Modeling and simulation methods in nanotechnology

Short description:

This series of lectures contains introduction to computational methods used in physics and biophysics for modeling and simulations of nanostructures. The selected examples are the following: tunneling through a potential barrier, modeling of the energy structure of graphene systems, optimalisation of strain effects in nanoscale, and molecular dynamics of nano-size biological systems. Description of the problems, is proceded by a short introduction to quantum mechanics. The presented methods are related to the numerical techniques studied during the course of numerical methods at the second grade. Here, students create (write) their own numerical codes, which solve the problems presented at the lectures or learn how to operate ready-to-use numerical packages.

Course 1. 3 hours

Introduction to elementary quantum mechanics

Short history of quantum physics

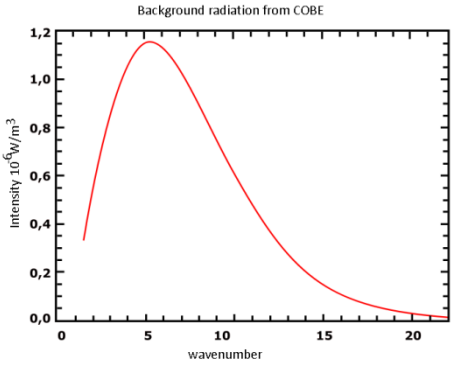
Classical physics describes macroscopic systems. The term "macro" concerns usually sizes above 1 micrometer. As classical systems we can consider individual objects having macroscopic masses and sizes (usually modeled as material points) or sets of such objects (e.g. planetary systems). The aim of theory is the description of the *state of a physical system.* In classical mechanics the state of a system built of many material points is uniquely determined by positions and momenta of all the elements of the system in any instant of time. The values of positions and momenta are found by solving equations of Newton's mechanics and dynamics, which in fact are second order differential equations. Classical state of a systems built of a very large number of weekly interacting particles is described within the framework of statistical physics (forgetting about their internal structure) using some macroscopic quantities as temperature, volume and pressure.

At the end of the XIX century there was a growth of experimental techniques both in the spectroscopy and in the electric research. The results of several experiments couldn't be explained on the basis of classical physics without introducing new theory describing microscopic nature of matter, i.e., nature of atoms and molecules. Some of these experiments and phenomena were the following: *black-body radiation, photoelectric effect and Franck-Hertz experiment*.

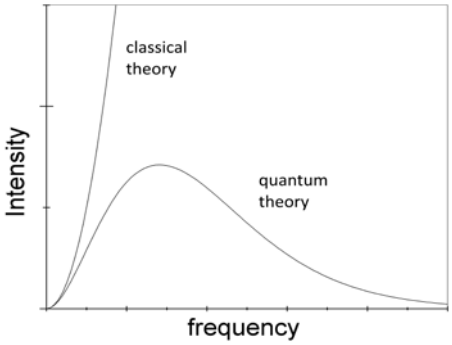
Black-body radiation.

A black body is a physical object that absorbs all incident electromagnetic radiation, to which it is exposed. Within the framework of classical physics the radiation of a black-body was described and governed by the Rayleigh-Jeans low, which says that the radiated energy should be proportional to the fourth power of radiation frequency, independently on temperature. However, the experiments proved that there is a maximum of radiation and it depends on temperature. This was known as

Wien's displacement low (Fig. 1). For example, radiation maximum from the Sun surface appears at 480 nm. Moreover, summing the Rayleigh-Jeans formula over all the frequencies we would get infinite energy. This is called *ultraviolet catastrophe* and it was the main reason for looking for a new physics describing micro world.



a)



b)

Fig. 1 a) Background radiation as the rest from the Big-Bang. From Wikipedia http://pl.wikipedia.org/wiki/Cia%C5%82o_doskonale_czarne *. b) Comparison of intensity of radiation of a black-body well described by Planck theory (curve with maximum) with classical prediction (monotonic curve).*

Oscillating charged particles are the source of radiation in classical physics (e.g., electrons in atoms). To explain discrepancies between the observations and Rayleigh-Jeans formula, Max Planck assumed in 1900 that the oscillating particles can emit radiation only in portions of energy called *quanta*, $E=h\nu$, where ν – radiation frequency ($\nu=c/\lambda$, λ - wavelengths, c – speed of light), and h is a constant having value $4{,}135\ 667\ 516 \cdot 10^{-15}$ eV. This totally revolutionized our view on the structure of matter, since it meant that the energy structure of atoms (of the elementary pieces of matter) is discrete, i.e. quantized, what means that only some definite values of the energy are admitted at this level.

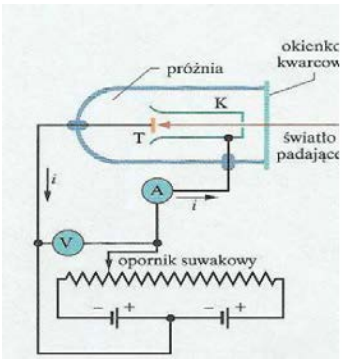
Photoelectric effect.

Photoelectric effect was discovered by Heinrich Hertz in 1887. Later, it was studied in detail by Philipp von Lenard and finally in 1905 explained by Albert Einstein based on quantum theory proposed by Planck.

In a vacuum tube containing two electrodes (Fig.2a), light beam (electromagnetic waves) is directed to a target metallic electrode (T). The electrons are ejected from metal and the applied potential (voltage) can accelerate or slow down the electrons on their way to cathode (K). For a given voltage V the electrons can be even turn back to the emitter (T). The energy of the turned back electrons equals $E=eVh$ (h being a path of electrons) and doesn't depend on the light intensity (Fig. 2b). The energy E depends linearly on frequency ν with some threshold Φ called *work function*. Einstein thus concluded that

$$h\nu = E + \Phi,$$

Φ is the energy needed to eject electrons from metal (is different for different materials).

a) 

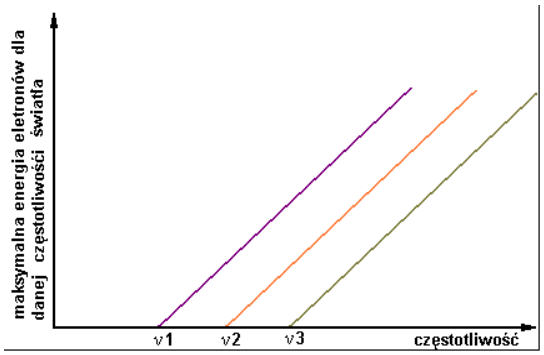
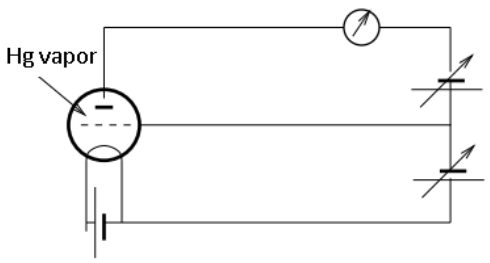
b) 

Fig. 2 a) Scheme of a set-up to observe photoelectric effect. b) Energies of emitted electrons vs. frequency of radiation for different metals of target (T). From http://library.thinkquest.org/28383/nowe_teksty/html/2_8.html

Franck-Hertz experiment.

The experiment involves a tube containing low pressure mercury vapor and three electrodes (Fig.3). The electrons emitted from cathode are accelerated by a mesh-grid electrode situated between cathode and anode. One measures the current between the electrodes vs. potential difference applied to a mesh-grid electrode. The electrons collide with the mercury atoms, for low voltage their kinetic energies are small and the collisions are elastic, while for higher voltages the scattering is inelastic and the electrons give part of their energy to excite mercury atoms. Every 4.88 eV the current drops dramatically down, what means that the energy needed to excite mercury atoms from the ground to the first excited state is just 4.88 eV; the conclusion: the energy structure of atom is discrete (quantized)

a) 

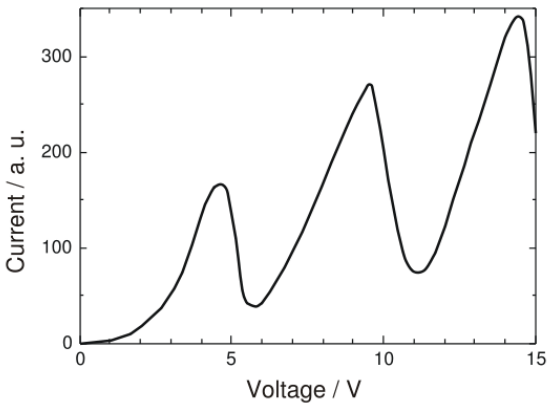
b) 

Fig.3 a) Scheme of a setup to observe Franck-Hertz experiment. b) Anode current versus voltage ad mesh-grid electrode. From http://en.wikipedia.org/wiki/Franck%E2%80%93Hertz_experiment

Diffraction of waves and electrons. Waves of matter.

The experiments described above reveal quantum nature of matter, at least as we speak about the energy structure of atoms and radiation processes (absorption and emission). However, another experiments show yet another striking property of matter: elementary pieces of matter, like e.g. electrons behave sometimes as waves.

The most fundamental feature of waves is their diffraction (and interference) on slits and edges. A typical picture of the intensity of light (observed at screen) after passing by two closely spaced slits (the width of slits compared with the wavelengths) is shown in Fig. 4. After constructive interference of spherical waves generated at slits we can observe bright and dark fringes at the screen (if the slits are point-like, the fringes are concentric). The pattern of fringes appears also for diffraction gratings. When the wavelengths is very small and it is difficult to perform a proper gratings (e.g. for X-rays), the waves can be diffracted on crystals, where the role of gratings is played by equally-spaced crystallographic planes built of regular mesh of atoms.

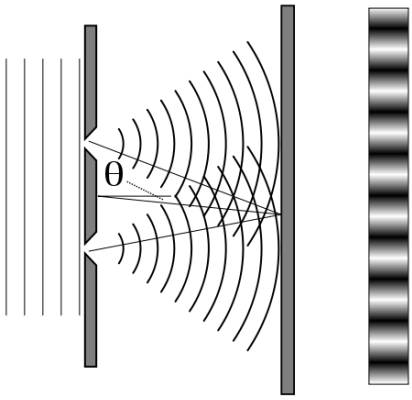


Fig. 4. Diffraction of a plane wave on a double slit and appearance of fringes on the screen. From http://en.wikipedia.org/wiki/Double-slit_experiment

However, a very similar picture one gets using electrons instead of X-rays. It is shown in Fig. 5.

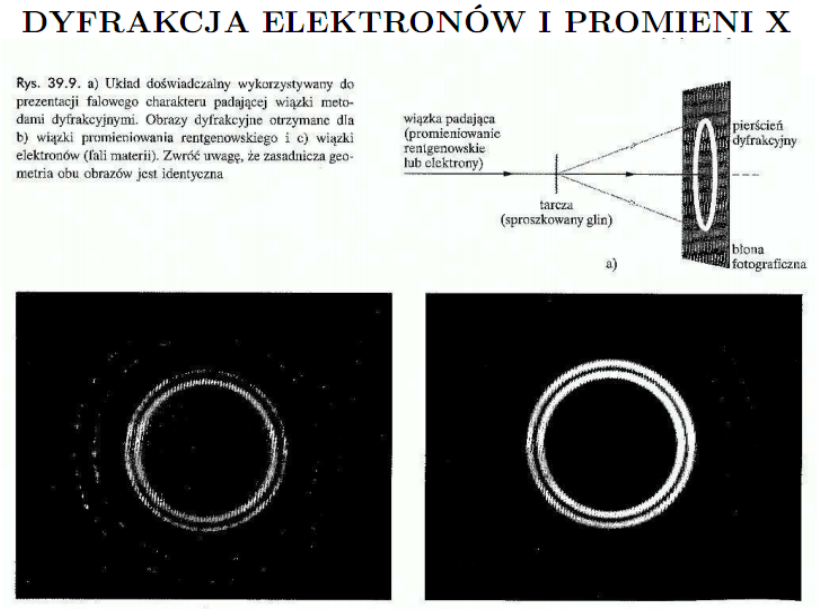


Fig.5 Comparison of diffraction of X-rays and electrons on Al crystal. From "Podstawy Fizyki", D. Halliday, R. Resnick, and J. Walker, PWN Warszawa 2007.

In the experiment with electrons, classical physics suggests that the particles should pass through one or another slit yielding for a dense beam of electrons a direct picture (a kind of reflection) of two slits (each electron "switches on" one pixel at the scintillating screen). However, the result of

experiment is completely different (see Fig. 6). Although each electron reaches some place at screen, the picture obtain after many, many electrons passes through the double-slit, is the same as for the waves!

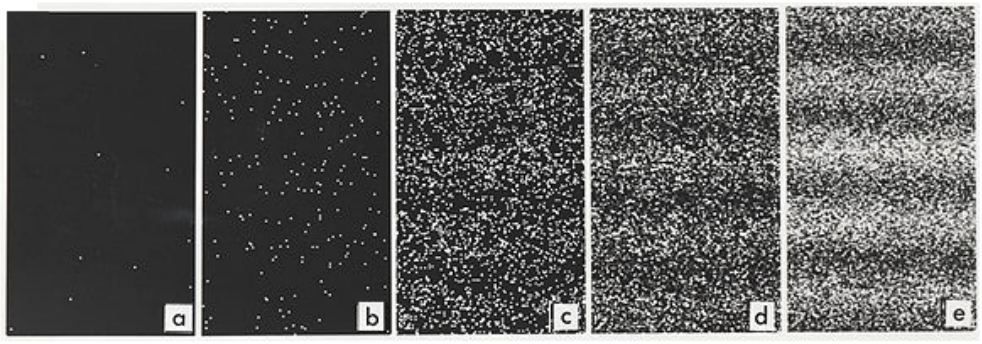


Fig. 6. Interference pattern obtained on luminescence screen in the experiment of electron beam passing by double slit. Pictures from a to d correspond to growing number of diffracted electrons. From http://en.wikipedia.org/wiki/Double-slit_experiment

In result, it means that we are not able to trace the electron on its way from the slit to the screen, i.e. we cannot know its trajectory – we cannot know its position and momentum in a given instant of time. It is not our disability, it is a most fundamental property of nature, which appears to be non-deterministic. This property is reflected in quantum mechanics by the so called *uncertainty principle*, discovered by Heisenberg. Bright and dark fringes (areas) at screen shown in Fig. 6, reveal that we can only use probabilistic theory to determine the state of electron. Any new physics, i.e. quantum physics, must therefore be a probabilistic theory.

Some fundamental axioms of quantum mechanics.

Axiom 1.

The state of a physical system is uniquely and fully described in quantum mechanics by a mathematical function called *wave function* (state function or state vector) $\Psi(\mathbf{x},t)$ where, $\mathbf{x}$ represents space-position variables of all the particles contained in the system and t represents time**.** The wave function provides information about the probability amplitude of position (or momentum). Its physical interpretation is the following:

$$\Psi^{*}\Psi d\mathbf{x} = |\Psi|^{2} d\mathbf{x},$$

gives probability of finding the state in a volume d**x,** while $|\Psi|^2$ is the probability density. Therefore, the integral of $|\Psi|^2$ over the entire space gives 1, since a given system is always somewhere in the space. We say that the wave function must be “square-integrable”.

Axiom 2.

Any measurable physical quantity A (that can depend on x and t) is represented by a hermitian operator **A**. The operators act on the wave functions, as the result we get another function (in some particular cases the same function) $\mathbf{A}\Psi=\Phi$**.** For example, taking derivative of a function can be considered as action of some operator on this function, or a constant that multiplies a function is the

simplest example of an operator. If $\mathbf{A}\Psi=a\Psi$, (a – a number), then Ψ is the so called eigenfunction of the operator **A** and a is called its eigenvalue.

Axiom 3.

The results of measurements of a quantity A are only eigenvalues of the corresponding operator **A** . Note that,

All the eigenfunctions of any operator (let us call them φ_i , i=1,2,…) span the basis in the functional space of square-integrable functions of a given type (given variables). The corresponding eigenvalues are a_i . Therefore, any function describing a state of a given system can be represented in this basis (as a linear combination)

$$\Psi = \sum_i c_i \varphi_i$$

Please note analogy to vector space. As in a vector space, one can define scalar product of φ_i and φ_j as

$$< \varphi_i \mid \varphi_j >= \int \varphi_i^* \varphi_j d\tau$$

where τ - represents all the space coordinates. The symbol <|> is the so called Dirac notation for scalar product.

A function is called normalized if $< \varphi_i | \varphi_i \geq 1$; two functions are orthogonal if their scalar product vanishes. If $\Phi=H\Psi$, then

$$< \varphi \mid \Phi >=< \varphi \mid H\psi >= \text{ozn} =< \varphi \mid H \mid \psi > .$$

Axiom 4.

If a given system is in the state described by the wave function Ψ, the probability of getting value a_i while measuring a quantity A equals:

$$P_i =|< \varphi_i \mid \Psi >|^2 = \left| \int \varphi_i^* \Psi d\tau \right|^2$$

Additionally,

A) average value of A obtained after a large number of measurements equals:

$$\overline{A} =< \Psi \mid A\Psi >= \int \Psi^* A\Psi d\tau$$

and is called *expectation value* of **A** (**A** being the corresponding operator).

B) operator of momentum $\vec{p} = (p_x, p_y, p_z)$ has also vector character. Its components are $p_x = -i\hbar \frac{\partial}{\partial x}$ and similarly for p_y and p_z .

Therefore, the kinetic energy operator reads:

$$T = \frac{mv^2}{2} = \frac{p^2}{2m} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \tag{1}$$

Axiom 5.

The wave functions are solutions of the Schrödinger equation

$$i\mathrm{h}\frac{\partial \Psi(\mathbf{x},t)}{\partial t} = H\Psi(\mathbf{x},t)$$

Where for a system composed of n particles, $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ and each $\mathbf{x_i} = (x_i, y_i, z_i)$.
The wave function fulfils also

$$H\Psi(\mathbf{x},t) = E\Psi(\mathbf{x},t) \tag{2}$$

known also as Schrödinger equation, where **H** is operator of the total energy of the system called also Hamiltonian. **H = T + V**, where **V** is the potential energy operator.
If V is time-independent the eigenstates of **H** are called *stationary* states**.** The Schrödinger equation (2) is in fact a second order differential equation, to solve it one has to impose boundary conditions on ψ , which cause that the solution of (2) exists usually for a discrete set of

$$E_i, \quad \Psi_i$$

Solution of the Schrödinger equation is the main aim of every quantum problem. Solving the equation (2) we find wave functions representing possible states of a given physical system. Therefore we obtain full information about the system. Solving the time-dependent Schrödinger equation we get information on the dynamics of a given system and we can foresee how it will behave in time.

Simple examples of physical systems

Very limited number of model physical systems can be solved exactly. In this lecture we will show only few of them. For all the other cases representing more complex physical systems, the equation (2) must be solved in an approximated way.

Particle in one-dimensional potential well (particle in a box).

Let us consider a particle of mass *m*, which can freely move in one direction only (let say x-direction) on the limited section (segment) of the lengths L. If independently on the particle energy it cannot leave this segment, in quantum mechanics its potential energy is represented by two infinitely high barriers set at the edges of the segment L (the so called *impenetrable* walls). If the particle could leave this segment for the energy V, we would call such a problem “a particle in a potential well of length L and finite potential barrier V”. Both cases are represented in Fig. 7.

