $v \approx \hbar k/m$, but at zone boundaries (137) the velocity vanishes, $v \to 0$.

We consider an electron under a constant force F. According to (146) the wave number $k = k_0 + Ft/\hbar$ grows linearly starting from its initial value k_0. If the electron is close to the bottom of the band, it behaves nearly as free electron. Starting from rest ($v = 0$, $k_0 = 0$) the velocity increases linearly in the direction of the force. The situation is different close to the upper end of the band. When k approaches the zone boundary $k = \pi/a$, the velocity decreases. This can be understood by scattering of the electron to the opposite direction. The state is a superposition of waves propagating in opposite directions (138), and v should be interpreted as the averaged velocity in the wave packet. When k reaches the zone boundary, $k = \pi/a$, it will be identified with $k = -\pi/a$. Then k will continue increasing from $k = -\pi/a$ according to (146), implying that it moves in the direction opposite to the force. Thus the electron will not jump from one band to another but remains in the same band. This is an example of adiabatic theorem, discussed in Quantum mechanics course. (The electron remains in the same band if the electric field is not too strong. A strong field can cause “Zener tunneling” to the next band.)

The electron motion could be clarified defining *effective mass.* Near the bottom of the band the energy depends quadraticly on k. Thus we can write

$$E \approx E_0 + \frac{\hbar^2(k - k_0)^2}{2m^*} \tag{147}$$

where m^* a constant called effective mass. The same formula could be used also close to the top of a band, with the difference that m^* is negative.

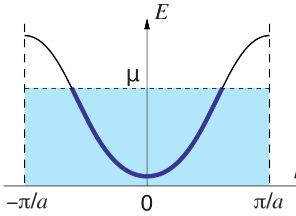
In region where the effective mass approximation is valid, we get from equations (145) and (146)

$$m^*\frac{dv}{dt} = F. \tag{148}$$

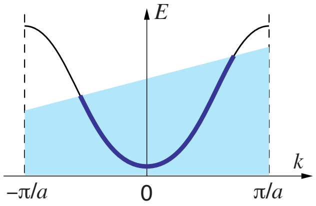
This is like Newton’s equation of motion, but now m^* can also be negative.

4.3 Electric conduction in metals

(S 16.1) We study a metal. We found previously that the electrons in the conduction band move at a high velocity $\sim v_F$. In the levels at wave vectors $\boldsymbol{k}$ and $-\boldsymbol{k}$, the motion is in opposite directions. These levels are equally likely to be occupied in equilibrium as they have the same energy. Thus the electric current vanishes in equilibrium, $\boldsymbol{j} = 0$.



electron flows to the right
and to the left cancel



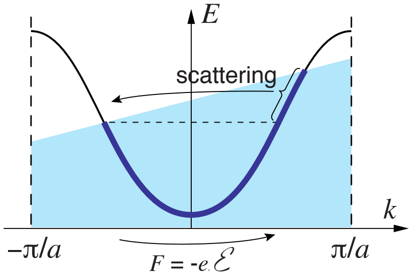
net electron flow to the right

In order to have electric current, $\boldsymbol{j} \neq 0$, 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.)

The imbalance of electrons can be achieved by external electric field. The force $F = -e\mathcal{E}$ accelerates the electrons so that their momentum $p = \hbar k$ grows according to (146). There are always imperfections of the lattice

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

These 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.



(S 3.0) A simple model of electric conduction was suggested by Drude (1900). The *Drude model* considers the momentum $\boldsymbol{p} = m\boldsymbol{v}$ that is the *averaged value over all conduction electrons*. Its equation of motion is assumed be

$$\frac{d\boldsymbol{p}}{dt} = -\frac{\boldsymbol{p}}{\tau} - e\boldsymbol{\mathcal{E}}. \tag{149}$$

This is Newton's equation of motion that has force $\boldsymbol{F} = -e\boldsymbol{\mathcal{E}}$ caused by electric field, and the additional term $-\boldsymbol{p}/\tau$ that describes scattering of electrons. 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 electric current density $\boldsymbol{j} = -en_e\boldsymbol{v}$, where n_e is the number density of the conduction electrons.

(The current density is defined as the electric current divided by the perpendicular area, through which the current flows. It is good to check that the units match in the formula $j = -en_e v$: $[j] = \mathrm{A/m^2} = \mathrm{C/s\,m^2}$, $[e] = \mathrm{C}$, $[n_e] = 1/\mathrm{m}^3$ and $[v] = \mathrm{m/s}$.)

(S 3.1.1) One particular case that can easily be solved is the response to a constant electric field $\boldsymbol{\mathcal{E}}$. Setting $d\boldsymbol{p}/dt = 0$, we get

$$\boldsymbol{j} = \sigma\boldsymbol{\mathcal{E}}, \tag{150}$$

where

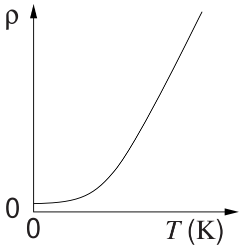
$$\sigma = \frac{e^2 n_e \tau}{m}. \tag{151}$$

Note that (150) is Ohm's law: the current depends linearly on the electric field. The more common form of Ohm's law $V = RI$ is obtained when we have a conductor of length L and cross sectional area A. Then the current $I = Aj$ and the voltage over the ends of the conductor $V = L\mathcal{E}$. Thus the resistance $R = L/\sigma A$. The quantity σ is the *conductivity* and its inverse $\rho = 1/\sigma$ is the *resistivity*.

The table below gives measured resistivities of some metals at the temperature $T = 273$ K. The relaxation times are calculated based on the measured resistivities.

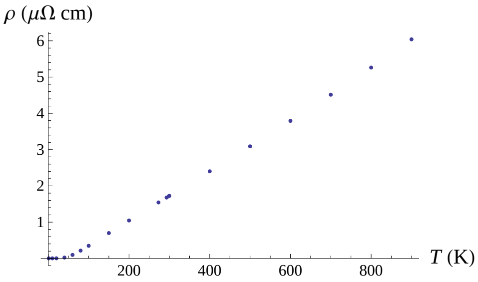
	ρ ($10^{-8}\Omega\,\mathrm{m}$)	τ (10^{-14} s)
Na	4.2	3.2
Cu	1.56	2.7
Fe	8.9	0.24
Al	2.45	0.80
Pb	19.0	0.14

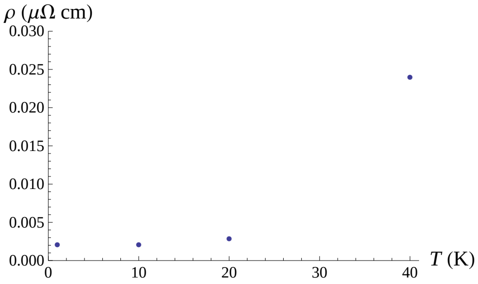
A typical dependence of resistivity $\rho = 1/\sigma$ on temperature is sketched in the following figure.



At room temperature the resistivity caused by lattice vibrations is dominant. With decreasing temperature it decreases towards zero. At low temperatures the resistivity comes from lattice defects as they cause resistivity that is temperature independent. The rule that the resistivities arising from the two sources can be added, $\rho = \rho_1 + \rho_2$, is called Mathiessen rule. Some metals become superconducting at low temperatures. This means that the resistivity vanishes abruptly below a temperature called critical temperature T_c (not shown in the figure)

The figures below show resistivity of copper, Cu, at different temperatures.





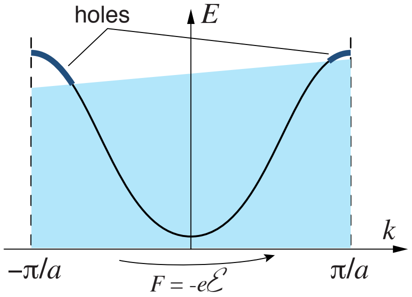
Full bands and holes

Above we looked at how electrons carry electric current in a partially filled band. Based on this we can also understand why a full energy band does not conduct. A full band has always electrons equally going in one direction than in the opposite direction. Thus the net current vanishes.

The fact that the full bands do not conduct is essential for the classification of solids into insulators, metals, and semiconductors (see above). The deep lying filled bands do not matter for conductivity.

There is an alternative way to describe electric conduction in a partially filled band. Instead of looking at the filled levels, we can concentrate on the unfilled levels. An elecron missing from a level is called *hole*. This way of looking is mostly useful when the energy band is almost full.

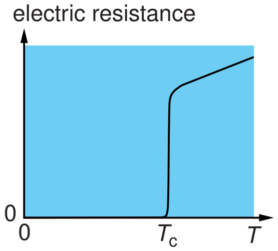
As above, consider a force that is driving the electrons towards increasing k. The situation with holes is as in the figure below.



In this case the velocity $v = (1/\hbar)dE/dk$ for holes is negative, on the average. That is, they move in the negative x direction. Compared to the case that the band would be full, the holes have a positive charge. Thus, instead of negatively charged electrons moving in the positive x direction, we can think the electric current to arise from positively charged holes that move in the negative x direction.

4.4 Superconductivity

(YF 42.8) By superconductivity we mean the phenomenon that the electrical resistivity of a material disappears below some temperature T_c.



Occurrence

- several metallic elements : Al, Nb, Sn, (but not in magnetic metals and in noble metals: Cu, Au, Ag)
- many alloys, e.g. Nb-Ti
- some compounds: Nb_3Ge, MgB_2, Y-Ba-Cu-O etc.

The temperature below which superconductivity occurs is called *critical temperature*, T_c. The list gives some critical temperatures.

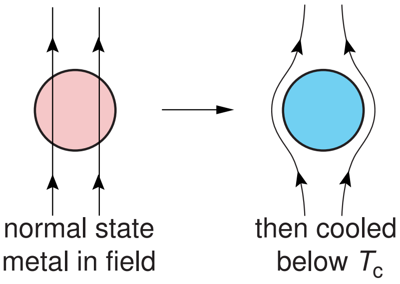
material	T_c (K)	$\mu_0 H_c(T=0)$ (mT)
Al	1.196	9.9
Hg	4.15	41
In	3.40	29.3
Pb	7.19	80.3
Nb	9.25	
Nb_3Ge	23	
MgB_2	39	
$YBa_2Cu_3O_{6+x}$	98	
$Tl_2Ca_2Ba_2Cu_3O_{10}$	125	

Infinte conductivity

In normal state metals the electric current is proportional to the electric field $\boldsymbol{j} = \sigma\boldsymbol{\mathcal{E}}$ (150). As stated above, in superconductors $\sigma = 1/\rho \to \infty$.

Meissner effect

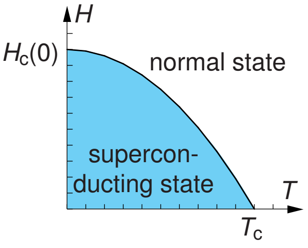
A more fundamental phenomenon than infinite conductivity is seen when a normal state metal is first placed in magnetic field B, and is then cooled into the superconducting state. It is observed that the magnetic field is expelled from the sample. This is called *Meissner effect.*



Critical field

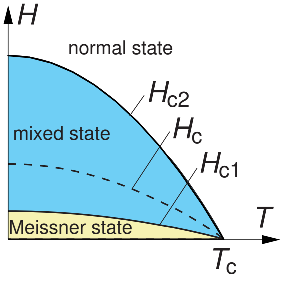
The Meissner effect is observed only if the magnetizing field H is not too large. It is observed that a transition between superconducting state and normal state takes place in *critical field* H_c, whose dependence on temperature is approximately

$$H_c(T) = H_c(0)\left[1 - \left(\frac{T}{T_c}\right)^2\right]. \tag{152}$$



Material showing this behavior are called *type I superconductors.*

Some superconducting materials have a mixed state between Meissner and normal states. These are called *type II superconductors.*

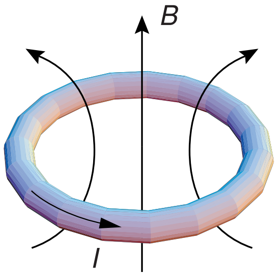


Persistent currents and flux quantization

Let us place a normal state ring in perpendicular magnetic field. When it is cooled below T_c, the magnetic field is expelled from the inside of the superconductor, but a magnetic flux through the ring remains. When the external field is removed, this flux remains unchanged. We deduce that a persistent current I is induced in the superconducting ring that generates the magnetic field $\boldsymbol{B}$. In addition, the magnetic flux $\Phi = \int d\boldsymbol{a} \cdot \boldsymbol{B}$ through the ring is quantized: it is an integer multiple of the flux quantum

$$\Phi_0 = \frac{h}{2e} = 2.07 \times 10^{-15}\ \mathrm{Wb}. \tag{153}$$

Here h is Planck's constant and e the elementary charge.