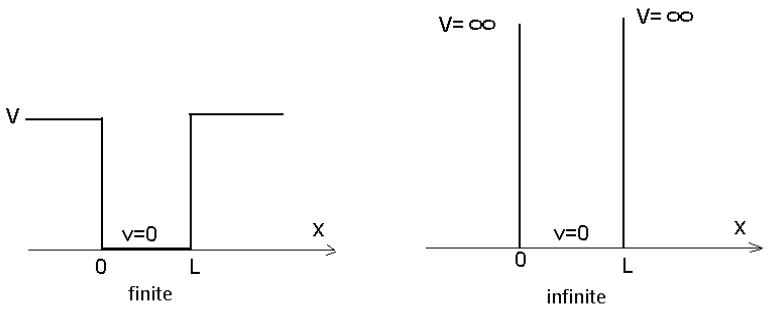


Fig. 7. Finite(left) and infinite (right) one-dimensional potential well of width L.

In both cases the Schrödinger equation depends only on one variable x. As differential equation it requires imposition of two boundary conditions to get its unique solution. In case of infinite potential barrier we have to require that the wave function vanishes at x=0 and x=L and also beyond the section [0,L] $\Psi(0) = \Psi(L) = 0$. It simply means that the particle cannot leave this segment. These conditions yield following eigenvalues:

$$E_n = \frac{h^2 n^2}{8mL^2}, \tag{3}$$

where *n* – a quantum number counting (numbering) the consecutive solutions. The corresponding wave functions are:

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

This is schematically shown in Fig. 8. It is very instructive to have a look at the problem of scale. For large, macroscopic sizes L and macroscopic masses, one gets in (3) a dense spectrum of eigenenergies. For example, for L=1 meter and *m*=1kG, the difference E_2-E_1 is about 10^{-66} J (10^{-47} eV) and is absolutely non-measurable, while for micro world like for electron in the 1nm trap, this energy difference is of the order of 1 eV and is measurable.

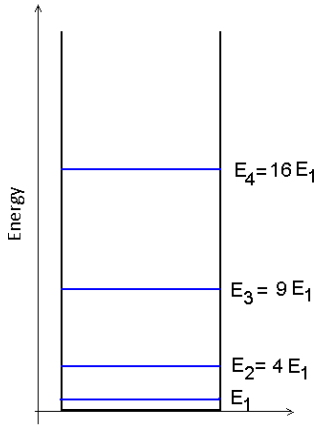


Fig. 8. Energy levels in one-dimensional infinite potential well.

Let us note, that transitions between the states, e.g. excitations from the ground state to higher energy levels (via photons) are possible only for photon energies equal:

$$h\nu = E_n - E_m$$

Such process is called absorption, and the reverse process is known as emission. Absorption and emission spectra will have discrete character (see Franck-Hertz experiment, Fig. 3).

Square modulus of ψ_n (probability density) are shown in Fig. 9, they remind "waves".

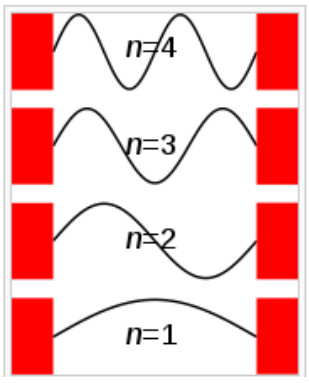


Fig. 9. Wave functions for several states in one-dimensional infinite quantum well. From http://en.wikipedia.org/wiki/Particle_in_a_box

Symmetry.

A physical system can be invariant under some symmetry operations (transformations). For example, two-atomic molecule like H_2 remains unchanged if we rotate it around the bond axis or if it is reflected in a plane perpendicular to the bond and placed half-way between both atoms. Infinite crystal, which is periodic system remains unchanged under *translations* by a lattice vector. If a system is invariant under some symmetry operation then the corresponding Hamiltonian is also invariant and there exists an operator representing this symmetry operation, which commutes with the Hamiltonian, **[H,S]=0**, i.e., the order of the product of both operators acting on any wave function is unimportant, **HS**ψ=**SH**ψ. It is easy to show that commuting operators have the same eigenfunctions. Therefore, if a physical system is somehow symmetric, the corresponding wave functions will need to reflects all symmetries and should be enumerated with the quantum numbers connected with eigenvalues of **S**.

Course 2. 3 hours.

Atoms, molecules and elements of the solid state theory

Hydrogen atom.

Hydrogen atom is the simplest system composed of two interacting charged particles: proton and electron. Since proton is about 1000 times heavier than electron, we can reasonably assume that it is much more inert than the electron, so we can treat proton as fixed in the coordinate system (at position 0). Therefore, hydrogen atom reduces to electron moving in the static Coulomb field of proton. Let us note that in classical physics the electron would simply fallen down to proton, so the Hydrogen atom could not exist…

Potential energy of electron in the electrostatic Coulomb field of proton is

$$V(r) = -\frac{e}{r} \qquad (4)$$

where e is the electron charge and r is its distance to nucleus. The energy is negative, what means that it can be balanced by the positive kinetic energy of the electron. The Hamiltonian consists of **T** (equation 1) and **V** (4). The solution of

$$H\Psi(\vec{r},t) = E\Psi(\vec{r},t)$$

leads to a discrete *spectrum* of eigenvalues (possible energies of electron in Hydrogen atom):

$$E_n = -\frac{me^4}{8h^2\varepsilon_0^2}\frac{1}{n^2} \qquad (5a)$$

In (5a) n is the principal (main) quantum number, enumerating energy levels of H. These energies are negative, which means that the electron is bound in the proton field. To realize it (to ionize the atom) one have to add (supply) positive energy equal $|E_1|$. These energy equals 13.6 eV and is very well measurable.

Let us note, that it is again a problem of a particle in a potential well, although now the well (4) is three-dimensional, spherical and has a "shape" of infinitely deep "funnel". Its shape is shown in Fig. 10, where some energy levels are also marked.

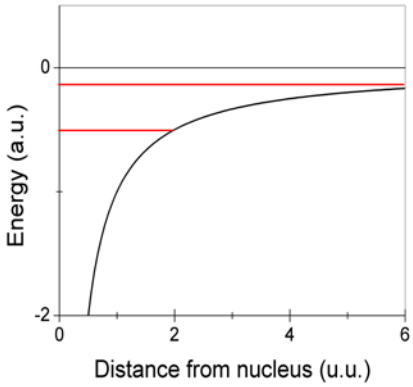


Fig. 10. Potential energy (black) and two lowest (red) energy levels in Hydrogen atom (in atomic units, i.e. putting m=h=e=1).

In general, when quantum well is two- or three-dimensional, the problem of degeneracy of energy levels occur. Degeneracy of energy levels appears when there are more than one eigenfunction for the same eigenvalue. It occurs also in case of the Hydrogen atom. The lowest energy level is not degenerate (not counting the electron spin). The first excited level is four-time degenerate, the second excited level is nine-times degenerate. Degeneracy appears when the system is invariant under some symmetry operations. Hydrogen atom has extremely high symmetry: it is invariant under any rotation around any axis passing by point-like proton. This is so called spherical symmetry and the resulting degeneracy is called *orbital*. This degeneracy is enumerated by the so called orbital (l) and magnetic(m) quantum numbers. The energy level n and all the corresponding eigenfunctions are called *shell*. In a given shell there are n different types of functions called *orbitals*, each one additionally degenerate 1x, 3x, 5x,… These different types we labels by letters *s, p, d,…* and call them *subshells*. All the orbitals (wave functions) depend on x,y,z, or on *r,* θ and ϕ if we work in spherical coordinates (Fig. 10a) are the following

$$\Psi(x,y,z) = \Psi(r,\theta,\varphi) = R_n(r)Y_{l,m}(\theta,\varphi) = \Psi_{n,l,m}\,, \qquad (5b)$$

where $Y_{l,m}$ – the so called spherical harmonics. n is the main quantum number, l – orbital quantum number and m – magnetic one. The energies depend only on n.

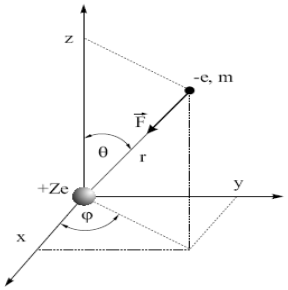


Fig. 10b Spherical coordinates.

Radial dependence (on r) of the orbitals is the same in each type. Square modulus of the orbital integrated over angular variables gives *radial distribution of the density* in a given state. For the first three shells they are shown in Fig. 11.a. Maximum of a given distribution marks distance from the nucleus at which the probability of finding electron in a given state is the highest. Surfaces at which total densities have constant values are called *isosurfaces*. For several orbitals they are shown in Fig. 11b. Please note, that due to orbital and magnetic degeneracy, any combination of functions with different l and m for a given n is also a good wave function. Therefore, the shapes shown in Fig. 11.b are "possible" buy not "unique".

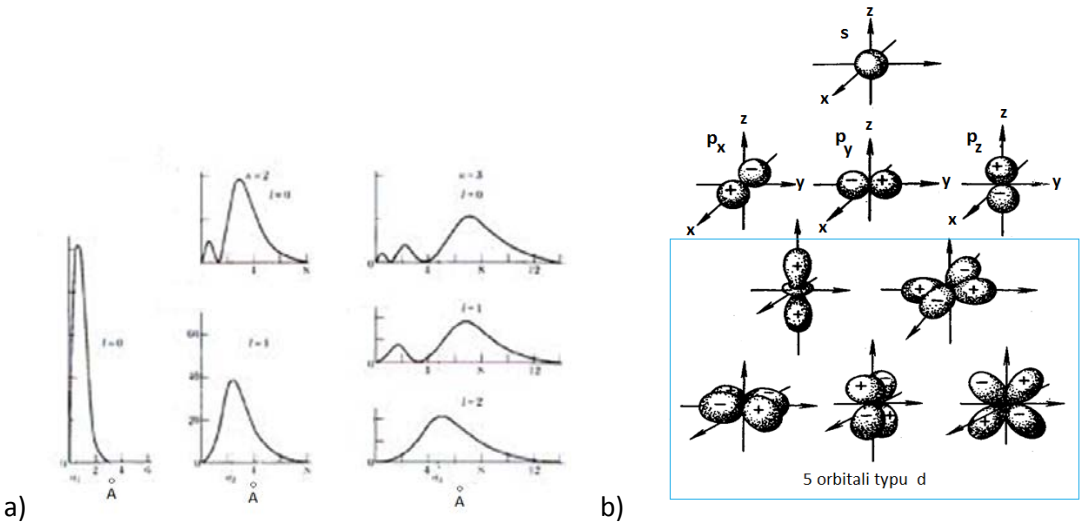


Fig. 11 Radial (a) and angular (b) density distributions for several lowest orbitals in Hydrogen atom. From any book of quantum chemistry or quantum mechanics.

In many electron atoms Coulomb interaction lifts the orbital degeneracy, now the energies depend on n and l, but still do not depend on m. Each energy level (given nl) is $2l$-1 times degenerated. The sequence of the energy levels is at the beginning 1s, 2s, 2p, 3s, 3p, 3d, … but changes for higher n. Due to spin of electron and the Pauli principle each atomic orbital can be occupied by two electrons (only). When several subshells are fully occupied the atoms (o rather the state) is called *closed-shell*. Nobel gasses are the examples. Such atoms are not eager to react with other atoms. Contrary, the *open-shell* systems "prefer" to "give" or to "receive" electrons, trying to close the electronic shells

and subshells, since it is energetically favorable. Such atoms react easily with the others forming molecules and crystals.

Hybridized orbitals and chemical bonds

Chemical bond.

Although atoms are electrically neutral they can bind forming molecules and crystals. A fixed junction between two or more atoms is called a *chemical bond*. There are different types of bonds. When one atom lacks one electron to close the external subshell, e.g. Cl ($1s^2 2s^2 2p^2 3s^2 3p^5$) and the other have only one electron at the outermost subshell, like Na ($1s^2 2s^2 2p^2 3s^1$), sodium will eagerly give one electron to chlorine, which eagerly will receive it. Two charged ions will appear, which will attract electrostaticly each other forming *ionic bond*. Another type of bond appears in metals: quasi-free electrons form negatively charged cloud, which keeps stable lattice of positively charged ions. The molecular or van der Waals bond occur between atoms of noble gases or large molecules (e.g. proteins), but also between layers of graphene in graphite. However, we will be mostly interested in *covalent* bond, which happens when hybridized orbitals of neighboring atoms strongly overlap.

Hybridized orbitals.

As mentioned above, one can create linear combinations of functions belonging to degenerate energy level. Each such combination is also a good wave function, i.e. an orbital having only different density distribution than the constituent functions. For the energy levels that are not exactly degenerated, but are energetically close we can assume they are degenerated and make linear combinations of the corresponding wave functions. Such situation takes place, for example in carbon C ($1s^2 2s^2 2p^4$), where the energies of 2s and 2p levels are close (at least comparing with 1s). Of 2s and $2p_x$ orbitals we can create two new hybridized orbitals sp. One as $2s+2p_x$ and the other as $2s-2p_x$. Since the orbital $2p_x$ has density localized along the 0X axis and the wavefunction is positive for x>0 (negative for x<0) both hybridized orbitals will be "directed" one along X+, the other along X-. It is illustrated in Fig. 12.

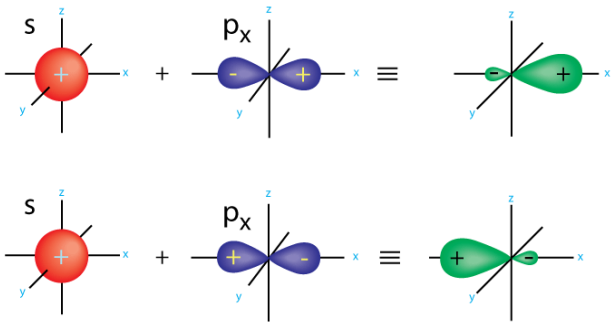


Fig. 12. Formation of sp hybridized orbitals. From http://www.grandinetti.org/comment/18

Such orbitals explain stability of bond in F_2 molecule, what is shown in Fig. 13. The overlapping sp hybridized orbitals of both atoms correspond to electron cloud "most probably" existing in the space between the atoms, which keeps the repealing F^{1+} ions. The sum of two overlapping sp orbitals forms the so called σ bonding molecular orbital.

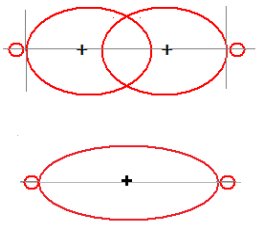
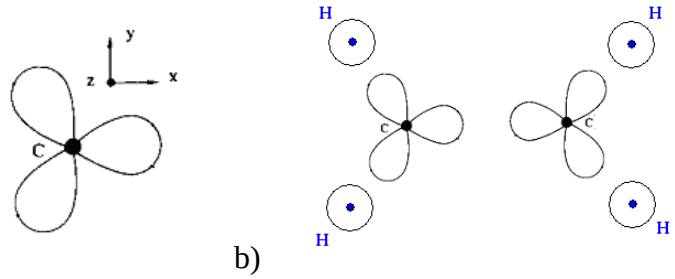


Fig. 13. Formation of covalent bond I F_2 molecule. Sum of two sp hybridized orbitals creates bonding molecular orbital σ.

Another hybridization, sp2, allows us to explain stability of many bonds, like e.g. these in benzene (and partially in C_2H_4). This hybridization occurs when one s and two p–type orbitals (e.g. p_x and p_y) are used to form three new orbitals (leaving the p_z orbital unchanged)s :

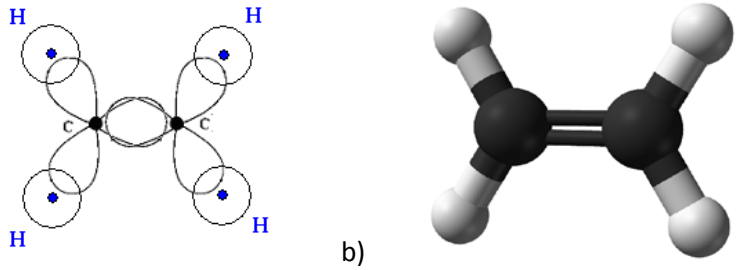
$$\phi_1 = \tfrac{1}{\sqrt{3}}(s + \sqrt{2}p_x)$$
$$\phi_2 = \tfrac{1}{\sqrt{6}}(\sqrt{2}s - p_x + \sqrt{3}p_y)$$
$$\phi_3 = \tfrac{1}{\sqrt{6}}(\sqrt{2}s + p_x - \sqrt{3}p_y)$$
$$\phi_4 = p_z$$

These orbitals are also directed into apexes of triangle, as shown in Fig. 14. Formation of covalent bonds in C_2H_4 molecule is presented in Fig. 15.



a) b)

Fig. 14. a) Hybridized orbitals sp2. b) 1s Hydrogen orbital and sp2 Hybridized orbital of carbon before formation of chemical bond in a molecule C_2H_4.



a) b)

Fig. 15. a) Formation of covalent bond in C_2H_4 molecule. b) A model of molecule C_2H_4. True bond between carbon atoms is not of σ type, but is of a π type formed from two sp3 hybridized orbitals of carbon. From http://en.wikipedia.org/wiki/File:Ethylene-CRC-MW-3D-balls.png

Periodic potentials

Crystals are physical systems with space-periodic structure. This structure is formed by atoms connected *via* chemical bonds. The crystal can be obtained by periodically repeating the so called *unit cell*, which contains one or more atoms forming the lattice. The center of each unit cell can be identified with *lattice nodes*. Fig. 16 shows some models of three-dimensional lattices (crystals) and a model of one-dimensional crystal is shown in Fig. 17. In both cases spheres mean atoms and (different colors for different types of atoms) and bars mean chemical bonds.

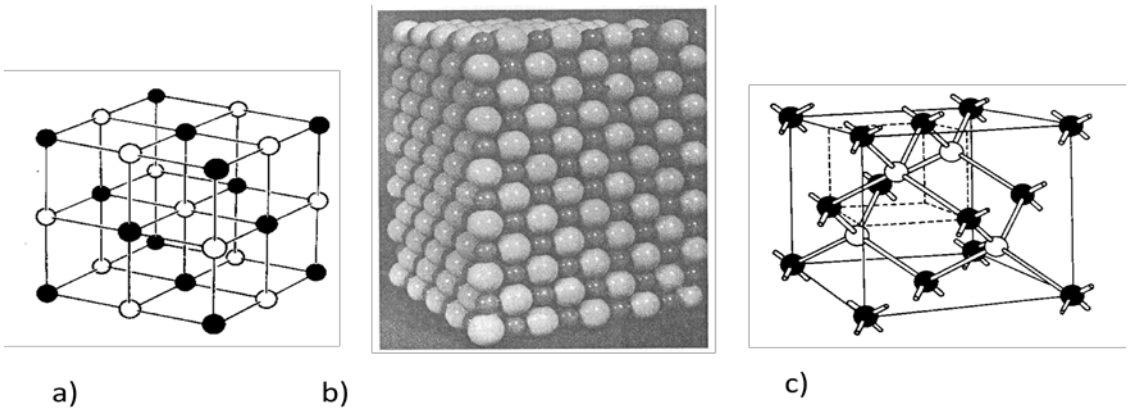


Fig. 16. Models of crystal lattices: a) and b) NaCl, c) GaAs. From "Wstęp do fizyki ciała stałego", C.W. Kittel, PWN, Warszawa 1999.