Theory of superconductivity

Many properties of superconductors can be understood based on *Bardeen-Cooper-Schrieffer (BCS) theory* (1957). It consists of two parts. 1) there is attractive force between electrons that is mediated by lattice vibrations. In superconducting metals this dominates the Coulomb repulsion between electrons. 2) Because of the attractive interaction, the electrons form pairs, so-called *Cooper pairs.* We study the latter part in more detail.

The many-body wave function takes the form consisting of pairs,

$$\psi_0(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2, \boldsymbol{r}_3\sigma_3, \ldots) = \phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) \times \phi(\boldsymbol{r}_3\sigma_3, \boldsymbol{r}_4\sigma_4)\phi(\boldsymbol{r}_5\sigma_5, \boldsymbol{r}_6\sigma_6)\ldots, \tag{154}$$

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

$$\phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) = -\phi(\boldsymbol{r}_2\sigma_2, \boldsymbol{r}_1\sigma_1). \tag{155}$$

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

We see that a large number of electrons behave according to a single pair wave function. This makes superconductivity to withstand perturbations that often destroy quantum coherence on a macroscopic scale. The pair wave function is said to be "macroscopic".

Consider a pair of particles. Assuming the center of mass $[\boldsymbol{R} = \frac{1}{2}(\boldsymbol{r}_1 + \boldsymbol{r}_2)]$ is at rest, the wave function can depend only on the relative location $\boldsymbol{r}_1 - \boldsymbol{r}_2$. We also assume the spin state is $\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow)$ (called "singlet"), which is believed to be the case in most superconductors. Thus

$$\phi(\boldsymbol{r}_1\sigma_1, \boldsymbol{r}_2\sigma_2) = \psi(\boldsymbol{r}_1 - \boldsymbol{r}_2)\frac{1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow). \tag{156}$$

If we now look at 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 $-\boldsymbol{k}$:

$$\psi(\boldsymbol{r}_1 - \boldsymbol{r}_2) = \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}_1} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_2} \tag{157}$$

This is understandable because the center of mass is assumed to be at rest, and therefore the total momentum should vanish. [In order to understand superconductivity the lattice potential is not essential. Therefore we have used sinusoidal waves (43) instead of the Bloch waves (142) in (157).]

A special case of the pair state (157) 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 (157) with many non-vanishing $c_{\boldsymbol{k}}$'s.

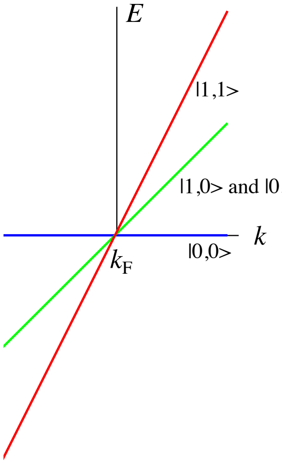
Let us look the pair state from the point of a fixed single $\boldsymbol{k} = \boldsymbol{k}_0$. The states included in (157) are $|0,0\rangle$ and $|1,1\rangle$. Here the two numbers mean the occupation of levels with wave vectors $\boldsymbol{k}_0$ and $-\boldsymbol{k}_0$, respectively. The state $|1,1\rangle$ means that the electrons in the pair occupy these momentum values. The state $|0,0\rangle$ 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

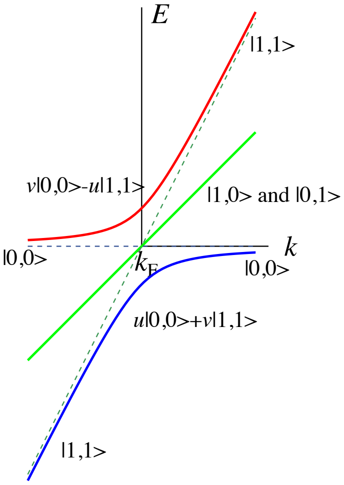
$$\begin{aligned} |0,0\rangle &: E = 0 \\ |1,0\rangle &: E = \xi_k \\ |0,1\rangle &: E = \xi_k \\ |1,1\rangle &: E = 2\xi_k \end{aligned} \tag{158}$$

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



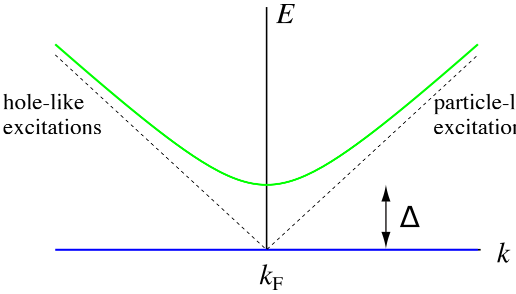
It is obvious that the system ground state consist of states $|1,1\rangle$ at $k < k_F$ and of states $|0,0\rangle$ at $k > k_F$. This is the Fermi sphere of noninteracting particles.

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 an avoided crossing, similarly as discussed in a different context above [Eq. (141)].



The ground state (blue in the figure) is now a superposition of states $|0,0\rangle$ and $|1,1\rangle$. 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 electric conduction at low temperatures.

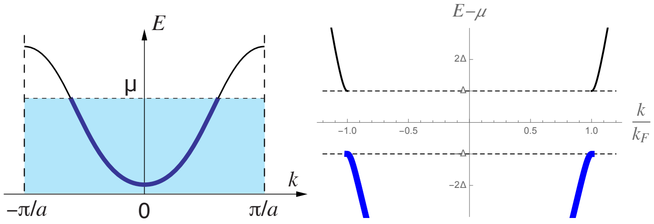
In addition to the superconducting ground state, the picture above gives also the elementary excited states. Namely, the lowest energy exited state is to break the pair at some $\boldsymbol{k}$ and replace it by state $|1,0\rangle$ or $|0,1\rangle$. The excitation energy is just the energy difference 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).

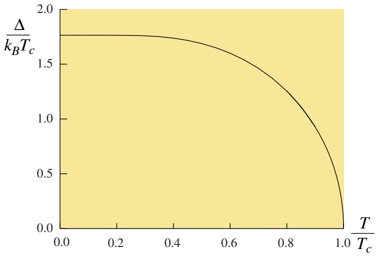
	T_c (K)	Δ (eV)
Al	1.2	1.7×10^{-4}
Sn	3.75	5.6×10^{-4}
Pb	7.22	13×10^{-4}

The superconducting energy gap Δ is illustrated in the figures below

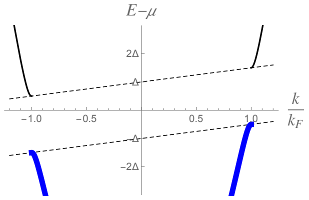


Because the gap is very small compared to the width of the conduction band, it will not be visible in the left panel showing the band scale. The right panel zooms in the energy range near the Fermi level, where the gap in the excitation spectrum becomes visible. At $T = 0$ all the states below the gap are filled and all the states above are empty.

Because of smallness of the gap Δ, electrons are excited across the gap already at temperatures of a few kelvin. The excitations limit the momentum states available for the pairs in (157). This leads to decrease of Δ with increasing T. The gap vanishes at $T = T_c$, and the metal is in the normal state at higher temperatures.



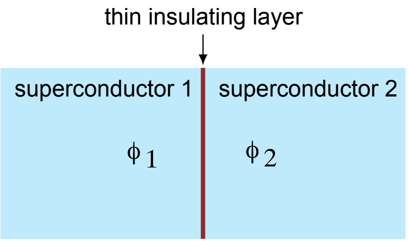
The superconducting gap differs from the crystal band gap that it need not be horizontal in the $E(k)$ plot. An electric field not only drives imbalance in the electron distribution, but also tilts the energy gap.



Since there are more electrons moving to the right than to the left, there is net electric current, $j \neq 0$. Lattice vibrations and other defects of the lattice cannot scatter the electrons since there are no empty states available at the same or lower energy. This explains the infinite conductivity in the superconducting state.

Also other properties of superconductors can be explained based on the Cooper pairs. The quantization of magnetic flux to $n\Phi_0$ (n integer) in a superconducting ring is a consequence that the pair wave function $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) \propto e^{i\phi}$, where the phase ϕ changes by $2\pi n$ in circling the ring.

Josephson effect



We study two pieces of a superconducting metal that are connected by a thin insulating layer. The electrons tunneling through the layer give rise to current. The current I depends on the phases ϕ_1 and ϕ_2 of the superconductors in the form

$$I = I_0 \sin(\phi_2 - \phi_1). \tag{159}$$

Under constant voltage V, the phase difference $\phi_2 - \phi_1$ changes as $\phi_2 - \phi_1 = \omega t$, where

$$\omega = \frac{2eV}{\hbar} \tag{160}$$

(This can be seen as $E = \hbar\omega$, where $E = 2eV$ is the energy associated with a pair tunneling form one side to the other.) Substituting into (159) we see that the current oscillates at frequency (160). Because the frequency can be measured accurately, this can be used as a voltage standard. For example, $V = 0.1$ mV corresponds to frequency $\nu = \omega/2\pi \approx 48$ GHz.

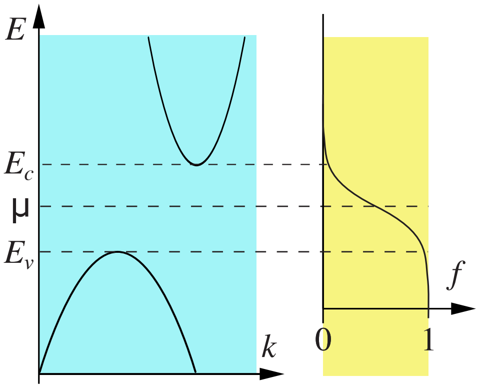
4.5 Semiconductors

(YF 42.6, S 17) Semiconductors were studied experimentally already in the 18 hundreds. They conduct electricity, but the conductivity is not as good as in metals. Opposite to the case of metals, their conductivity increases (resistivity decreases) with increasing temperature. The conductivity also depends strongly on impurities, in contrast to a metal. Semiconductors have a rectifying property: the observed conductivity in one direction is different from the one in the opposite direction. In the following we aim to understand these properties of semiconductors. The understanding is based on the band theory and the semiclassical theory of electron dynamics.

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 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

The region near the band gap is illustrated in the figure below. The highest energy in the valence band is denoted by E_v and the lowest energy in the conduction band by E_c. Thus the band gap $E_g = E_c - E_v$. The figure below illustrates an *indirect band gap*, where E_v and E_c occur at different k's. A material where they are aligned in k is called to have a *direct band gap*.



The right hand side of the figure shows the Fermi function (133). 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 $k_B T$, f is small but not negligible at E_c. Thus there are some electrons in the conduction band. Similarly, there is a small difference of f from unity at E_v. This means that there are holes in the valence band.

Electric conduction in semiconductors

The electric conduction in semiconductors takes place both by electrons in the conduction band and holes in the valence band. As in a metal, the average velocity of the charge carriers is linear in the electric field $\mathcal{E}$. We write this separately for electrons in the conduction band (index n meaning negative charge) and holes in the valence band (index p meaning positive charge),

$$\boldsymbol{v}_n = -eB_n\boldsymbol{\mathcal{E}}, \quad \boldsymbol{v}_p = eB_p\boldsymbol{\mathcal{E}}. \tag{161}$$

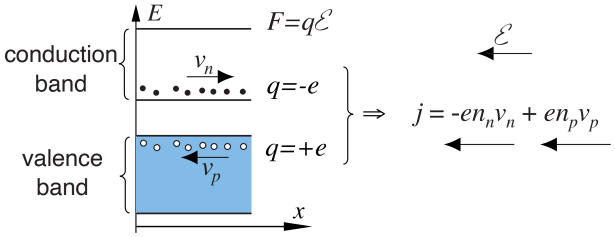
The coefficients B_n and B_p are called *mobilities*. Note that $\boldsymbol{v}_n$ is the same as in Drude model except that now we have written B_n instead of τ/m. The holes have different effective m and τ, and thus their mobility B_p can be different. For the electric current we need the density of electrons in conduction band n_n and holes in the valence band p_p. The current density

$$\boldsymbol{j} = -en_n\boldsymbol{v}_n + en_p\boldsymbol{v}_p = e^2(n_nB_n + n_pB_p)\boldsymbol{\mathcal{E}}. \tag{162}$$

From this we get the conductivity

$$\sigma = e^2(n_n B_n + n_p B_p). \tag{163}$$

We see that the conductivity of the electrons in the conduction band and the holes in the valence band add (not subtract), as was discussed above. This is illustrated in the figure below (where q denotes the charge).