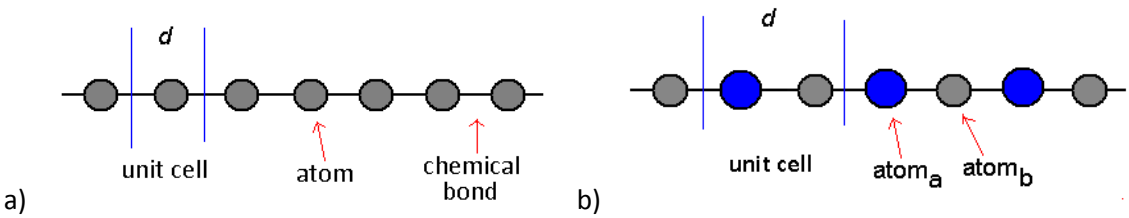


Fig. 17. A model of one-dimensional lattice, a) made of the same atoms, b) made of two different atoms. Case b) is so called "lattice with basis"- unit cell contains more than one atom.

Macroscopic crystal is a very complex system. It contains millions of strongly interacting atoms (electrons and protons). The exact quantum-mechanical description of such a system is impossible, not only because of computational complexity, but also due to problems with physical interpretation of functions, which depends on millions of variables. Therefore we usually make several approximations. In the first one we separate electrons from nuclei, similarly as what we did in the case of Hydrogen atom. In the next step, still very complex many-electron Hamiltonian is reduced to a one-particle Hamiltonian. This is called *one electron approximation*. It is assumed that we can describe the state of a single (individual) electron in a molecule or crystal, using one-electron function; further, the function describing the entire system can be approximately composed of one-electron functions. These one-electron functions are called atomic, molecular or crystal *orbitals*. They are solutions of one-electron Schrödinger-like equation; now each electron interacts with the lattice of positive ions and with all the other electrons. For crystals, this interaction potential is approximated by periodic term, which for one-dimensional chain of atoms (Fig. 17) reads: $V(x+d) = V(x)$, where d - period of lattice (the size of unit cell in this case).

Bloch theorem and band structure of energy in crystals

It is easy to understand that since crystal is periodic, the entire information about its properties is embedded in a single unit cell, plus boundary conditions imposed on the wave function. This condition is nothing else, but *translational symmetry*. In turn, translational symmetry means that if we move crystal by any lattice vector (a vector connecting any two lattice nodes) the crystal (lattice) doesn't change. It is obvious that this property can concern only infinite crystals. However, for crystals containing large number of unit cells, translation by a single unit will change almost nothing, except its edge. If we are interested in the properties of crystals originating from its *bulk* character, but not in the surface features, we can reasonably treat such large macroscopic crystal as infinite, We will show latter that this approximation will have some influence on possible values of the wave vector.

Now, let us have a look at the most important theorem concerning periodic systems, namely the Bloch theorem, which says that the wave function describing a state of electron in a periodic potential must fulfill the following condition (here for one-dimension only)

$$\Psi(x+d) = e^{ikd}\Psi(x) \tag{6a}$$

where k is the so called wavevector and d is translation vector. There is another form of this theorem

$$\Psi(x) = e^{ikx}u(x) \tag{6b}$$

where u(x) is a periodic functions, $u(x+d)=u(x)$.

Bloch theorem results from translational invariance of a given system. The Hamiltonian **H** commutes with the translation operator **T** and both operators have common eigenfunctions. Therefore, the eigenfunctions of **H** can be enumerated as eigenfunctions of **T**. A quantum number that enumerates eigenvalues of **T** is called ***k*** – wave vector, which in one dimension reduces to scalar *k*. In fact, if the right-hand side of (6b) depends on *k*, so $\psi(x)$ must me enumerated by *k*. The result of action of **T** on ψ is

$$T_d\Psi(x) = \Psi(x+d) = e^{ikd}\Psi(x).$$

Since e^{ikd} is a dimensionless quantity, the dimension of *k* is reciprocal lengths. e^{ikd} is periodic with periodicity equal 2π, so it is sufficient if *k* changes only in the interval $(0, 2\pi/d)$ or $(-\pi/d, +\pi/d)$. This interval is called *Brillouin zone (BZ)*. For infinite crystal *k* takes continuous values, but if the crystal is finite and contains e.g. N unit cells (in one dimension), *k* takes discrete values $0, \ldots, 2\pi n/(dN), \ldots$, very closely spaced for a large N (i.e. almost continuous index).

Independently of *k*, each wave function is also the solution of (2) with periodic potential, which are enumerated using "usual" quantum number, e.g. *n*. Therefore, for a periodic potential the eigenvalues of **H** have to labeled as $E_n(k)$, where for each *n*, *k* changes in $(-\pi/d, +\pi/d)$. The energy

spectrum have a structure of *bands*. Bands are enumerated by *n* and in each band the energy changes like $E_n(k)$, we call it *dispersion relation*. The number *k* enumerates N one-electron states belonging to a given band.

Metals and semiconductors

For finite crystals bands contain large numbers of one-electron levels densely located on the energy scale. The number of electrons is also finite. All these electrons occupy one-electron states (energy levels). In fact we usually use only the electrons from the outermost (valence) atomic shells, sine the periodic potential in equation (2) concerns only valence electrons. There are several possible situations, which are shown in Fig. 18.

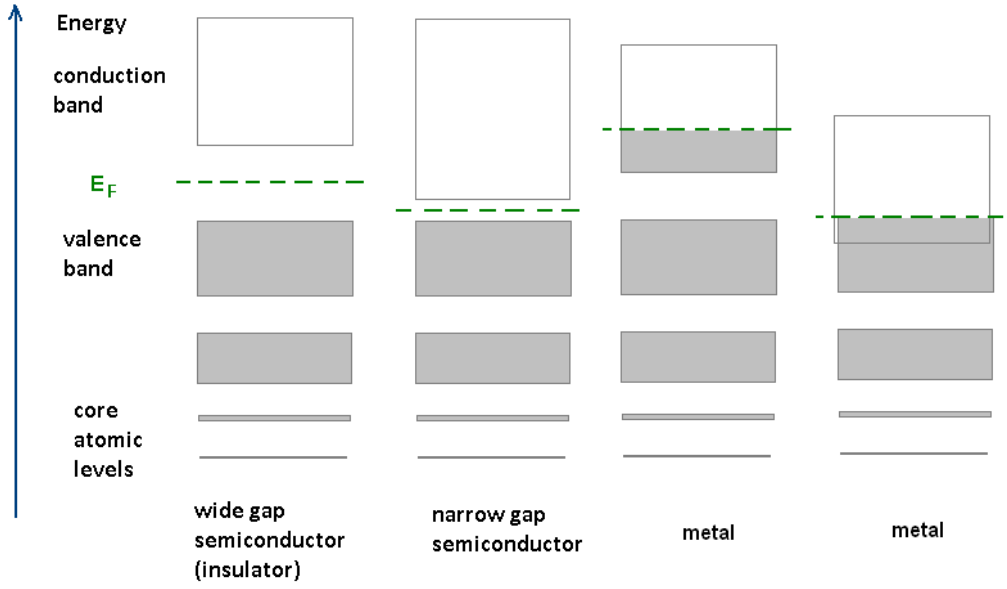


Fig. 18. Energy bands in insulators, semiconductors and metals.

Occupancy of bands determine whether a given crystal is metal, semiconductor or isolator. If the bands are separated by *energy gaps*, the number of valence electrons decides on the crystal character. Since due to spin each one-electron level can be occupied by two electrons, so for even number of electrons several bands are completely occupied and the other ones are empty. The highest (energetically) fully occupied band is called *valence band* and the lowest unoccupied – *conduction band*. When the energy band between these two bands is large (several eV) such crystal is an isolator, when the gap is small, we deal with semiconductor. It these two bands overlap or if the number of electrons is odd we get metal and when the bands merely touch each other we deal with *semimetal*. Graphene is such a case of semimetal.

In metals, the highest occupied level (in T=0K) marks the so called Fermi level (or Fermi energy) E_F. For semiconductors E_F lies in the middle of the energy gap. Metal can conduct electric current, because just above the Fermi level there is continuum of unoccupied one-electron levels, which can be easily occupied by electrons slowly increasing their energies in the electric field. This is impossible in semiconductors and insulators, since high energy is required to move (energetically) electron from valence to conduction band. Such transitions are usually performed optically.

Course 3. 3 hours.

Methods of calculation of the energy bands. Graphene structures.

Model space.

Instead of solving differential equation (2) one can approximately expand the wave function in a finite orthonormal basis set of N functions φ

$$\Psi = \sum_{i=1}^{N} c_i \varphi_i \tag{7}$$

and approximate the energy as

$$E = <\psi \mid H \mid \psi> . \tag{8}$$

In the basis φ_i, the expression $<\psi|H|\psi>$, strictly speaking the set of numbers

$H_{ij} = <\varphi_i|H|\varphi_j>$, takes the form of matrix and the problem $H\Psi(\boldsymbol{x}) = E\Psi(\boldsymbol{x})$ (eq. 1) is equivalent to the problem of finding eigenspectrum of Hamiltonian matrix H

$$\begin{bmatrix} H_{11} & \dots & H_{1N} \\ \dots & \dots & \dots \\ H_{N1} & \dots & H_{NN} \end{bmatrix} \begin{bmatrix} c_1 \\ \dots \\ c_N \end{bmatrix} = E \begin{bmatrix} c_1 \\ \dots \\ c_N \end{bmatrix} \tag{9}$$

The solutions reduces to diagonalization of matrix H, which is done by unitary transformation S, such, that $SHS^T = D$, where D is diagonal matrix. The eigenvalues of H are situated at the diagonal of D. We get N values E_i (some of them can be the same – degeneracy), which in line with the *variational principle* are interpreted as upper approximations (bounds) to the exact energies of the system represented by the Hamiltonian $\boldsymbol{H}$. For periodic systems all H_{ij} depend on k, since the wave function depends on k. Therefore, solving (9) for periodic system we get energy bands $E_i = E_i(k)$. However, we have to solve (9) for many values of k from the BZ.

Graphene.

Graphene is a two-dimensional rhombic (hexagonal) lattice with two-atomic basis (see Fig.19).

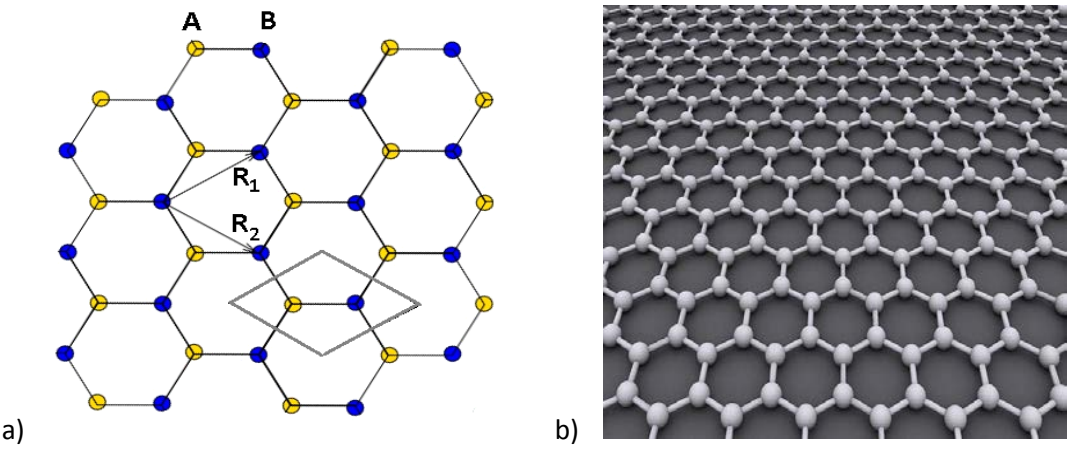


Fig. 19. a) Hexagonal lattice of graphene. Unit cell contains two carbon atoms. R_1 and R_2 are two basis vectors of this lattice (any other lattice-translation vector can be built as a linear combination of R_1 and R_2). b) A model of graphene. From http://en.wikipedia.org/wiki/Graphene

It was for the first time isolated from graphite in 2005 by Geim and Novoselov, who in 2010 were awarded Nobel Prize for this achievement. They obtained single layers of graphene using a Scotch tape for removing consecutive layers from the "bulk" graphite. It is quite easy to recognize a single layer since, contrary to any other graphite system, it is transparent. Nowadays graphene is obtained also using CVD methods or by evaporation of silicon from layers of SiC. Examples of single layers of graphene obtained using STM techniques are shown in Fig. 20.

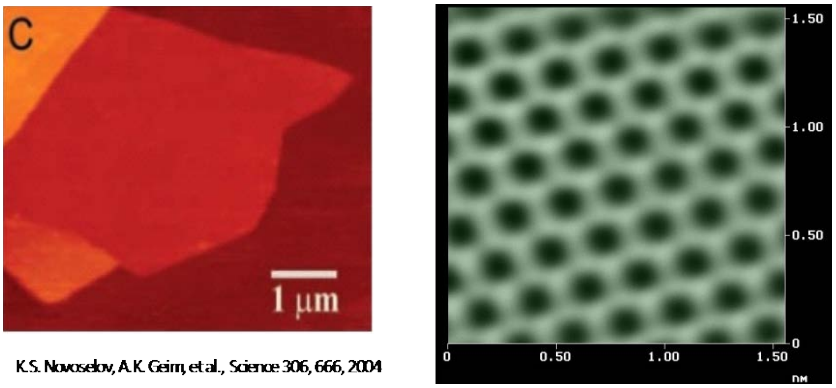


Fig. 20. Microscopic pictures of graphene. From
http://www.polymermicroscopy.com/eng_afm_graphit.htm
and http://www.graphene.manchester.ac.uk/story/image-gallery/

Stability of graphene is guaranteed by the in-layer covalent σ bonds, formed between neighboring carbon atoms. These bonds are formed by overlapping sp2 hybridized orbitals. However, each carbon atom supports yet another electron occupying p_z orbital (perpendicular to the graphene layer). These orbitals also form bonds, the so called π bonds (π molecular orbitals) , bonding or anybonding, depending on whether we add or subtract p_z orbitals from neighboring atoms. Since graphene is periodic in two dimensions, each energy level corresponding to π-type molecular orbital transforms into a band. Such bands are called π-bands. Both bands $E_i(\boldsymbol{k})$, i=1,2 (π and π^*) depend on the wavevector, which now is a two-dimensional vector (k_x , k_y), The Fermi level is situated between them and this is the reason why these bands are responsible for electrical properties of graphene.

Tight binding method.

The tight binding method is a practical realization of a model space method. We expand one-electron functions (7) describing valence electrons (π electrons in graphene) onto the basis of p_z atomic orbitals. The unit cell contains only two such orbitals and due to periodicity we first have to construct crystal (delocalized) orbitals of them. They are called the *Bloch sums (orbitals)*, since they fulfill the Bloch theorem. They depend on *k*. Such two Bloch orbitals simplify the problem: dimension of the Hamiltonian matrix is equal 2 (but each matrix element depends on *k*). Such a simple matrix one can easily diagonalize analytically. As the result we get two solutions E(k), one is called π-bonding $E_+(k)$ and another antibonding $E_-(k)$.

$$E(k_x, k_y) = \pm t\left\{1 + 4\cos\left(\frac{\sqrt{3}k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + 4\cos^2\left(\frac{k_y a}{2}\right)\right\}^{1/2} \tag{10}$$

Brillouin zone is also two-dimensional and looks like a hexagon. The energies of both bands are shown in Fig. 21.

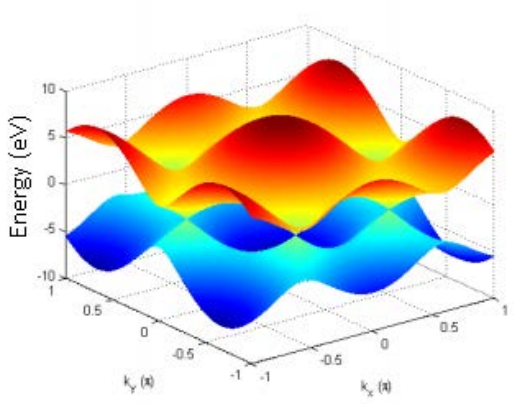


Fig. 21. π bands in w graphene. From PhD thesis by M. Pelc, Toruń, 2012.

The Hamiltonian matrix has four elements. The diagonal ones correspond to two p_z orbitals at the same node $H_{ij} =< \varphi_i|H|\varphi_j >$. It is the so called *on site* integral. Since both diagonal elements are the same we can assign them arbitrary value, which can be treated as arbitrary choice of the origin at the energy scale. Customarily we assign them value zero. Off-diagonal elements depend on $H_{ij} =< \varphi_i|H|\varphi_j >$, where *i,j* correspond to neighboring atoms. This integral is also treated as an empirical parameter, its value (taken from the experimental data) is about 2.7 eV. After diagonalization we get (10).

Finite graphene structures.

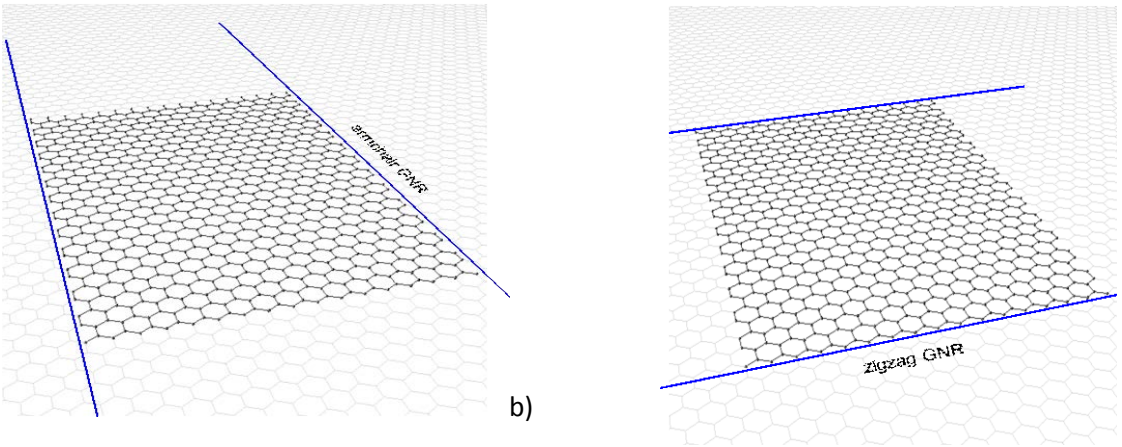
Finite graphene structures are in fact large molecules built of carbon atoms, they are not periodic, so for them we cannot keep the concept of a unit cell. Therefore, a basis in (7) must contain p_z orbitals localized at all atoms. *k* is no longer a good quantum number (we need not to solve equation 9 for many *k* values, but instead we the Hamiltonian matrix may be really big. Since we usually work in the

nearest neighbors approximation, which neglects all the integrals connecting non-neighboring atoms, the *H* matrix is almost empty – non-zero elements appear only close to diagonal. After diagonalization we obtain a set of eigenvalues, which represent energy levels of a given "molecule".