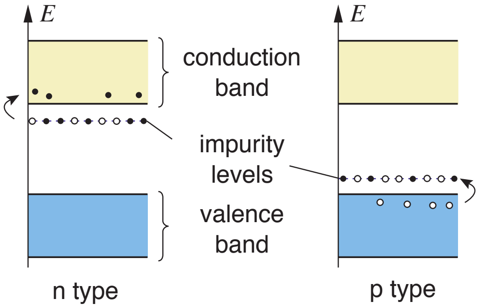


The most essential difference in the conductivity between a metal and a semiconductor is the following. In a metal (151) the density of charge carriers n is the number of electrons in the conduction band. This is constant for each metal. In particular it does not depend on temperature. In a semiconductor (163), the charge carriers densities n_n and n_p are suppressed at low temperatures and grown exponentially with temperature. This effect is generally stronger than any temperature dependence in the mobility $B = \tau/m$. Thus we can now understand that the conductivity of semiconductors is smaller than in metals, and it grows with temperature, in contrast to the case of metals.

Extrinsic semiconductor

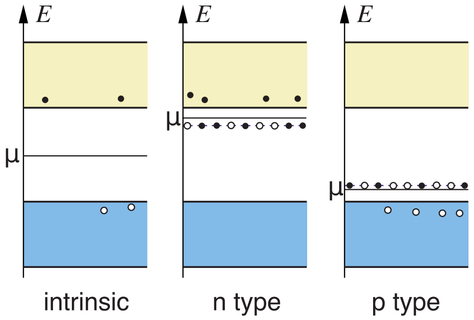
Certain type of impurities can strongly affect the conductivity of semiconductors. Such an impure semiconductor is called *extrinsic*. Let us consider silicon or germanium, which have 4 valence electrons per atom. We replace one atom by phosphorous or arsenic (P, As), which have 5 electrons in the outermost shell. The extra electron can move in the conduction band. However, the impurity atom has a positive charge e, that attracts this electron. As a consequence, the electron can be trapped in the neighborhood of the impurity atom. A simple model to consider this is that the fixed positive charge and the mobile electron form something like a hydrogen atom. The difference is that the electron moves in the background of the semiconductor atoms. This could be taken into account that 1) the mass of the orbiting electron is the effective mass in the conduction band, and 2) the Coulomb potential caused by the positive charge is reduced by the relative dielectric constant of the semiconductor material. Because of these reasons, the bound state energy is only slightly below E_c. (exercise).

Let us think that there are several impurities. The extra electrons of these atoms are placed on levels that are formed slightly below the conduction band. From there they are easily excited to the conduction band, where they can contribute to electric conduction. Such a semiconductor is called negative type, or n type.



Correspondingly there are impurities that have less than 4 electrons in the outermost shell, for example aluminium or gallium (Al, Ga). They make bound hole state slightly above the the valence band. When electrons from the valence band are excited to these levels, mobile holes are formed. Such a semiconductor is called positive type, or p type.

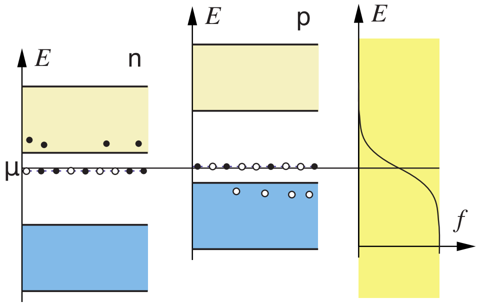
In intrinsic semiconductors the Fermi level is close to middle of the gap. In n-type semiconductor the Fermi level is shifted up to be close to the “donor” impurity levels.



Correspondingly, in p-type semiconductors the Fermi level is shifted down to be close to the “acceptor” impurity levels. Both type of impurities can dramatically increase the number of charge carriers and thus the conductivity of the semiconductor compared to the intrinsic case.

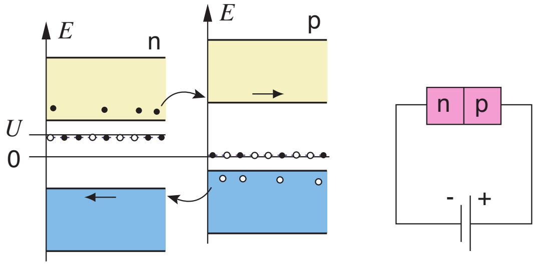
N-p junction

Consider an interface between n and p type semiconductors. Because the Fermi level is higher in the n-type side, electrons start to flow to the p-type side. Therefore a negative charge will form on the p side and positive charge remains on the n side. This increases the relative potential of the p side. An equilibrium is achieved when the Fermi levels on both sides are equal. Thus, the whole system is described by the same Fermi distribution.



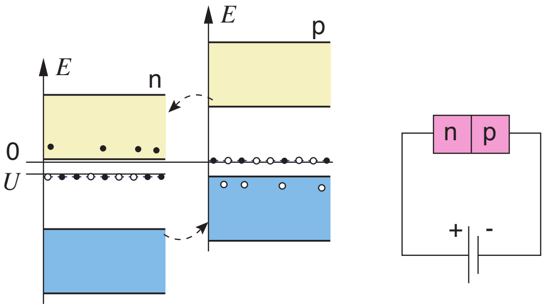
A corresponding potential difference develops between any two conducting materials. It is called *contact potential*.

We next bias the junction by an external voltage V that is negative on the n side. According to the definition of voltage, the energies on the n side are shifted by $U = -eV$. This allows electrons from the n side to flow to the p side. Similarly, holes from p side can flow to n side.



forward bias ($U > 0$)

If the voltage is biased in the reverse direction, one could think that the currents would be reversed. However, there are very few electrons on the p side conduction band and very few holes on the n type valence band. This means that the current in this reverse direction is less than in the case of forward bias.

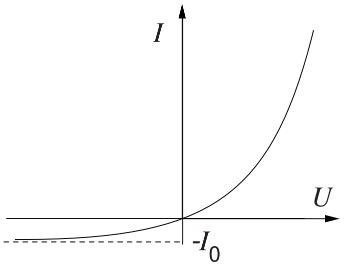


reverse bias ($U < 0$)

We conclude that the junction acts as a current rectifier: the current in one direction is larger than in the opposite direction at equal voltage.