Periodic graphene structures.

Nanoribbons.

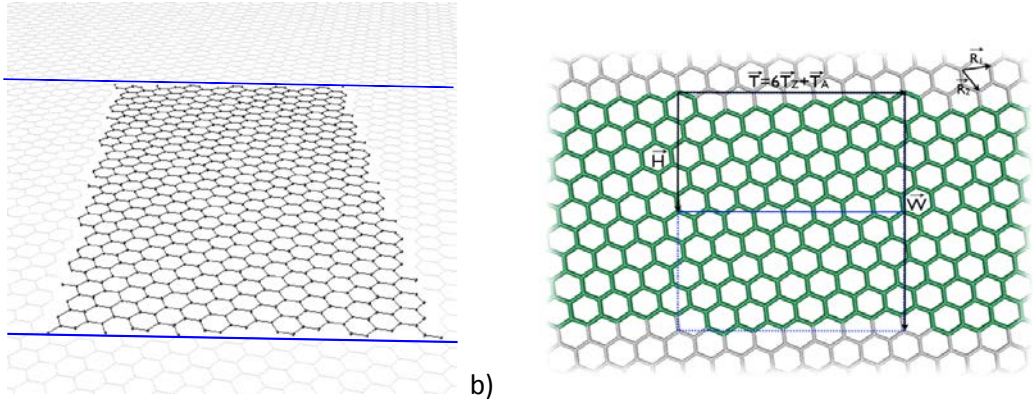
Graphene nanoribbons (GNR) and carbon nanotubes (CN) belong to periodic structures made of graphene. Ribbons are formed by cutting-off a strip from the graphene sheet. When the cut goes along σ bonds we get *armchair* ribbon; when the cut goes in perpendicular direction we obtain *zigzag* ribbon. Examples of such ribbons are shown in Fig. 22.



a) b)

Fig. 22. Graphene nanoribbons. a) armchair, and b) zigzag.

Cutting graphene in any other direction we get *chiral* ribbon; two examples are shown in Fig. 23. Such a ribbon may have quite irregular, but periodic, edges.



a) b)

Fig. 23. Examples of chiral ribbons. In b) unit cell is marked as space confined by vectors T and W.

Any GNR can be uniquely identified (determined) by three elements: the translation vector T, the width W and details of the edge (edges) in the GNR unit cell. The unit cell of GNR contains all the atoms enclosed in the rectangle made of T and W (see Fig. 23 b).

It is easy to built Hamiltonian matrix for such a system. The procedure is the following:

1. first, enumerate all the nodes in the unit cell and do the same for neighboring unit cells,
2. set all the matrix elements equal to zero,
3. in off-diagonal elements connecting atoms belonging to the same unit cell put value *t*,

4. in matrix elements connecting i-th node of a given unit cell with j-th node of the right unit cell put value *t x exp(ikd),* while in elements connecting node "k" of a given unit cell with node "l" of the left unit cell put *t x exp(-ikd)*,
5. check whether the entire matrix is hermitian,
6. since *k* changes in the interval 0-2π/d, the factors *exp(±ikd)* can be substituted by *exp(±ik)* and *k* can vary from 0 to 2π.

Nanotubes.

Graphene nanoribbons can be rolled up to form carbon nanotubes, as it is shown in Fig. 24.

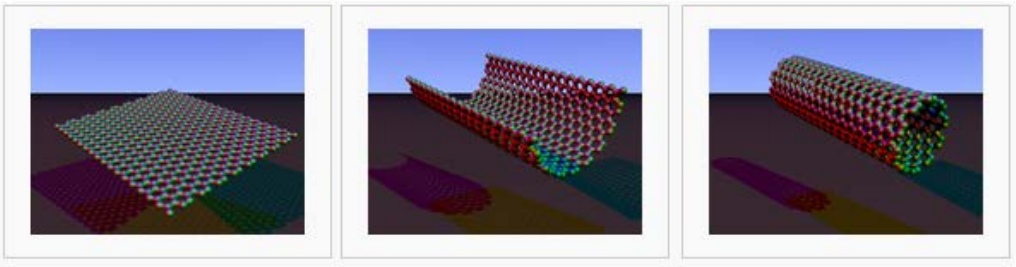


Fig. 24. Formation of a nanotube from a graphene nanoribbon.
From http://en.wikipedia.org/wiki/Carbon_nanotube

In case of nanotubes the carbon atoms from one edge of the ribbon must be connected with carbon atoms from the opposite edge. This is shown in Fig. 25. One has to remember about that when the corresponding Hamiltonian matrix is formed.

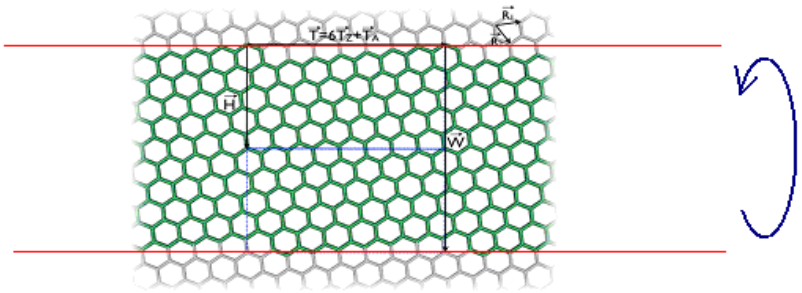


Fig.25. Formation of nanotube from GNR by rolling up and connecting atoms from opposite edges.

Here are three different types of nanotubes, armchair, zigzag and chiral, they are presented in Fig. 26. Now, the type of tube is determined by the shape of edge when the tube is cut perpendicular to its axis.

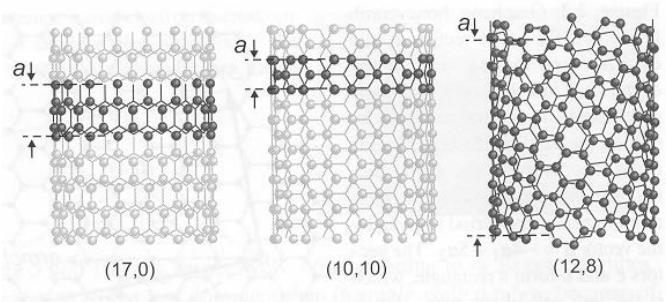


Fig. 26. Three types of carbon nanotubes. From the left: zigzag, armchair and chiral. From "Physical Properties of Carbon Nanotubes",M.S. Dresselhause, G. Dresselhause, and R. Saito, Imperial College Press, London 1998.

Any carbon tube can be uniquely defined by *chiral vector*, called also a vector of circumference. It is in fact vector W shown in Fig. 25 (usually it is named C_h). Components of chiral vector in the basis defining graphene lattice (vectors R_1 and R_2 in Fig. 19a) reduce to two integer numbers (n,m). These two numbers uniquely define any carbon nanotube.

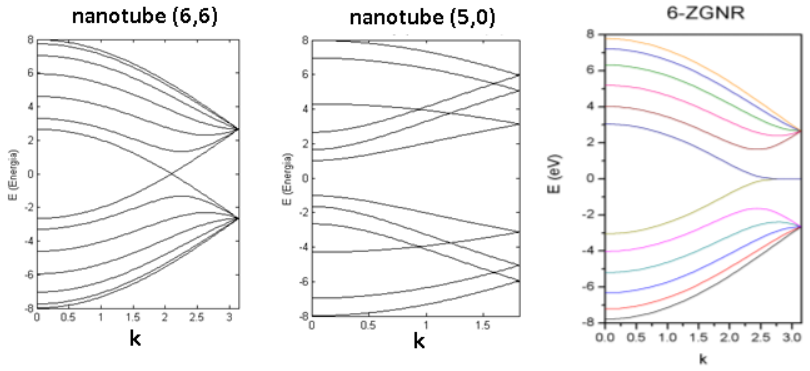


Fig. 27. Band structure of three different graphene systems. From the left: armchair tune (6,6), zigzag tube (5,0) and zigzag GNR of width equal 6 graphene unit cells.

Since GNRs and CNs are quasi-one-dimensional systems, their energy structure has a band character $E_i(k)$. Three examples are shown in Fig. 27. Depending on the diameter and chirality of a tube, and in case of GNR – on its widths and edge-shape, these systems can be metallic or semiconducting. We can see in Fig. 27, that CN (6,6) and zigzag GNR are metals – there is no energy gap between the valence and conducting bands (remember that E_F, which is at zero energy, separates both bands). In contrast, the tube (5,0) is a semiconductor with the energy gap as large as ~ 2 eV. Additionally, two different nanotubes can be connected (joined) by the so called *topological defects 5/7*. One example is shown in Fig. 28. Junctions between the tubes can be exploited to built *quantum dots* and *superlattices*.

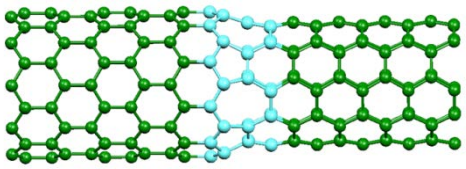


Fig. 28. Junction between zigzag (2n,0) and armchair (n,n) tubes. Junction is defined by n pentane/heptagon pairs.

Graphene ribbons and carbon nanotubes may have potential applications in future carbon-based nanoelectronics. Prototype transistors have already been fabricated using both the tubes and the ribbons. They have sizes of several nanometers.

Resonance states of particles in nanostructures

Basic information on properties of nanostructures, or rather, on a behavior of particles placed in them, can be obtained by modeling the influence of the nanostructures on particles by relatively simple potentials. How to guess such a potential? Let us think about a case of a nonostructure consisting of several regions of different semiconducting materials (conductors and isolators also possible) and consider an electron moving in such a structure. For each of these materials there is a characteristic distribution of bands of allowed energies. In particular there is a conduction band for each material, placed in a proper place in the energy scale. An electron having energy within the conduction band "feels free". However, the region built of a given material is finite, of size of a few or a few tens of nanometers. So the electron sooner or later comes across a border with another material in which its energy coincides with the energy gap (a range of forbidden energies). Then the electron cannot easily enter in such a material. However this is possible. Probability of entering the other material depends on the electron energy and on the conduction energy difference on the border between the materials. Only an ideal isolator is closed for the electron. In this case the conduction band is placed at infinite energies. A model potential for electrons in such nanostructures is assumed to be equal to the bottom energies of conduction bands of material filling given space regions.

In general the situation can be complicated depending on a spatial distribution of various materials forming the nanostructure. The simplest case is a system of few parallel layers few nanometers or few tens of nanometers thick placed one after another along a given direction and of large size in the perpendicular direction so that they can be considered as infinite in this dimension. Then the potential does not change in the planes of layers. It changes along the axis of distribution of layers only. The model can be confined to this dimension only. The problem of finding the energy structure consists in solving the Schroedinger equation (the energy operator eigenequation) for a one-dimensional system of square wells and barriers. Which part is a well which one is a barrier depends on relative values of the potential.

What information we receive? From mathematical point of view we obtain eigenvalues of the energy operator (Hamiltonian). From physical point of view we learn what energies are allowed for the electron in the nanostructure. These can be discrete, completely determined, corresponding to bound states of the electron in the nanostructure, so that it cannot escape. Knowing them one can determine the optical properties of the nanostructure, e.g. one can predict what light can be emitted or absorbed by the nanostructure and what should be the frequency of the electromagnetic wave that could set the electron bound to the nanostructure free. States that are not strictly bound but quasibound (or pseudobound) are also possible. These are states in which the electron seems to be bound but can escape from the nanostructure within a finite time without receiving extra energy from the environment. As suggested above the electron can penetrate a space region in which its energy is below the conduction band. This means that it is possible for the electron to get out of the nanostructure. Such states also correspond to characteristic energies. However they are not strictly discrete but rather "diluted" or broadened. Pseudobound states are very important for determination of conducting properties of nanostructures (They determine whether and when the electrons can get inside and outside the nanostructure).

This part of the teaching materials is devoted to quasibound states, also known as resonance states. We begin with a consideration of a simple one-dimensional model : a square well-barrier system. Obtaining analytical solutions for this case we get known basic properties of the resonance states. Then we present two methods developed specially for searching for the resonance states: the stabilization method and the complex coordinate rotation method.

Well-barrier system

Here we are considered with a one-dimensional well-barrier system which seems to be the simplest case illustrating the fact of the existence and properties of resonance states. First we investigate a rectangular potential well for which we find bound states. Then we add a rectangular barrier to show how the bound states of the potential well transform into resonance states and how their properties are transformed.

Bound states in a rectangular potential well

The particle under consideration (an electron) lives in a straight line. Its location is determined by one coordinate: x. It interacts with the nanostructure (the rest of the world) so that it gets the potential $V(x)$. The nanostructure under consideration is a semiconductor nanolayer of thickness a. We assume the bottom of the conduction band of this layer as 0 on the energy scale. On the right side of the layer we have another infinite semiconductor whose conduction band is placed at V_1(higher). On the left side there is an infinite ideal isolator of infinite potential (the conduction band lies infinitely high). The potential is shown in Fig. 29 and can be written as

$$V(x) = \begin{cases} \infty, & \text{dla } x < 0, \\ 0, & \text{dla } 0 \le x \le a, \\ V_1, & \text{dla } x > a. \end{cases}$$

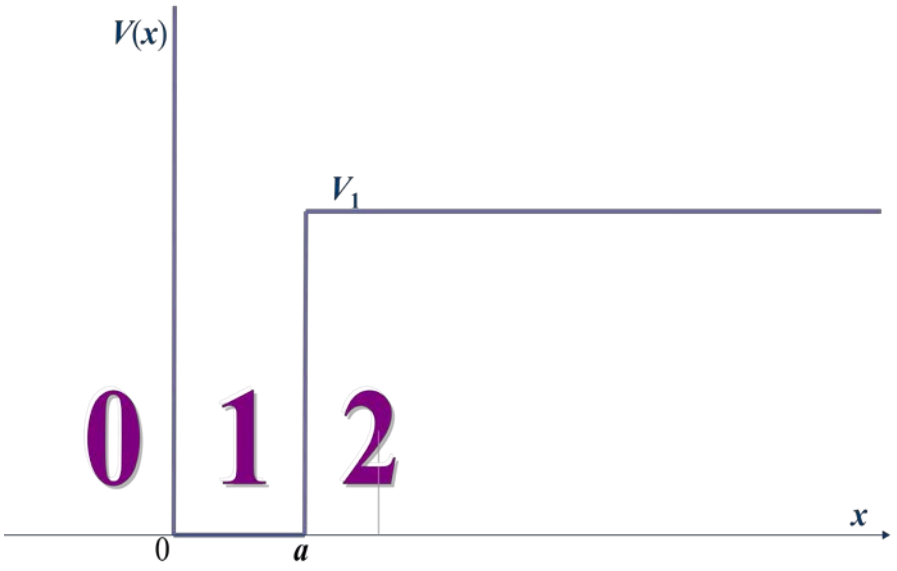


Fig. 29. Rectangular potential well. It can model the system of a semiconductor nanolayer (region number 1) between an infinite semiconductor (region 2) of a conduction band placed by V_1 above the bottom of the conduction band in the nanolayer and an infinite isolator (region 0). Only bound states are expected in the energy range between 0 and V_1.

We are going to solve the Schroedinger equation for this system. Let us focus on the energy range from 0 to V_1. One cannot expect any physical solutions below zero because they would implicate negative values of the kinetic energy of the particle. Similarly, the particle cannot escape to the infinity having the total energy in the range $0 \leq E \leq V_1$ because its kinetic energy would be negative. Such an escape is not possible. Thus for the total energy restricted to $0 \leq E \leq V_1$ we expect bound states only.

Because the potential is defined in a specific way, in three regions, we solve the Schroedinger equation in each region separately and then we combine the three solutions into one for the whole range of x. For region „0" we have

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Phi_0(x) + (\infty - E)\Phi_0(x) = 0.$$

The equation can be satisfied in two cases: either $\Phi_0(x) = 0$ everywhere or the second derivative of Φ_0 is infinite whenever Φ_0 differs from 0. The latter is absurd. However the former possibility is both mathematically simple and proper and physically obvious and meaningful: the wavefunction equal to zero implicates the probability distribution equal to zero. This is what we thougth about the system: the electron cannot enter the isolator (region "0"). Thus we have got the solution for region "0":

$$\Phi_0(x) = 0.$$

Inside the well (in the line segment $[0, a]$) the Schroedinger equation reads

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Phi_1(x) + E\Phi_1(x) = 0$$

or

$$\frac{d^2}{dx^2}\Phi_1(x) + k_1^2\Phi_1(x) = 0, \text{ where } k_1^2 \equiv \frac{2mE}{\hbar^2},\ k_1 > 0.$$

The equation is identical to the harmonic oscillator equation. Its general solution can be written as

$$\Phi_1(x) = A\sin k_1 x + B\cos k_1 x.$$

We still need to solve the problem in region „2". Equation

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Phi_2(x) + (V_1 - E)\Phi_2(x) = 0$$

is equivalent to

$$\frac{d^2}{dx^2}\Phi_2(x) - k_2^2\Phi_2(x) = 0, \text{ where } k_2^2 \equiv \frac{2m(V_1 - E)}{\hbar^2},\ k_2 > 0.$$

It differs from the harmonic oscillator equation with the sign at the factor k_2^2. The general solution

$$\Phi_2(x) = Ce^{-k_2 x} + De^{k_2 x},$$

is a linear combination of two exponential functions: decreasing to zero and exploding at infinity. The latter one is unacceptable for physical reasons. Probability of finding the particle at infinity (far from the well) must be zero. Therefore from among all mathematical solutions we chose only these physically proper by setting $D = 0$. (We will see that this physical restriction causes the discrete character of allowed energies). Eventually the solution for region "2" is

$$\Phi_2(x) = Ce^{-k_2 x}.$$

Now the solutions for particular region need to get stitched up to form a continuous function. This will assure that the probability density of finding the particle at x is a continuous function of x. Thus we demand that the function and its first derivative (if possible) are continuous at mend points $x = 0$ and $x = a$, i.e.

$$\begin{aligned} \Phi_0(0) &= \Phi_1(0), \\ \Phi_1(a) &= \Phi_2(a), \\ \left.\frac{d\Phi_1}{dx}\right|_{x=a} &= \left.\frac{d\Phi_2}{dx}\right|_{x=a}. \end{aligned}$$

(In the case under consideration one cannot assure continuity of the first derivative at $x = 0.$) These conditions lead to a set of linear equations for coefficients A, B i C:

$$\begin{aligned} 0 &= A\sin 0 + B\cos 0 \\ A\sin k_1 a + B\cos k_1 a &= Ce^{-k_2 a} \\ k_1 A\cos k_1 a - k_1 B\sin k_1 a &= -k_2 Ce^{-k_2 a}. \end{aligned}$$

From the first one obtains immediately $B = 0$. Two remaining equations

$$\begin{aligned} A\sin k_1 a - Ce^{-k_2 a} &= 0 \\ k_1 A\cos k_1 a + k_2 Ce^{-k_2 a} &= 0 \end{aligned}$$

should give us A i C. If these equations are linearly independent then the obvious solution $A = 0$ i $C = 0$ is the only one. However it cannot be accepted from physical point of view: it means that probability of finding the particle at any position is zero, i.e. the particle does not exist (?!).

Are there any physical solutions ? Yes, under the condition that the equations are linearly dependent. To make them linearly dependent we demand that

$$\begin{vmatrix} \sin k_1 a & -e^{-k_2 a} \\ k_1\cos k_1 a & k_2 e^{-k_2 a} \end{vmatrix} = 0.$$

This is a constraint put on energy E hidden in k_1i k_2. It may happen that for some energy values of the range above 0 and below V_1 the condition above is fulfilled. Then only these values of energy are allowed for the particle in the well; only for them there exist states of the particle bound in the well.

The above equation is nonlinear and therefore solving it is not a trivial task. Expanding the determinant and assuming (for simplicity) $\xi \equiv k_1 a$ and $\eta \equiv k_2 a$ we obtain

$$\eta = -\xi\cot\xi.$$

From definitions of k_1 and k_2 we have got

$$\xi^2 + \eta^2 = \frac{2m}{\hbar^2} V_1 a^2 \equiv R^2.$$

Fig. 30 contains plots of these two relations between ξ and η, performed for $m = 1$ a. u., $a = 2$ a. u., $V_1 = 5$ a. u. i $\hbar =$ a. u. (We use arbitrary units specific for the system under consideration: The mass unit is equal to the mass of our particle. Energy unit is such that the Planck constant divided by 2π is 1). From the figure we see that (in this case) we have two intersection points at which both relations are satisfied. It is easy to show that in a general case the number of solutions (intersection points) n is determined by $(2n-1)\frac{\pi}{2} \leq R < 2n+1)\frac{\pi}{2}$ and is always finite. It depends on the size of the well: on the width a and the depth V_1. For a narrow and/or shallow potential well there might be no solutions.

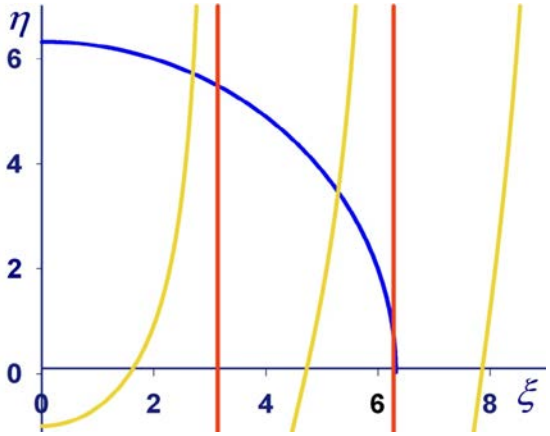


Fig. 30. Plots of $\eta = -\xi \cot \xi$ (yellow) and $\xi^2 + \eta^2 = R^2$ (blue). The intersection points correspond to nonzero solutions of the Schroedinger equation. In red the vertical asymptotes are given at $\xi = m\pi$.

The energy values corresponding to these solutions, the Hamiltonian eigenvalues, can be deduced from the coordinates, ξ and η, of the intersection points. We use a bit different, though graphical, method. We plot function

$$f(E) = \eta(E) + \xi(E) \cot \xi(E),$$

for $0 < E < V_1$ and search for its roots, zero-value points. Such a plot is given in Fig. 31 (for the parameter values as assumed above). We find two roots: $E_1 \cong 0.91155$ a. u. i $E_2 \cong 3.5005$ a. u. Using them one after the other in the system of linear equations for A and C, which are linearly dependent now, we obtain:

$$C_i = A_i e^{k_2(E_i)a} \sin[k_1(E_i)a], \text{ for } i = 1,2\,.$$

Finally, the wave functions corresponding to E_i can be written as

$$\Psi_i(x) = A_i \begin{cases} 0, \text{ for } x < 0, \\ \sin[k_1(E_i)x], \text{ for } 0 \leq x \leq a, \\ e^{-k_2(E_i)(x-a)} \sin[k_1(E_i)a], \text{ for } x > a. \end{cases}$$

They are plotted in Fig. 32.

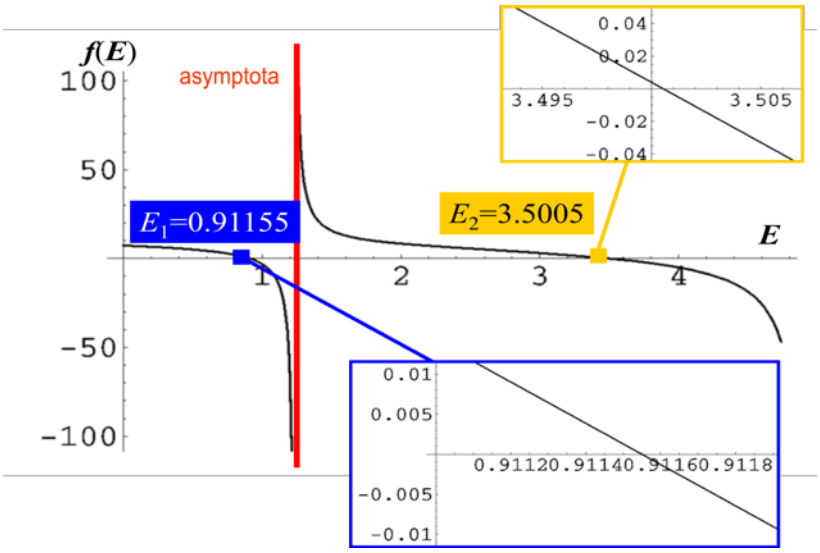


Fig. 31. Function $f(E) = \eta(E) + \xi(E)\cot\xi(E)$. Its roots correspond to energy eigenvalues (energies allowed for our system). The line in red is an asymptote.

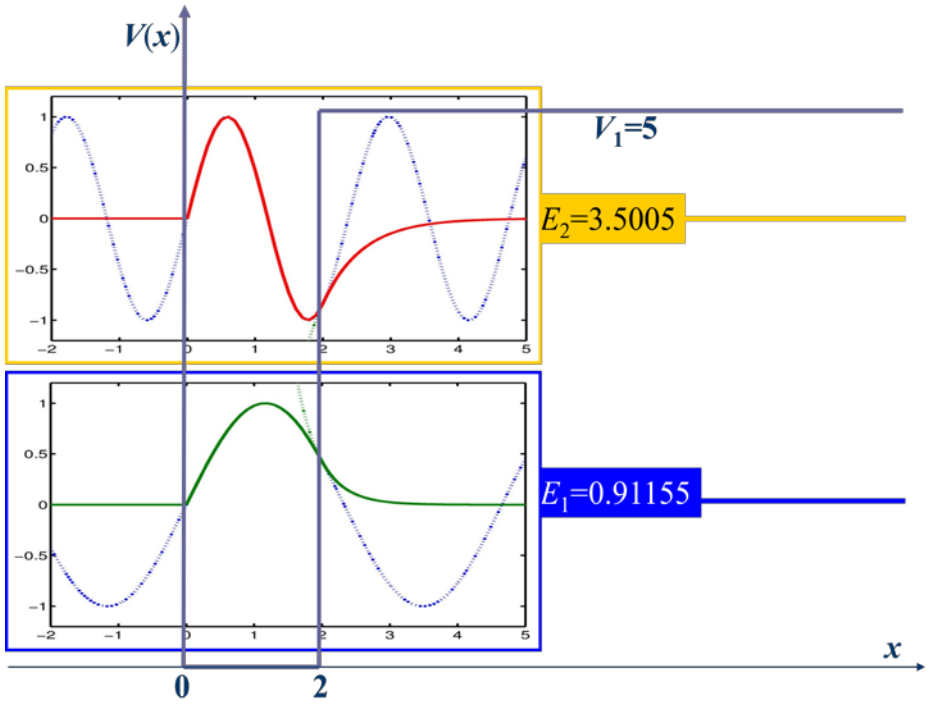


Fig. 32. Wave functions for two bound states of the particle in the potential well under consideration. They are placed on the background of the potential plot. The aim of this traditional way of presenting them is just to assign them to proper eigenenergies and confront their space diffuseness with the width of the well. The only common axis is the x-coordinate axis. Each of the plots has its own function-value axis (also independent of the energy axis).

Resonance states of a potential well-barrier system

We modify the potential. Region "2" is finite now, it extends to $x = b$. It constitutes a barrier for the particle. Then, behind this barrier, we have a region of another material of the conduction band bottom at $V_2 < V_1$. The potential energy of the electron

$$V(x) = \begin{cases} \infty, \text{for } x \le 0, \\ 0, \text{for } 0 \le x \le a, \\ V_1, \text{ for } a < x \le b, \\ V_2, \text{for } x > b \end{cases}$$

is plotted in Fig. 33.

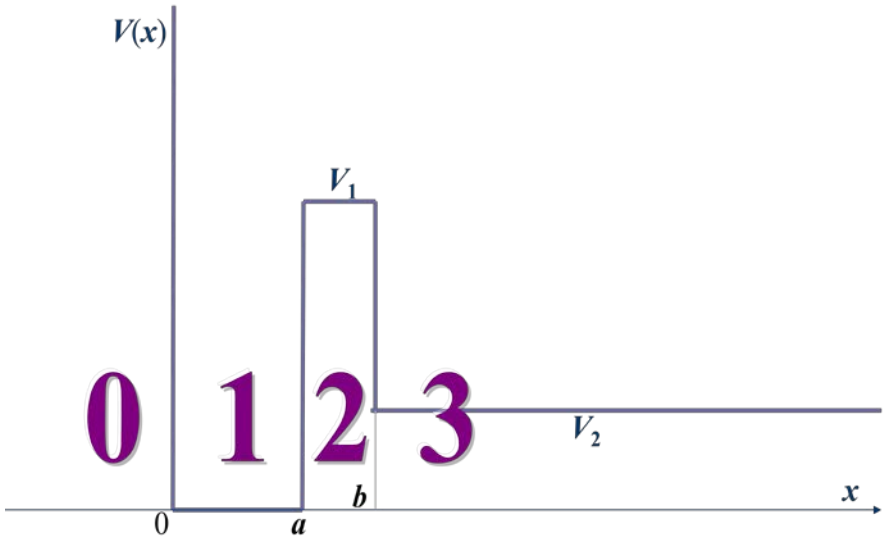


Fig. 33. Well-barrier potential. Infinite region "2" of Fig. 2.1 has been restricted to the segment $[a, b]$. Now it constitutes a finite potential barrier between the well ("1") and an infinite semiconductor region ("3") of the conduction-band bottom shifted by V_2 with respect to the conduction-band bottom of "1". The particle located in "1" and having the energy above V_2 can get out of the well by tunneling through the barrier "2" to the outside region "3" and then escape to the infinity. Such a state of the particle is not bound but resonant.

We are interested in the energy range $V_2 < E < V_1$. As before we solve the Schroedinger equation in each particular region separately. The solutions for regions "0" and "1" are the same as for the previous case. Region "2" is finite now so the increasing exponential function is not a "physical danger". Thus

$$\Phi_2(x) = Ce^{-k_2 x} + De^{k_2 x}$$

is an acceptable general solution for this region. Situation in the outer region "3" is different than in the previous case: the particle energy is considered to be above the potential energy in this region so there is no reason to keep the particle in a finite (small) distance from the well. The Schroedinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Phi_2(x) + (V_1 - E)\Phi_2(x) = 0$$

or

$$\frac{d^2}{dx^2}\Phi_3(x) + k_3^2\Phi_3(x) = 0, \text{ where } k_3^2 \equiv \frac{2m(E-V_2)}{\hbar^2},\ k_3 > 0.$$

Again this equation is identical to the harmonic oscillator equation and the general solution can be taken in the form of a linear combination of sine and cosine functions of frequency k_3 or equivalently

$$\Phi_3(x) = Fe^{-ik_3x} + Ge^{ik_3x} = F(e^{-ik_3x} + Se^{ik_3x}).$$

The latter form is more plausible because it offers a clear physical interpretation. Functions e^{ik_3x} and e^{-ik_3x} are eigenfunctions of the momentum operator ($\hat{p} = -i\hbar\frac{d}{dx}$) corresponding to the eigenvalues $k_3\hbar$ and $-k_3\hbar$, respectively. Hence the former one describes a particle of positive momentum running away from the well-barrier system. Whereas the latter one corresponding to negative momentum describes the particle moving toward the barrier- well. Therefore e^{ik_3x} is called an outgoing wave, whereas e^{-ik_3x} is known as incoming wave. If one considers the physical situation of beams of many particle-projectiles shot against the barrier-well and scattered on it, then the incoming wave describes the beam of particles attacking the barrier and the outgoing wave represents the beam of scattered particles. Measures of beam intensities are $|F|^2$ and $|G|^2$, respectively. Their ratio , $|S|^2$, is the reflection coefficient. This is why S is the most interesting parameter of the solution in the outside region.

Stitching of the particular "regional" solutions into a continuous and smooth global wave function relies in formal conditions:

$$\begin{gathered}\Phi_0(0) = \Phi_1(0),\\ \Phi_1(a) = \Phi_2(a),\\ \left.\frac{d\Phi_1}{dx}\right|_{x=a} = \left.\frac{d\Phi_2}{dx}\right|_{x=a},\\ \Phi_2(b) = \Phi_3(b),\\ \left.\frac{d\Phi_2}{dx}\right|_{x=b} = \left.\frac{d\Phi_3}{dx}\right|_{x=b}.\end{gathered}$$

Their explicit forms are

$$\begin{gathered}0 = A\sin 0 + B\cos 0\,,\\ A\sin k_1a + B\cos k_1a = Ce^{-k_2a} + De^{k_2a},\\ k_1A\cos k_1a - k_1B\sin k_1a = -k_2Ce^{-k_2a} + k_2De^{k_2a},\\ Ce^{-k_2b} + De^{k_2b} = F(e^{-ik_3b} + Se^{ik_3b}),\\ -k_2Ce^{-k_2b} + k_2De^{k_2b} = -ik_3F(e^{-ik_3b} - Se^{ik_3b}).\end{gathered}$$

They form a set of five linear equations for six unknown (arbitrary up to now) coefficients: A, B, C, D, F, S. Besides the obvious zero solution, there are also other solutions physically meaningful. This means, that no extra conditions need to be imposed on energy; for every value of the energy of the assumed range there exist physical solutions of the Schroedinger equation, i .e. all energy values of $V_2 < E < V_1$ are allowed for the particle.

Someone who is interested in energies only could stop here. However we expect that the wavefunctions may change significantly with energy. Therefore in the following we investigate how

the solutions depend on energy, $\Psi(x;E)$, in the range $V_2 < E < V_1$. Instead of solving the set of linear equations completely for all the unknowns, let us think in which parameters we are really interested. Then we shall search for answers to the questions of interest only. Thinking on the problem in terms of a beam of particles which scatter on the barrier-well device getting reflected or tunneling inside the well we find the barrier part of the solution not very interested. We will not solve the problem for C and D.

The wavefunction inside the well is determined by coefficient A (because $B = 0$). Probability of finding the particle inside depends on A. Evaluation of the probability is not simple in this case because the wavefunction is not normalized to unity (it is not square-integrable; it does not vanish at infinity). If it was then the probability of finding the particle of energy E inside the well could be evaluated as $P = \int_0^a |\Psi(x;E)|^2\, dx$. Assuming that the "missing normalization factor" is independent of energy value we can accept the ratio

$$\frac{\int_0^a |\Psi(x;E_1)|^2\, dx}{\int_0^a |\Psi(x;E_2)|^2\, dx} \cong \frac{|A(E_1)|^2}{|A(E_2)|^2}$$

as a relative probability of finding the particle in the well. The meaning of it is how many times the probability at energy E_1 is bigger than the probability at energy E_2. (Farther approximation in the above formula follows from the fact that $k_1(E)$ is rather a smooth function of E in comparison to rapid changes of $A(E)$).

The most important information on the solution in outer region is the coefficient S. Coefficient F may be considered as independent. (There are five equations for six unknowns. The solution is ambiguous, one can only set down relations of five of them to the sixth one, considered as independent. The choice of F to be independent one is natural: it is F which determines the beam of incident particles).

We solve the equations with respect A and S. Here they are:

$$A = -F\,\frac{4ik_2k_3e^{2k_2(a+b)}e^{-ik_3b}}{k_2\sin k_1a[e^{2k_2b}(k_2-ik_3)-e^{2k_2a}(k_2+ik_3)]+k_1\cos k_1a[e^{2k_2b}(k_2-ik_3)+e^{2k_2a}(k_2+ik_3)]}$$

$$S = \frac{k_2\sin k_1a[e^{2k_2a}(k_2-ik_3)-e^{2k_2b}(k_2+ik_3)]-k_1\cos k_1a[e^{2k_2a}(k_2-ik_3)+e^{2k_2b}(k_2+ik_3)]}{k_2\sin k_1a[e^{2k_2b}(k_2-ik_3)-e^{2k_2a}(k_2+ik_3)]+k_1\cos k_1a[e^{2k_2b}(k_2-ik_3)+e^{2k_2a}(k_2+ik_3)]}.$$

The coefficients are complex and written as complicated formulae, from which it is difficult to see the dependence on energy. We shall show it in a graphical form. For this purpose we assume the same values of parameters as previously: $m = 1\text{a.u.}$, $a = 2\text{a.u.}$, $V_1 = 5\text{a.u.}$, $\hbar = 1\text{a.u.}$, and additionally $V_2 = 0\text{a.u.}$, and $b = 3\text{a.u.}$

Let us begin with the well. $|A(E)|^2$ which determines the relative probability of finding the particle inside the well is plotted as a function of energy in Fig. 34. In the energy range assumed for consideration, from $V_2 = 0$ to $V_1 = 5$ there are two outstanding maxima. They appear in the vicinities of the only two allowed energies of the particle trapped in the well of the previous case. In

that case the particle could exist inside the well only with energy E_1 or E_2, strictly determined. Now, in the presence of a finite potential barrier the particle can appear inside with an arbitrary energy. However, the probability is negligible for almost all the values of energy. Only in the vicinity of E_1 and E_2 it increases by several orders of magnitude. In the lower part of Fig. 34 both peaks are shown in details. Their profiles are identical. It is known from elsewhere, that their shape is described by a Lorentz curve. These two peaks differ significantly in width measured in the middle of their heights. The one close to E_1 is very narrow. Its half width is $\Gamma_1 = 0.0022$ a. u. The half width of the other one at E_2 is larger $\Gamma_1 = 0.056$ a. u. Half widths are of important physical meaning: they are probabilities of a decay of the state per unit of time, i.e. probabilities that within one unit of time the particle escapes outside the well via tunneling trough the barrier. (The values of widths given above have no special meaning, the choice of parameter values was rather accidental. However, we put stress on them because in what follows they will come out again from another consideration).