We make a simple model. We only look at electrons, the behavior of holes is analogous. The flow of electrons p→n depends only on the number of electrons in the p-type conduction band. The current $-I_0$ caused by this is independent of the potential U. The flow of electrons n→p depends on the number of those electrons on n side, whose energy exceeds the p-type E_c. Because in this region the Fermi distribution is given by the exponential (exercise), the current caused by these is $I_1(U) = I_1(0)e^{U/k_BT}$. The total current is the sum of the two contributions. Taking into account that in equilibrium $U = 0$, the total current has to vanish, we get

$$I(U) = I_0[e^{U/k_BT} - 1]. \tag{164}$$



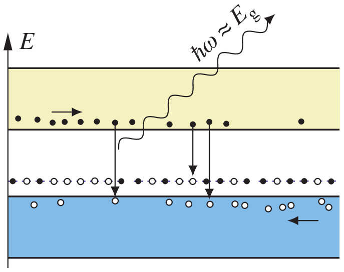
At large negative U there will be electric breakdown (not shown in the figure).

Applications of n-p junction

The n-p junction, diode, is used as rectifier. A rectifier transforms an alternating electric current to an oscillating direct current. A slightly more complicated component with two nearby junctions (for example n-p-n) is called bipolar transistor. It is important component in electronics as it can be used to amplify signals: with a small current to the "base" (the middle semiconductor) one can direct a larger current through both junctions.

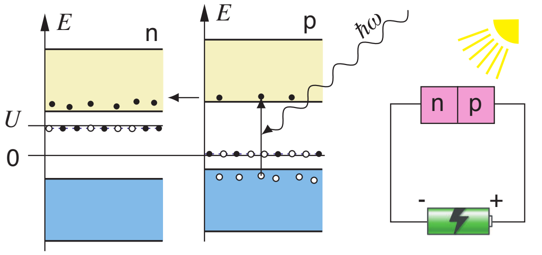
Another type of transistor is a MOS-FET (metal-oxide-semiconductor, field effect transistor). It it widely used in logic circuits. It also can have n-p-n structure, where the conductance through the p layer can be changed by a voltage on a nearby metal electrode (insulated by the oxide).

Let us consider the p side of the diode. In forward bias there will be extra electrons coming from the n side. In equilibrium, the current in p type is by holes. The transformation between these is called *recombination*, where the electrons in the conduction band drop to the holes in the valence band. The energy $\approx E_g$ per electron is released in recombination. This can be emitted as electromagnetic radiation. This is the working principle of *light emitting diodes* (LED) and diode lasers.



recombination in p type

N-p junction is also used in *solar cells*, or more generally in *photovoltaic cells*. The idea is to absorb the electromagnetic radiation and convert it to electric current and voltage. When an insulator or semiconductor is shined with radiation, whose energy quantum $\hbar\omega$ exceeds the energy gap E_g, electrons are excited from the valence to the conduction band. These electrons move randomly in the conduction band before they after some time recombine with with the holes to form radiation again. In case the excitation takes place in the p side of a n-p junction, the electron can drift to the n side. There it cannot recombine any more (because there are no holes) and it cannot easily return to the p side (if the electron energy is relaxed below p-side E_c). This causes a negative charge on the n side, that causes voltage and current in an external circuit.



principle of solar cell

battery is charged

4.6 Transitions

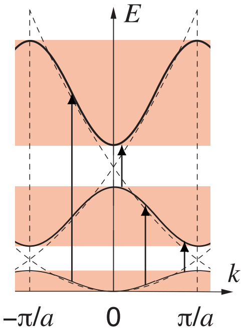
We considered transitions between atomic levels before. We argued that in a transition (absorption or emission) the energy is conserved. That is, the photon energy is equal to the energy different between the two levels (36). The same rule applies also for transitions in solid state. For definiteness, consider absorption and let i denote the initial electron level and f the final level. The energy conservation can then be written

$$E_f(\boldsymbol{k}_f) = E_i(\boldsymbol{k}_i) + \hbar\omega_{\text{ph}}. \tag{165}$$

where ω_{ph} is the photon angular frequency. In a crystal there is additional condition that also the crystal momentum has to be conserved,

$$\boldsymbol{k}_f = \boldsymbol{k}_i + \boldsymbol{k}_{\text{ph}}. \tag{166}$$

The photon wave number $k_{\text{ph}} = |\boldsymbol{k}_{\text{ph}}| = 2\pi/\lambda$. At frequencies of visible light and lower, the wave length λ is much longer than the atomic spacing. Thus the photon momentum is negligible in comparison to the typical crystal momentum $\sim \pi/a$. This means that the transitions are almost vertical in a band diagram.



Some interband transition are depicted in the figure. A transition is possible only if the initial level is filled. Similarly, the final level has to be initially empty since the Pauli principle prohibits transition to a filled level. Below we consider transitions in insulators, semiconductors and metals separately.

Let us consider insulators. In order for the photon to be absorbed, its energy has to be lager than the band gap, $\hbar\omega > E_g$. At lower frequencies the the material is transparent. Visible light corresponds to energies from 1.7 eV (red) to 3.1 eV (violet). From the table on page 38 we see that diamond, table salt (NaCl), corundum (Al_2O_3) and quartz (SiO_2) have energy gap larger than 3.1 eV, and thus they are transparent for visible light. The same hold for many other insulators.

The reason that many insulator materials do not look transparent is that they have impurities or other defects that scatter or absorb light. An impurity can add an electron level within the band gap. Depending on their energy, they can give the material a characteristic color. If some wave length is absorbed effectively, the material shows the complementary color. The figure shows a colorless and two colored quartz crystals.

In addition to electronic transitions, also lattice vibrations cause absorption. This absorption is in the infrared region.

In semiconductors with direct band gap, the absorption takes place when photon frequency exceeds the band gap. The case of indirect band gap is more complicated. Absorption at band gap frequency is still possible in a process which involves also an optical phonon. It takes care that the momentum is conserved, but such a process is not as effective as direct transition. In light emitting diodes, a direct band band gap is preferred.

In metals, the conduction electrons have a strong effect on optical phenomena. Because of them, electromagnetic radiation does not penetrate into the conductor. Thus metals are not transparent. In the visible range, a major part of photons is reflected from the metal surface. This is the reason that metals are shiny, a clean metal surface reflects the incident light. A small part of the incident light is absorbed in interband transitions. For example, the yellowish colors of copper (Cu) and gold (Au) can be understood as absorption of the violet end of spectrum in transitions from a d band below the Fermi level to the conduction band.

An excited electronic state can relax so that a photon of energy $\hbar\omega = E_i - E_f$ is emitted in the transition from state i to a lower energy state f. The state can also relax by other mechanisms, for example, by emission of several photons or phonons. Some examples are the following. *Fluorescence*: emitted radiation with frequency smaller that the absorbed radiation than caused the excited state. Can be understood by relaxation by two photons, out of which only one is observed. Fluorescent substances are used for optical brightening. *Phosphorescence*: delayed emission where the excited electron goes to a long lived state, and relaxes only after longer time by emitting a photon.

4.7 Other electronic properties

Electronic heat capacity

(S 4.2) We studied above the heat capacity caused by lattice vibrations. In addition, also electrons contribute to heat capacity. In principle, the contribution from conduction electrons in a metal could be of the same order of magnitude as from the lattice vibrations. However, at room temperature $k_B T \ll E_F$. Because of this, only small fraction ($\approx k_B T/E_F$) of electrons is excited. This means that the electronic heat capacity is small and in many cases can be neglected.

Heat conductivity

(S 3.2) Heat conductivity κ is defined by formula

$$j_Q = -\kappa \frac{dT}{dx}, \tag{167}$$

where j_Q heat current density (unit W/m^2). This is nonvanishing when the temperature T depends on location x. (The minus sign indicates that the heat current is towards decreasing temperature.)

We mentioned above the heat conductivity caused by lattice vibrations (phonons). It takes place both in insulators and in conductors. In conductors also conduction electrons transport heat. Because of this, the heat conductivity of metals is often higher than that of insulators. Without proper justification, we mention that in many metals the heat conductivity is approximately proportion to the electric conductivity σ (150) so that

$$\kappa \approx \frac{\pi^2 k_{\mathrm{B}}^2 T}{3e^2}\sigma. \tag{168}$$

This dependence is called *Wiedemann-Franz law.* It implies that the scattering that limits the electrical conduction (quantified by the relaxation time τ, page 33), also limits the heat conduction.

Magnetism

(S 19-23) Seems we have to skip this topic.

4.8 Electron-electron interaction

We have studied electron bands in solids. This is based on the independent-electron approximation, which approximates the electron-electron interaction by an effective potential. The theory seems to work well in many cases discussed. The reason is partly explained by *screening* of the Coulomb interaction. This means that a charge placed in a material causes the electrons to adjust in such a way that the electrostatic potential is exponentially weaker ($\propto e^{-\alpha r}/r$) compared to unscreened Coulomb potential ($\propto 1/r$). This makes the potential of short range and thus not that serious a problem.

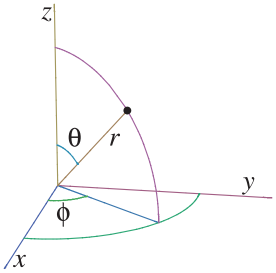
The independent-electron approximation allows to understand many properties of solids, as discussed in this course. Exceptions to this are the following topics: a) atoms, b) covalent bond, c) superconductivity, and d) ferromagnetism.

5. Conclusion

I hope you have enjoyed this course. See the web page of the course for some information about the exam.

Appendixes

A. Spherical coordinates



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

$$\boldsymbol{r} = r\sin\theta(\cos\phi\,\hat{\boldsymbol{x}} + \sin\phi\,\hat{\boldsymbol{y}}) + r\cos\theta\,\hat{\boldsymbol{z}}. \tag{169}$$

Alternatively

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta. \tag{170}$$

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

$$\begin{aligned} \hat{\boldsymbol{r}} &= \frac{\partial \boldsymbol{r}}{\partial r} = \sin\theta\cos\phi\,\hat{\boldsymbol{x}} + \sin\theta\sin\phi\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}} \\ \hat{\boldsymbol{\theta}} &= \frac{1}{r}\frac{\partial \boldsymbol{r}}{\partial \theta} = \cos\theta\cos\phi\,\hat{\boldsymbol{x}} + \cos\theta\sin\phi\,\hat{\boldsymbol{y}} - \sin\theta\,\hat{\boldsymbol{z}} \\ \hat{\boldsymbol{\phi}} &= \frac{1}{r\sin\theta}\frac{\partial \boldsymbol{r}}{\partial \phi} = -\sin\phi\,\hat{\boldsymbol{x}} + \cos\phi\,\hat{\boldsymbol{y}}. \end{aligned} \tag{171}$$

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

$$\boldsymbol{\nabla} = \hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}. \tag{172}$$

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

$$\boldsymbol{\nabla}\Phi = \hat{\boldsymbol{x}}\frac{\partial \Phi}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial \Phi}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial \Phi}{\partial z}. \tag{173}$$

We now claim that in spherical coordinates

$$\boldsymbol{\nabla} = \hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}. \tag{174}$$

We can recognize here the same factors as in the definition of the basis vectors (171). The gradient of a scalar function $\Phi(r,\theta,\phi)$ is

$$\boldsymbol{\nabla}\Phi = \hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}. \tag{175}$$

The Laplace operator is defined as $\nabla^2 = \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}$. In cartesian coordinates we get from (172)

$$\begin{aligned}
\nabla^2\Phi &= \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\Phi \\
&= \left(\hat{\boldsymbol{x}}\frac{\partial}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial}{\partial z}\right)\cdot\left(\hat{\boldsymbol{x}}\frac{\partial \Phi}{\partial x} + \hat{\boldsymbol{y}}\frac{\partial \Phi}{\partial y} + \hat{\boldsymbol{z}}\frac{\partial \Phi}{\partial z}\right) \\
&= \hat{\boldsymbol{x}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial x^2} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial x\partial y} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial x\partial z}\right) \\
&\quad + \hat{\boldsymbol{y}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial y\partial x} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial y^2} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial y\partial z}\right) \\
&\quad + \hat{\boldsymbol{z}}\cdot\left(\hat{\boldsymbol{x}}\frac{\partial^2 \Phi}{\partial z\partial x} + \hat{\boldsymbol{y}}\frac{\partial^2 \Phi}{\partial z\partial y} + \hat{\boldsymbol{z}}\frac{\partial^2 \Phi}{\partial z^2}\right) \\
&= \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}
\end{aligned} \tag{176}$$