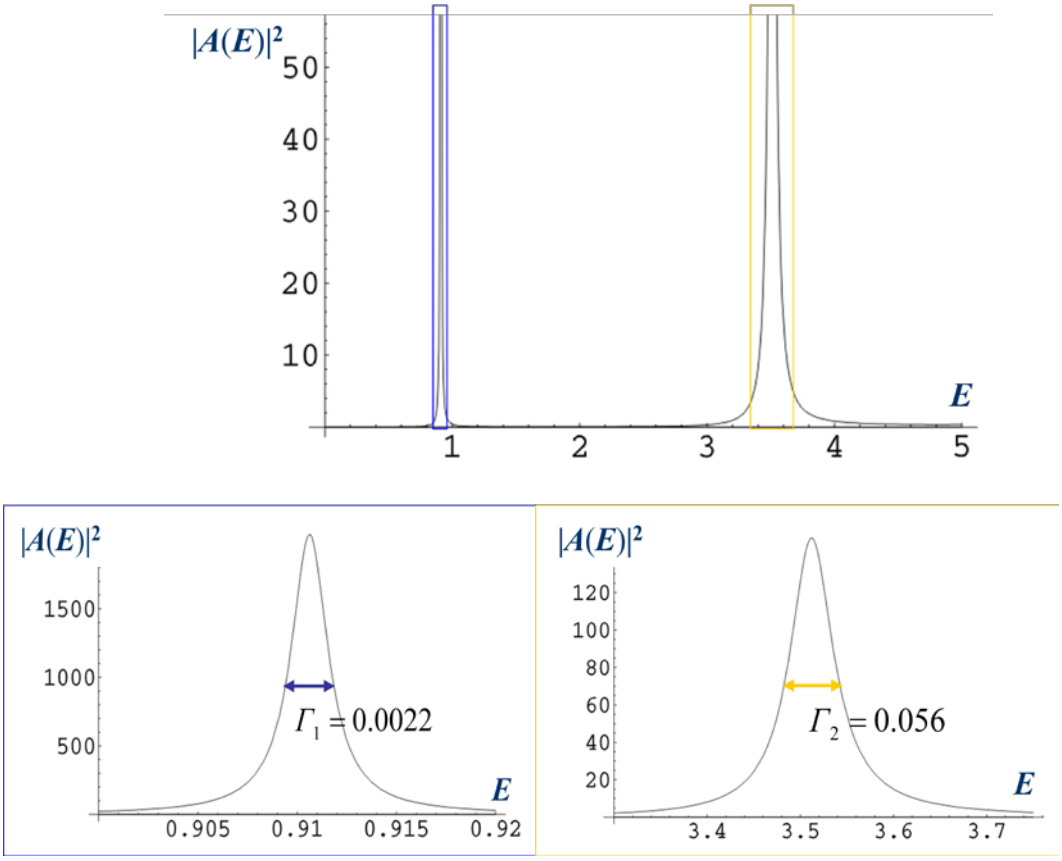


Fig. 34. Quantity $|A(E)|^2$, proportional to the probability of finding the particle inside the well. The upper part covers the whole energy range under consideration, from $V_2 = 0$ to $V_1 = 5$. Two areas marked there with color rectangles are enlarged in the lower panels.

It is evident that trapping the particle in the well is of resonant character. Probability that it happens is large only in a neighborhood of an eigenenergy of the well. Size of the neighborhood is determined by the width Γ. This is analogous to oscillating systems which can be made oscillating with large amplitudes only for frequencies close to the normal frequencies of the system.

If trapping the particle in the well is resonant then also reflection from the barrier (outside) must show a violent resonance behavior at E_1 and E_2. In fact, this can be seen in Fig. 35 where the real part of S^{-1} is plotted versus energy. On an oscillating but smooth background there appear two violent structures at E_1 and E_2. When magnified they exhibit the same character but the structure at E_1 is narrow whereas the other, at $E_{2,}$ is broad. Similar resonant structures should show up in the imaginary part, argument and modulus of S. We will check the modulus in a way which may be surprising.

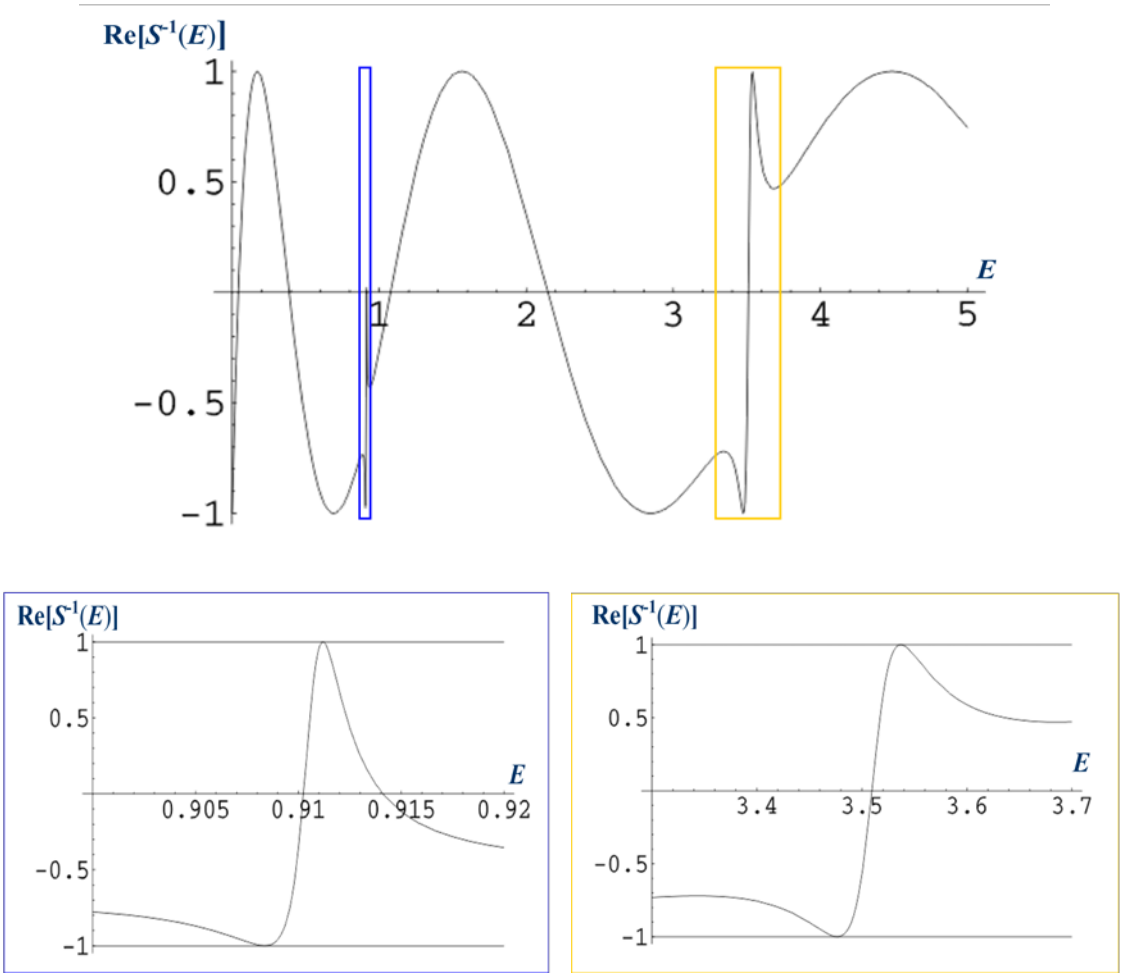


Fig. 35. Real part of S^{-1}. The upper part covers the whole energy range under consideration, from $V_2 = 0$ to $V_1 = 5$. Two areas marked there with color rectangles are magnified in the lower panels.

Let the argument of function $S(E)$ be complex instead of real. From the mathematical point of view this is just natural. On the other hand, from the physical point of view complex values of energy may seem to be unacceptable. Basic courses in quantum physics used to show that the eigenenergies as eigenvalues of a Hermitian operator must be real. That is right but for the case of normalized bound states. It does not hold for resonance state functions which do not vanish at infinity. Thus there is no formal reason to forbid E to be complex. I assure those who feel uneasy about this that we will find a proper physical interpretation for both the real and imaginary parts of such a complex energy.

For technical reason which shall be obvious in a moment we consider $|S|^{-1}$ instead of $|S|$. We use contour plots, Fig.36, because the argument E is two-dimensional. There are two small regions in Fig.36 where $|S|^{-1}$ drops down evidently. They are located near $\mathrm{Re}(E) = E_1$ and $\mathrm{Re}(E) = E_2$, although not at the real axis, $\mathrm{Im}(E) = 0$. They lie below the axis. One of the minima is placed at $\breve{E}_1 \cong (0.91062 - i0.001125)$ a. u. , and the other at $\breve{E}_2 \cong (3.51168 - i0.02818)$ a. u. It is not unexpected that the real parts of these particular complex numbers are close to the eigenenergies of the particle in the well, $E_1 = 0.91155$ a. u. i $E_2 = 3.5005$ a. u. Let us note, however, that they are not the same. The differences are not due to any inaccuracy or approximation. They follow from the fact that the system has been modified. They are considered as „continuum shifts" due to the energy continuum open for the particle on the outer side of the barrier (in the well-barrier system all the energy values above $V_2 = 0$ are allowed).

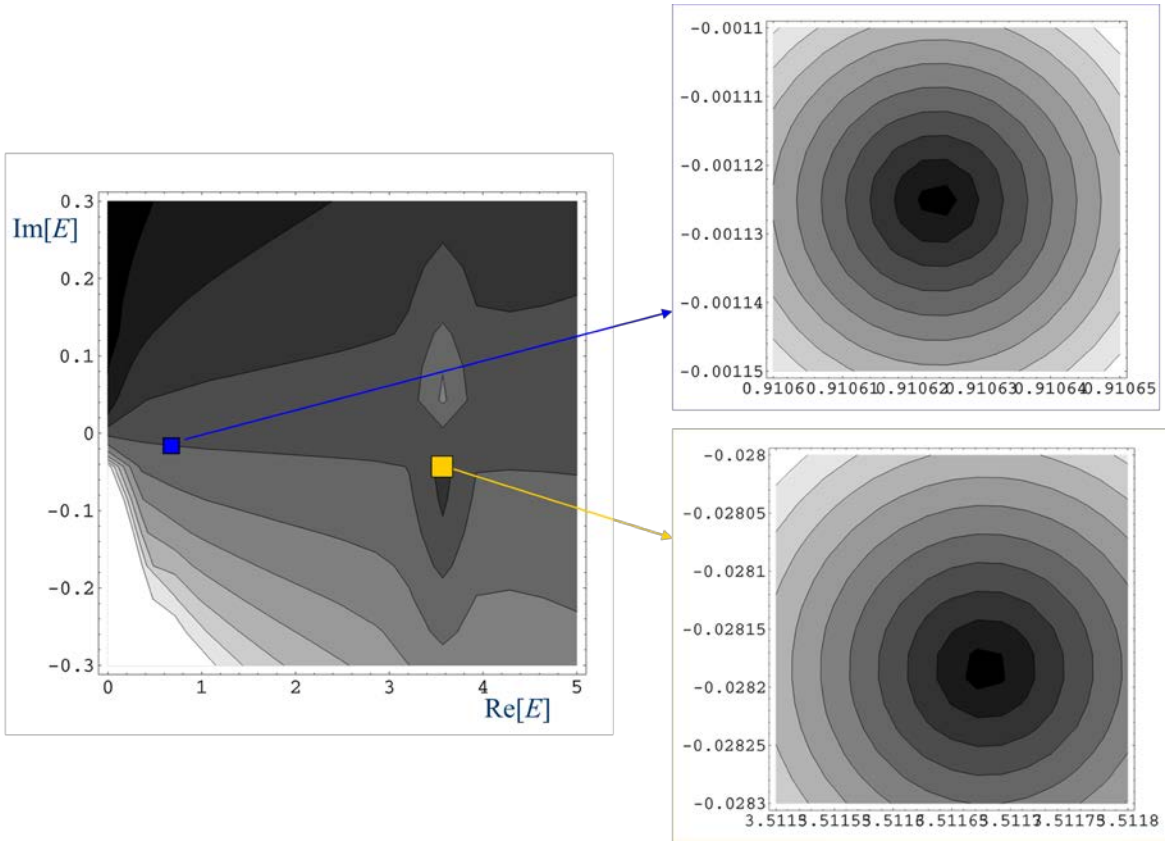


Fig. 36. Inverse of the modulus of scattering matrix, $|S|^{-1}$, as a function of complex variable E. The darker tone the lower value of $|S|^{-1}$. The plot on the left covers the energy range under consideration, $V_2 = 0 < \mathrm{Re}(E) < V_1 = 5$. The regions marked with color rectangles are magnified on the right.

What about imaginary parts? Let us note that when doubled they are equal to the half widths of the resonant structures in $|A(E)|^2$. Thus we have fund specific complex eigenvalues of the Hamiltonian whose real parts correspond to positions of the resonance levels, E_r, and the imaginary parts are equal to half of the half widths of the levels, Γ: $\breve{E} = E_r - i\frac{\Gamma}{2}$.

How deep are the minima in Fig. 36? Do they reach zero? We use 3D plots in Figs. 37 and 38 to examine this. They show $\log|S^{-1}(E)|$ for the lower and higher resonance, respectively. Looking at

them from different perspectives one can read precise position of the minima. As shown in these pictures they reach as low as 10^{-11}. Actually, these minima are points where $S^{-1} = 0$. (That is why we preferred to plot S^{-1} instead of S). This means that these points there is no incoming wave in the state function. There is the outgoing wave only. Such specific asymptotic conditions can be adopted as a definition of resonance states. Thus resonances are pseudobound states which can decay. Therefore their wave functions contain exclusively the outgoing wave which describes the escaping particle.

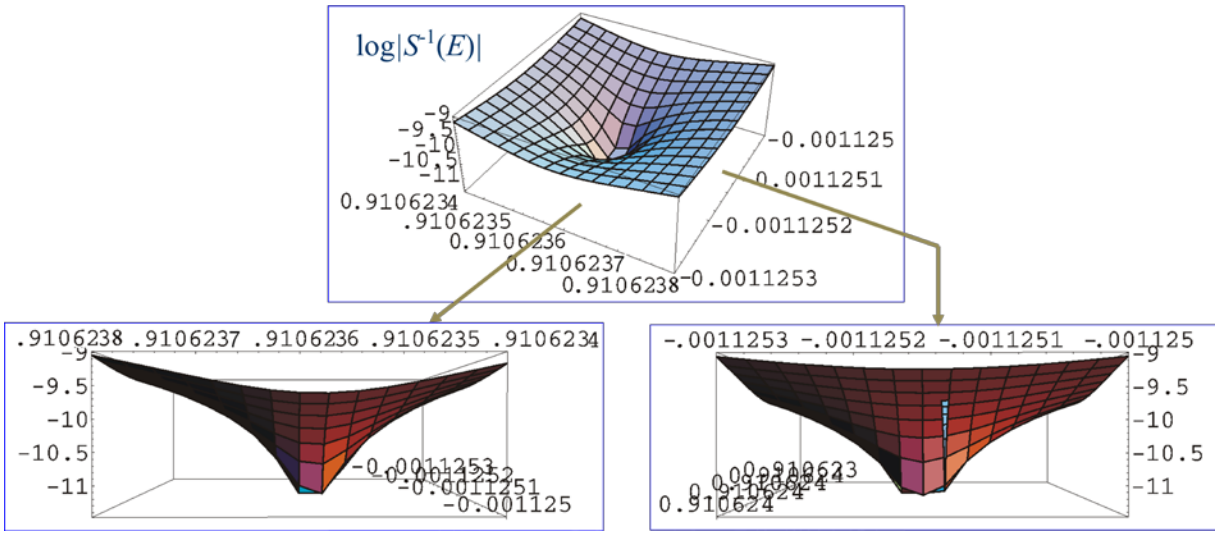


Fig. 37. $\log|S^{-1}|$ as a function of complex variable E in a small vicinity of the lower lying resonance of the well-barrier system under consideration; the lower panels present the views along the imaginary axis (on the left) and along the real axis (on the right) so that the real and imaginary parts, respectively, of the minimum point can be clearly seen and deciphered.

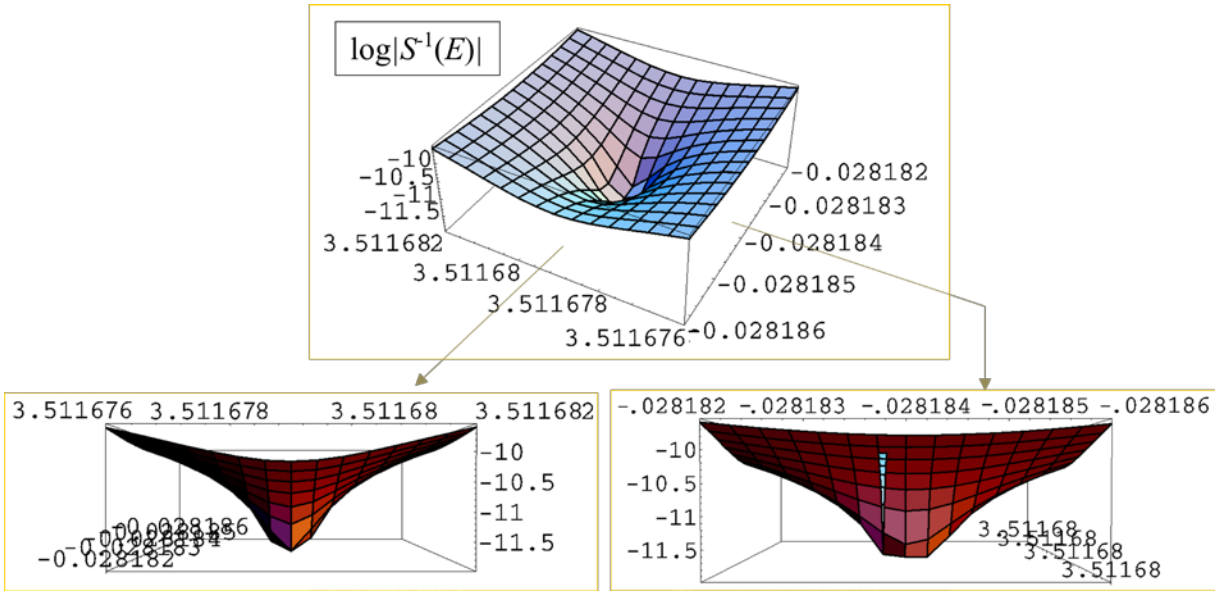


Fig. 38. $\log|S^{-1}|$ as a function of complex variable E in a small vicinity of the upper resonance of the well-barrier system under consideration; the lower panels present the views along the imaginary axis (on the left) and along the real axis (on the right) so that the real and imaginary parts, respectively, of the minimum point can be clearly seen and read.