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

$$\begin{aligned}
\nabla^2\Phi &= \boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\Phi \\
&= \left(\hat{\boldsymbol{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\right) \\
&\quad \cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&= \hat{\boldsymbol{r}}\frac{\partial}{\partial r}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&\quad + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial \theta}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right) \\
&\quad + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\cdot\left(\hat{\boldsymbol{r}}\frac{\partial \Phi}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial \Phi}{\partial \theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial \Phi}{\partial \phi}\right)
\end{aligned} \tag{177}$$

The added complication here compared to the cartesian case (176) is the extra factors $1/r$ and $1/r\sin\theta$ but also that the unit vectors (171) depend on θ and ϕ. We calculate

$$\begin{aligned}
&\frac{\partial \hat{\boldsymbol{r}}}{\partial r} = 0, \quad \frac{\partial \hat{\boldsymbol{r}}}{\partial \theta} = \hat{\boldsymbol{\theta}}, \quad \frac{\partial \hat{\boldsymbol{r}}}{\partial \phi} = \sin\theta\hat{\boldsymbol{\phi}}, \\
&\frac{\partial \hat{\boldsymbol{\theta}}}{\partial r} = 0, \quad \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \theta} = -\hat{\boldsymbol{r}}, \quad \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \phi} = \cos\theta\hat{\boldsymbol{\phi}}, \\
&\frac{\partial \hat{\boldsymbol{\phi}}}{\partial r} = 0, \quad \frac{\partial \hat{\boldsymbol{\phi}}}{\partial \theta} = 0, \quad \frac{\partial \hat{\boldsymbol{\phi}}}{\partial \phi} = -(\sin\theta\hat{\boldsymbol{r}} + \cos\theta\hat{\boldsymbol{\theta}}).
\end{aligned} \tag{178}$$

In the same order as in (177), we get 6 nonvanishing terms

$$\begin{aligned}
\nabla^2\Phi &= \frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \theta^2} \\
&\quad + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{\cos\theta}{r^2\sin\theta}\frac{\partial \Phi}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 \Phi}{\partial \phi^2}.
\end{aligned} \tag{179}$$

Verify that this is equivalent to

$$\begin{aligned}
\nabla^2\Phi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial \Phi}{\partial \theta}\right) \\
&\quad + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 \Phi}{\partial \phi^2}.
\end{aligned} \tag{180}$$

B. Reduced mass

Let us consider two particles. We assume that the potential depends only on their relative distance $\boldsymbol{r}_1 - \boldsymbol{r}_2$: The energy of the system is

$$E = \frac{1}{2}m_1\dot{\boldsymbol{r}}_1^2 + \frac{1}{2}m_2\dot{\boldsymbol{r}}_2^2 + U(\boldsymbol{r}_1 - \boldsymbol{r}_2), \tag{181}$$

where the dot denotes time derivative. We define vectors for the center of mass and the relative vector,

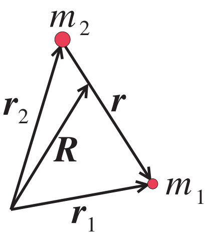
$$\boldsymbol{R} = \frac{m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2}{m_1 + m_2} \tag{182}$$

$$\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2. \tag{183}$$

Solving these for $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ gives

$$\boldsymbol{r}_1 = \boldsymbol{R} + \frac{m_2}{m_1 + m_2}\boldsymbol{r} \tag{184}$$

$$\boldsymbol{r}_2 = \boldsymbol{R} - \frac{m_1}{m_1 + m_2}\boldsymbol{r}. \tag{185}$$



Taking the time derivative and subtituting to (181) gives

$$E = \frac{1}{2}(m_1 + m_2)\dot{\boldsymbol{R}}^2 + \frac{1}{2}\frac{m_1 m_2}{m_1 + m_2}\dot{\boldsymbol{r}}^2 + U(\boldsymbol{r}). \tag{186}$$

The first term is the kinetic energy coming from the motion of the center of mass. The two last terms describe the relative motion, where the mass appearing in the kinetic energy is the reduced mass

$$\frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2} \quad \Leftrightarrow \quad m = \frac{m_1 m_2}{m_1 + m_2}. \tag{187}$$

This makes it plausible that the two-body problem is reduced to a single-body problem with a fixed potential and the reduced mass. (The definite proof would make use of the Lagrangian instead of E, but would otherwise be the same as above.)

C. Group velocity

Consider the wave

$$\xi(x,t) = A\sin(kx - \omega t) = A\sin(k(x - \frac{\omega}{k}t)). \tag{188}$$

Drawing it at different times, we can see that it propagates at velocity

$$v_{\text{phase}} = \frac{\omega}{k}, \tag{189}$$

which is called the *phase velocity.* In many cases the angular frequency ω depends linearly on the wave number, $\omega = \text{constant} \times k$. In this case $v_{\text{phase}} = \text{constant}$. That is, all waves move with the same velocity (independently on the frequency or the wave length).

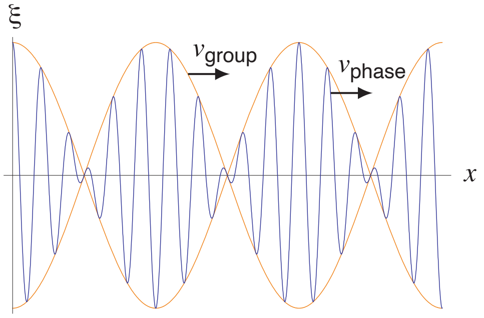
The alternative case is *dispersion.* In this case the dependence $\omega(k)$ is a more complicated than a linear function. Because the phase velocity (189) is not a constant, waves of different frequency move at different velocities. We study a superposition of two waves, whose wave numbers $k \pm \delta k$ are close to each other.

$$\xi = A\cos[(k+\delta k)x-(\omega+\delta\omega)t]+A\cos[(k-\delta k)x-(\omega-\delta\omega)t]. \tag{190}$$

Using trigonometric formulas we can write this in the form

$$\xi = 2A\cos(kx-\omega t)\cos(\delta kx-\delta\omega t). \tag{191}$$

Here the amplitude of the wave of short wave length ($\lambda = 2\pi/k$, the former cosine factor) is modulated by a wave of long wave length ($\lambda = 2\pi/\delta k$, the latter cosine factor). It represents a train of wave packets.



The modulating wave now gives the velocity of the wave packet,

$$v_{\text{group}} = \frac{\delta\omega}{\delta k} \approx \frac{d\omega}{dk}. \tag{192}$$

which is called the *group velocity.*