Assembling two resonance parameters, the energy position and width, in one complex "energy" $\check{E} = E_r - i\frac{\Gamma}{2}$ has an important meaning in the context of the state evolution, it offers good explanation of decay of the resonance state. According to one of the quantum-mechanics postulates given in Chapter I, the evolution of every states runs in accordance to the Schroedinger equation dependent on time. Particularly, the eigenstates of the Hamiltonian evolve as follows

$$\psi(x,t) = e^{-\frac{i}{\hbar}Et}\Psi(x), \text{ where } \hat{H}\Psi(x) = E\Psi(x).$$

Then the probability distribution $|\psi(x,t)|^2$ for the resonance state changes in time exponentially

$$|\psi(x,t)|^2 = \left|e^{-\frac{i}{\hbar}(E_r - i\Gamma/2)t}\Psi(x)\right|^2 = e^{-\frac{\Gamma}{\hbar}t}|\Psi(x)|^2 .$$

Hence in the resonance state the probability of finding the particle inside the well decreases in time exponentially. It gets e times smaller after $\tau = \frac{\hbar}{\Gamma}$, which is called the everage live time.

Approximate methods of finding positions and widths of resonances

Two variational methods are presented: the stabilization method and the complex coordinate rotation mathod. Both consist in expanding an approximate resonance wavefunction in a given basis of normalized functions and in optimization with respect to an extra nonlinear parameter specific for the method. From the practical computation point of view, the most difficult problem is the background energy continuum corresponding to unbound states in which the resonance energy is diluted. Because of this, variational computations relying in expansion in a basis set provide a set of real energies from which it is difficult to select the right ones corresponding to resonance states. Another problem is that resonance wavefunctions are not normalizable to unity. Therefore expanding them in a set of normalized functions may be a serious approximation.

The stabilization method allows for selection of the right roots corresponding to resonances via specific optimization. It is however encumbered with the use of square-integrable basis function to represent resonance-state function. The complex coordinate rotation method allows for finding two parameters of the resonance level: the position and the width as the real and imaginary parts of a complex eigenvalue. It uses transformation which makes the resonant wavefunction square integrable (normalizable to unity) so that using of square-integrable function basis is legitimate.

Stabilization method

We have got a basis of N functions $\{\varphi_i(\vec{r})\}_{i=1}^{N}$. They are normalized and do not contain free parameters. If a resonance function is expressed in such a basis then the only variational parameters (allowing for optimization of the expansion so that the Schroedinger equation is satisfied with the best possible accuracy) are the linear expansion coefficients. We introduce an extra nonlinear parameter, one for all the basis functions, which will give us a tool for identification of resonance roots. This is a real parameter α of scaling coordinates, $\vec{r} \to \alpha\vec{r}$. We introduce the scaled coordinates into the basis functions obtaining a basis of new functions which can be tuned with α, $\{\varphi_i^{\alpha}(\vec{r}) \stackrel{\text{def}}{=} \varphi_i(\alpha\vec{r})\}_{i=1}^{N}$. This basis is used to expand the resonance wavefunction

$$\psi(\vec{r}; c_1, \cdots c_N, \alpha) = \sum_{i=1}^{N} c_i \varphi_i^{\alpha}(\vec{r}).$$

The energy corresponding to ψ is to be fund as the expectation value of the hamiltonian:

$$E(c_1, \cdots c_N, \alpha) = \frac{\langle \psi(\vec{r}; c_1, \cdots c_N, \alpha) | \hat{H} | \psi(\vec{r}; c_1, \cdots c_N, \alpha) \rangle}{\langle \psi(\vec{r}; c_1, \cdots c_N, \alpha) | \psi(\vec{r} c_1, \cdots c_N, \alpha) \rangle}.$$

As satisfaction of Schroedinger equations is considered the optimal $\psi(\vec{r}; c_1, \cdots c_N, \alpha)$ is that one for which

$$\left.\frac{\partial \mathrm{E}(\mathrm{c}_1, \cdots \mathrm{c}_\mathrm{N}, \alpha)}{\partial \mathrm{c}_\mathrm{i}}\right|_{\mathrm{c_i}^{\mathrm{opt}}} = 0 \text{ and } \left.\frac{\partial E(c_1, \cdots c_N, \alpha)}{\partial \alpha}\right|_{\alpha^{opt}} = 0.$$

Let us work with the former condition. Since the expectation value depends on c_i in a way which is determined analytically we can consider explicitly the condition obtaining the following matrix equation:

$$\mathbf{H}(\alpha)\mathbf{C}^{\mathrm{opt}}(\alpha) = E(\alpha)\mathbf{S}(\alpha)\mathbf{C}^{\mathrm{opt}}(\alpha),$$

where $\mathbf{C}^{\mathrm{opt}}(\alpha)$ is a column vector of constituents $c_i{}^{opt}$, $\mathbf{H}(\alpha)$ is the so called Hamiltonian matrix built of elements $H_{ij}(\alpha) = \langle \varphi_i^\alpha | \hat{H} | \varphi_j^\alpha \rangle$, and $\mathbf{S}(\alpha)$ is a matrix of scalar products of basis functions $S_{ij}(\alpha) = \langle \varphi_i^\alpha | \varphi_j^\alpha \rangle$. So we need to solve an algebraic problem (generalized secular equation) for matrices of size $N \times N$. We obtain N solutions: N real eigenvalues $E_k(\alpha)$. Some of them may correspond to bound states, many are associated with unbound states and some with resonance states. The latter are of our interest. Unfortunately we are not able to distinguish them from those belonging to unbound states.

Now we take care of condition $\left.\frac{\partial E(c_1,\cdots c_N,\alpha)}{\partial \alpha}\right|_{\alpha^{opt}} = 0$. Since the dependence of solutions $E_k\left(c_1^{opt},\cdots c_N^{opt};\alpha\right)$ on parameter α cannot be established analytically we have to find the optimal point for which $\frac{\partial E_k(\alpha)}{\partial \alpha} = 0$ in a numerical way: 1) We perform series of computations for many equidistant values of α; 2) We renumber the eingenvalues $E_k(\alpha)$ obtained for a given α from the smallest one to the biggest one; 3) We plot every eigenvalue E_k as a function of α. An example of picture we should obtain is shown in Fig. 39.

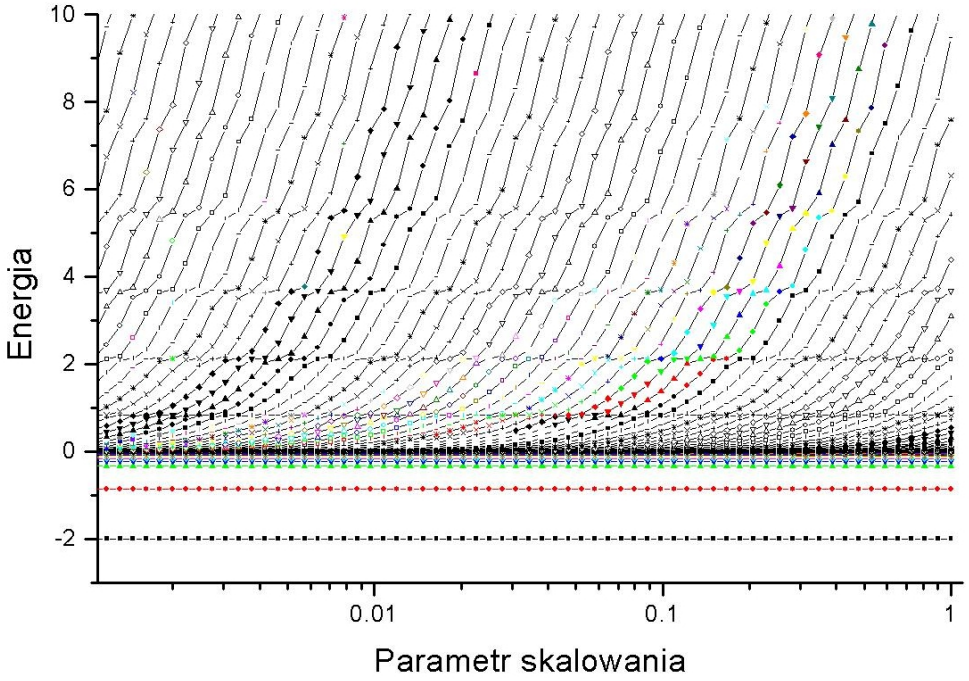


Fig. 39. Stabilization of energy of some system with respect to the coordinate scaling parameter α. The horizontal lines below zero correspond to bound states. (They are horizontal because the basis perfectly reproduces the resonant wave functions and corresponding to them energies of bound states, independently of α). Energy continuum starts at zero. Above zero every root increases with α increasing. However every one stabilizes at some levels (corresponding to resonances) for a small range of α, until next root comes to the level and overtakes the role of representing the energy of resonance..

As the functions expand due to coordinate scaling $\alpha\vec{r}$, every E_k decrease monotonically toward the threshold where the energy continuum starts (placed at zero in Fig. 39). However at some levels they stabilize. These levels are attributed to resonances. Their positions, i.e. resonance

energies, can be read with some accuracy from the plot or obtained numerically $\frac{dE_k(\alpha)}{d\alpha} = 0$. They can be determined within a procedure described below which except for the energy positions determines the widths of resonances.

Systematic scaling of coordinates indicates systematic changes of eigenvalues. It is kind of scanning of the energy continuum. If the number of scaling parameter α and number of basis functions are big enough then one obtains quite a number of eigenvalues whose density with respect to energy represents well density of states of the system. In practice we represent it by a histogram: we divide the energy range under consideration into small segments of equal width, we count the roots (for every value α) which have fallen in given segments. It is rather obvious that in such a histogram there appear maxima in the neighborhood of levels of stabilization, i.e. in the neighborhood resonance levels. One of them, the one corresponding to the level of about 2.1 a.u. of Fig. 39 is shown in Fig. 40.

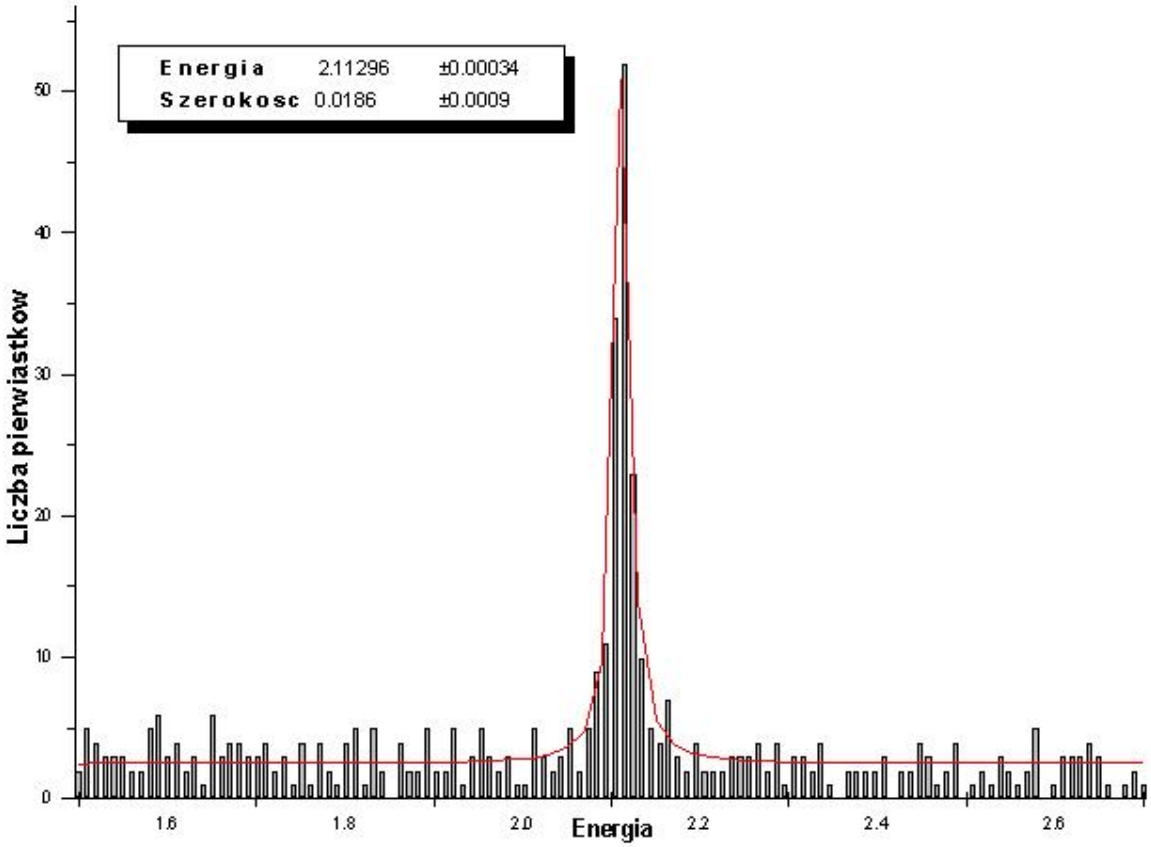


Fig. 40. Density of states in the continuum range in the vicinity of one of resonance levels of Fig. 39 represented by distribution of roots obtained by means of the stabilization method. The red curve is the Lorentz profile fitted to the histogram. The resonance energy position and width are fitting parameters. Their optimal values are given.

Actually these maxima are related to the maxima shown in Fig. 34. They have resonance profiles which can be described by the Lorentz curves. Therefore Lorentz curves

$$\rho(E) = \rho_o + \frac{D}{(E - E_r)^2 + \frac{1}{4}\Gamma^2} .$$

are fitted to the distribution of roots. From among four of fitting parameters (ρ_o, D, E_r, Γ) two are especially important for determination of a resonance state. These are the position of the resonance level, E_r, and its width Γ. Finally their values obtained by fitting the Lorentz curve to the resonance structure in the density of states (histogram) are considered to be the most trustful values of resonance parameter that can be obtained by the stabilization method.

Complex coordinate rotation method

The complex coordinate rotation method is similar the stabilization method. Instead of Real scaling a complex scaling is used; the scaling factor is $\alpha = e^{i\theta}$, where $\theta > 0$ is a real parameter. Transformation $\vec{r} \rightarrow e^{i\theta}\vec{r}$ is called complex coordinate rotation or complex scaling. Hence the method is also known as the complex scaling method.

The method is known in several variants. In the simplest version the complex coordinate transformation is applied to the coordinates in the Hamiltonian, not in the wavefunctions. It has been mathematically proven that the spectrum of eigenvalues of the transformed Hamiltonian $\hat{H}(e^{i\theta}\vec{r})$ is in strict relation to the one of original operator and has physical interpretation. In particular:

1) The discrete eigenvalues of the original Hamiltonian $\hat{H}(\vec{r})$ corresponding to bound states are preserved under the complex rotation of coordinates.
2) The continuous energy spectrum corresponding to unbound states is rotated with respect to its threshold into the complex plane by -2θ.
3) The energies of resonances are complex, separated from the continuum and in a range of θ independent of it. Their imaginary parts are negative.

The above properties are illustrated in Fig. 41.

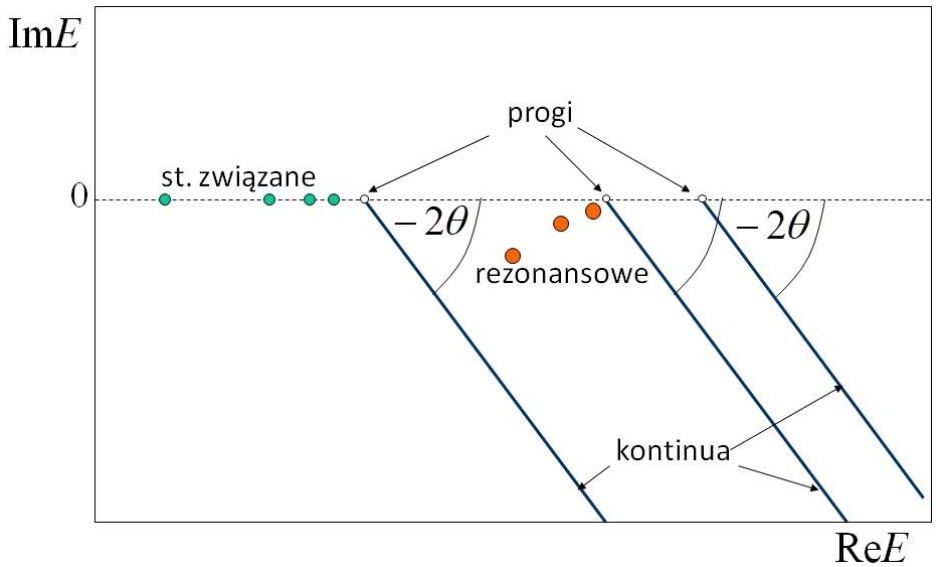


Fig. 41. The eigenspetrum of the complex coordinate rotated Hamiltonian. The energies corresponding to bound states and continuum thresholds are affected by the rotation. The continuum spectrum semilines are rotated by 2θ to the lower halfplane. The resonance eigenvalues appear in the lower halfplane and are separated from the continua.

An advantage of the method is fact that both resonance parameters are fund in one complex number, $\breve{E} = E_r - i\frac{\Gamma}{2}$, which can be quite easily distinguished from the resonance energies of continuum. Another important advantage of the complex coordinate rotation transformation is that the resonance wavefunctions become square-integrable. Therefore using basis sets of square-integrable functions to represent resonances is definitely appropriate in contradiction to the stabilization method.

In practice, because of a finite size of the space where we are searching for resonance state approximation the approximate resonance energy depends slightly on θ. So the rotation angle is used as a variational parameter.

Finally, a recipe for complex coordinate rotation calculation is the following:

1) Rotate the coordinates in the Hamiltonian so as to obtain $\hat{H}(e^{i\theta}\vec{r})$.
2) Search for a solution of eigenequation of operator $\hat{H}(e^{i\theta}\vec{r})$ in the form of expansion $\psi(\vec{r}; c_1, \cdots c_N) = \sum_{i=1}^{N} c_i \varphi_i(\vec{r})$, where the basis functions φ_i are given.
3) Energy corresponding to the above expansion is to be found as optimized expectation value
$$E(c_1, \cdots c_N, \theta) = \frac{\langle \psi(\vec{r}; c_1, \cdots c_N) | \hat{H}(e^{i\theta}\vec{r}) | \psi(\vec{r}; c_1, \cdots c_N) \rangle}{\langle \psi(\vec{r}; c_1, \cdots c_N) | \psi(\vec{r} c_1, \cdots c_N) \rangle}.$$
4) Optimal sets of coefficient c_i and corresponding to them energies are obtained by solving the matrix equation
$$\mathbf{H}(\theta)\mathbf{C}^{\mathrm{opt}}(\theta) = E(\theta)\mathbf{S}\mathbf{C}^{\mathrm{opt}}(\theta),$$
where $\mathbf{C}^{\mathrm{opt}}(\theta)$ is a column vector whose components are ${c_i}^{opt}(\theta)$,
$\mathbf{H}(\theta)$ is the complex rotated Hamiltonian matrix of built elements
$H_{ij}(\theta) = \langle \varphi_i | \hat{H}(e^{i\theta}\vec{r}) | \varphi_j \rangle$ and $\mathbf{S}$ is the matrix of scalar products of basis functions $S_{ij} = \langle \varphi_i | \varphi_j \rangle$. Because matrix $\mathbf{H}(\theta)$is nonHermitian the eigenvalues $E(\theta)$ obtained from the equation are complex**.**
5) Most of the eigenvalues should depend on θ in the way characteristic for unbound states: they should rotate into the complex plane by angle -2θ. Only few eigenvalues are expected to be independent of θ or rather depend on θ very slowly. Repeat solving of equation of 4) for different values of θ, so as to find the optimal resonance energies according to $\left.\frac{dE(\theta)}{d\theta}\right|_{\theta^{opt}} = 0$. This means that from among eigenvalues obtained as explained in 4) the ones which are the most stable against variation of θ should be chosen as the optimal approximations of resonance energies.
6) The final answer is: the real parts of optimal energy eigenvalues obtained as described above are the positions of resonance states and the imaginary parts of them are half widths of resonances.

An example of results obtained by the complex coordinate rotation method is presented in Fig. 42. All the obtained eigenvalues of real part in the range of interest are put in the complex plane. The results were computed by using three slightly different basis sets and for thirty values of θ forming an arithmetic sequence from 0 to $\frac{\pi}{4}$. There appear θ-trajectories (on which the eigenvalues move when θ is changing) of two types. Trajectories forming arcs correspond to the continuum energies rotating systematically with θ increasing. They definitely differ from three θ-trajectories

(one for each of three basis) of the other type which demonstrate much slower dependence of roots on θ. Although these three θ-trajectories start at three evidently different points on real axis for $\theta = 0$, they converge to one point at which the roots stabilize against variation of θ. This is the optimal complex energy of the resonance.

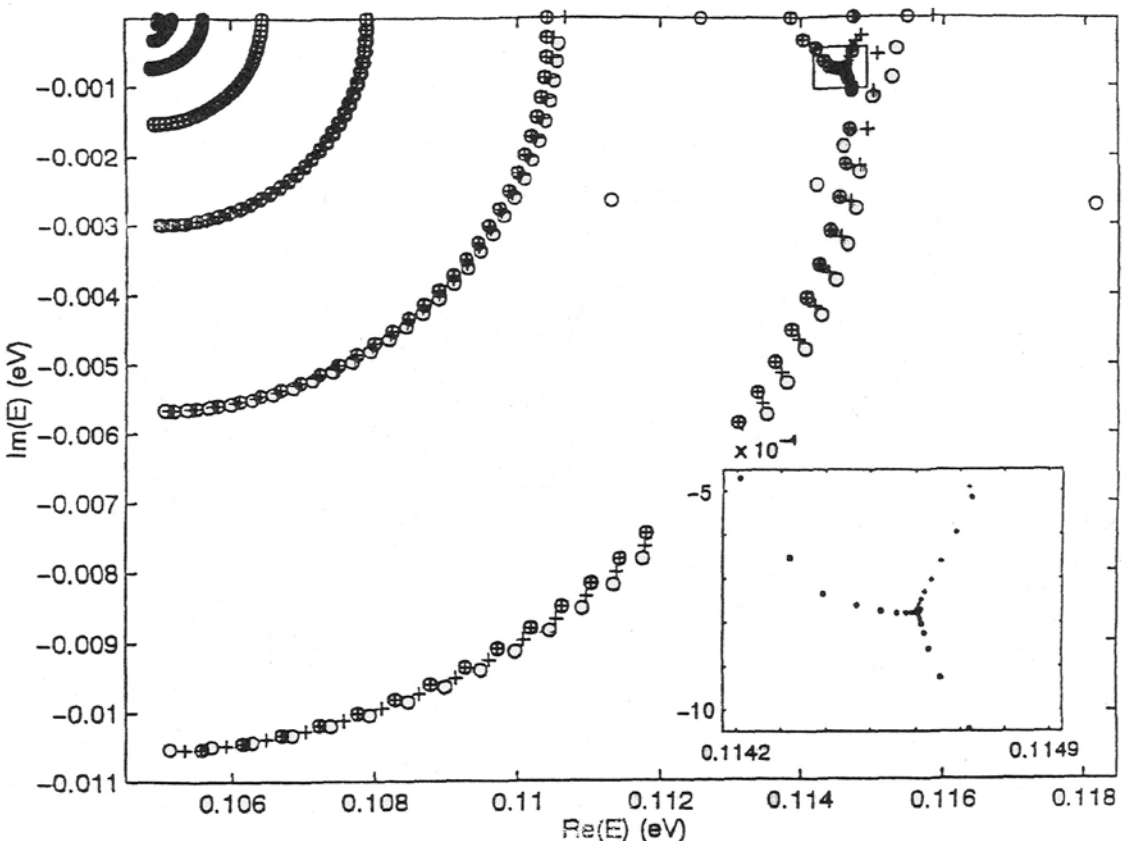


Fig. 42. θ-trajektories obtained for a nanostructure of symmetric system of semiconductor nanolayers: two barriers and a well between them [M. Bylicki, R. Oszwałdowski, W. Jaskólski, J. Phys.: Condensed Matter 8 (1996) 6393]. The region of convergence of resonance θ-trajectories marked with the rectangle is magnified in the inset.

Self-assembled quantum dots as an example of strained system

Self-assembled quantum dots

Molecular beam epitaxy (Fig. 43) is one of the most important nanostructure growth methods. In this approach, in the ultra-high vacuum conditions, subsequent layers of atoms or molecules are deposited on a properly prepared substrate. Temperature stabilized evaporators, so called effusion cells (also known as Knudsen cells), are sources of the atomic or molecular beams (hence the name: MBE). MBE combined with RHEED spectroscopy (Reflection High-Energy Electron Diffraction) allow for a precise control of layer deposition with a precision up to a single atomic monolayer.

Fig. 43. MBE chamber (From the Internet).

The chemical composition can be modified during the growth process, by shutting or opening effusion cells. For example, after deposition of several gallium arsenide (GaAs) monolayers one can follow with a different compound, e.g. aluminum arsenide (AlAs). Both the substrate and the deposited material have different electronic properties, in particular different the energy gap, thus using MBE it is possible to grow nanostructures such as quantum wells or superlattices.

Not all semiconductor compounds can be mixed (grown) together in a defect (dislocation) free manner. In particular, crystal lattice constants (related to bond lengths) of the substrate and the deposited material cannot vary too much. Earlier, we have mentioned GaAs and AlAs, which are good example of lattice matched materials with less than 0.2% lattice constant difference.

On the other hand, a quite interesting situation occurs for a customarily used material combinations, e.g. for InAs/InP the lattice mismatch is about 3%, while for InAs/GaAs it is as much as 7%. Surely, because of the lattice mismatch, there has to be a process of bonds alignment (Fig. 44) between the substrate and the deposited layer.

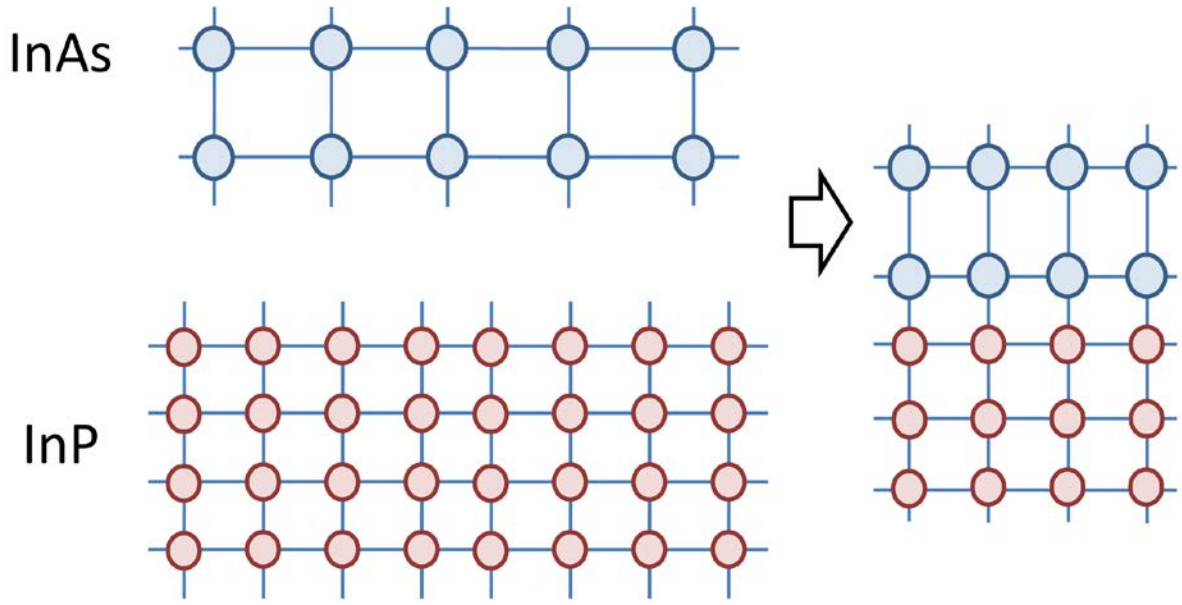


Fig. 44. Schematics illustrating bond lengths alignment for an epitaxially deposited layer of lattice mismatch materials.

For a combination of materials with a substantial lattice mismatch, in the process of Stransky-Kranstanov growth, the first deposited layer (so called wetting layer) adopts its in-plane lattice constants to match the substrate lattice constant (it is so called pseudo-morph). For a case shown on Fig. 45, the layer material (InAs) of larger lattice constant is squeezed in the junction plane with a simultaneous expansion in the growth (out-of-plane) direction, while the substrate remains virtually unmodified. In such case we say that the layer is highly deformed, or strained, because interatomic distances are modified compared to unstrained bulk crystal. More quantitative description of strain will be presented later.

Further growth of lattice mismatch deposit would lead to strong strain of the newly formed layer. In practice, in the Stransky-Kranstanov growth mode, it is more energetically favorable to minimize strain energy by a formation on non-uniform “islands” of deposited material instead of a homogenous layer (Fig. 45). These „islands” have diameters on the order of 10 to 30 nanometer and height varying from 1 to 5 nanometer and are located in a non-regular, random-like, pattern on a substrate surface. The shape is quite similar to lens, while “island” parameters, such as chemical composition, dimension or surface density can be controlled to a large degree by growth conditions.

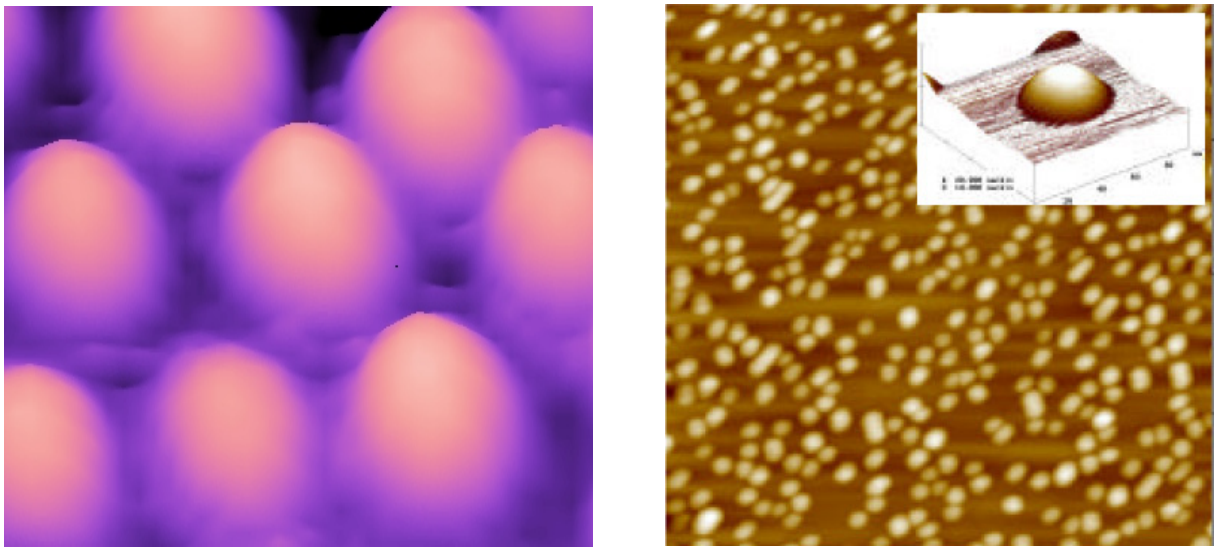


Fig. 45. Atomic force microscope pictures showing formation of „islands" (quantum dots) on a lattice mismatched substrate (from the Internet).

Usually, once „islands" are grown the process is stopped and the substrate material is deposited again. In effect, one gets a structure composed of semiconductor intrusions surrounded by another material semiconductor matrix (Fig. 46 and Fig. 57).

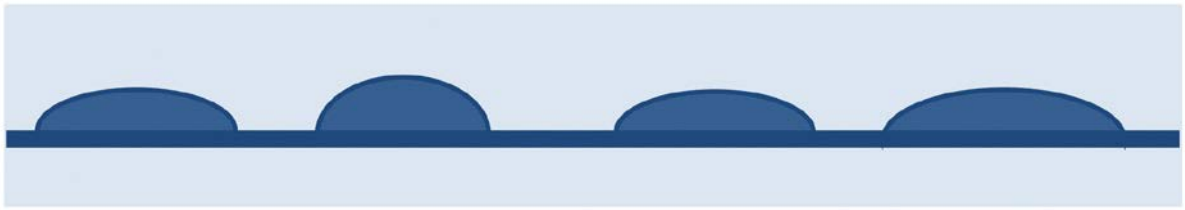


Fig. 46. Schematic cross-section of quantum dot layer.

A proper selection of chemical composition, e.g. large band gap surrounding matrix (GaAs or InP) and small band gap „island" material (InAs) leads to spatial confinement of charge carriers (electrons and holes) in the "island" volume. These kinds of nanostructures are called self-assembled quantum dots, where „self-assembled" corresponds to the growth process, while "quantum dot" corresponds to small volume spatially limited in all three dimensions.

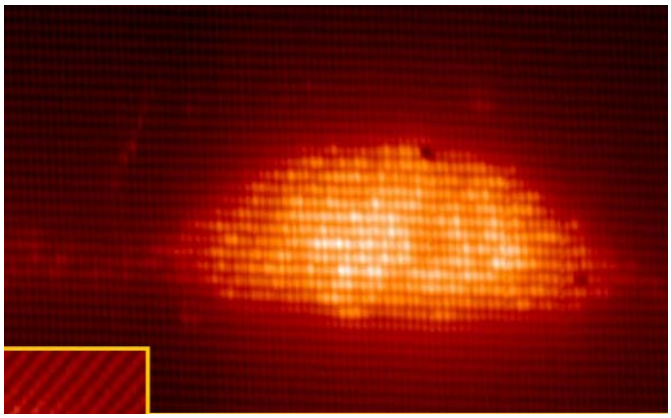


Fig. 47. Cross-section scanning tunneling microscope picture (X-STM) of self-assembled InAs/GaAs lens type quantum dot (from: papers by P.M. Koenraad et al.).

The control of growth parameters allows also for tailoring quantum dot shape different then lens. In particular, indium-flush technique allows for growth of disc type self-assembled quantum dots (Fig. 48).

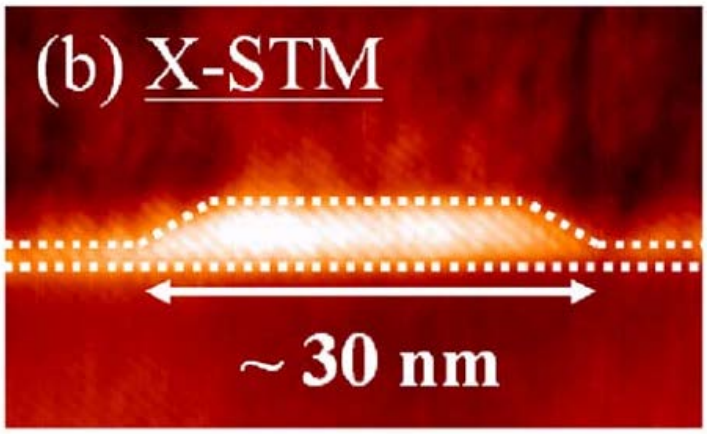


Fig. 48. Cross-section scanning tunneling microscope picture (X-STM) of self-assembled InAs/GaAs disc type quantum dot (from: papers by P.M. Koenraad et al.).

Apart from single quantum it is possible to grow double quantum dots (Fig. 49) or even vertically stacked multiple quantum dots systems (Fig. 50).

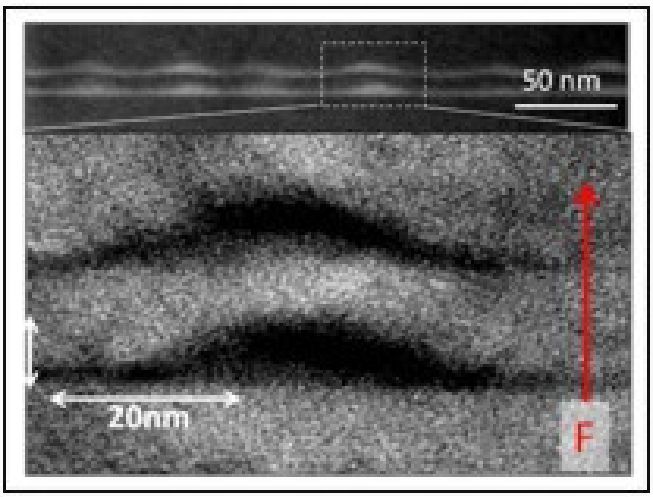


Fig. 49. Transmission electron microscopy (TEM) cross-section of a double InAs/GaAs quantum dot system (from the Internet).

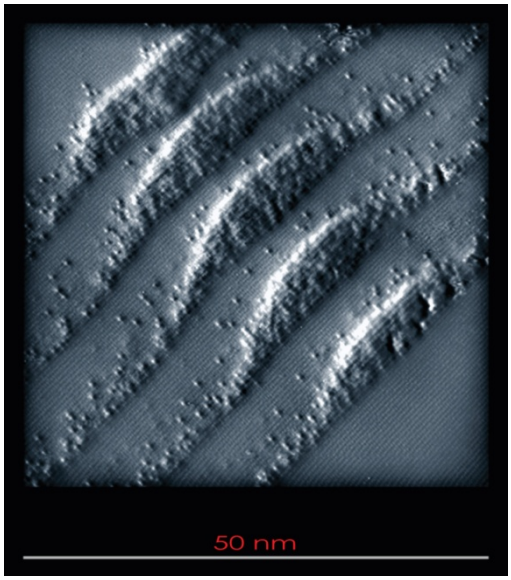


Fig. 50. X-STM picture of a vertical stack of InAs/GaAs self-assembled quantum dots (from the Internet).

Spectral properties of single quantum dots resemble that of natural atom (discreet energy spectra), hence a term “artificial atoms” is sometimes used to describe them. By analogy, a system of two or more quantum dots is sometimes called an „artificial molecule”.

6. Types and sources of strain

Strain tensor

So far we have discussed self-assembled quantum dots which form due to lattice mismatch of the quantum dot and the substrate materials. The wetting later and the quantum dot material is significantly strained (Fig. 51), thus one can suspect that strain effects will play a major role and affect quantum dot spectral properties.

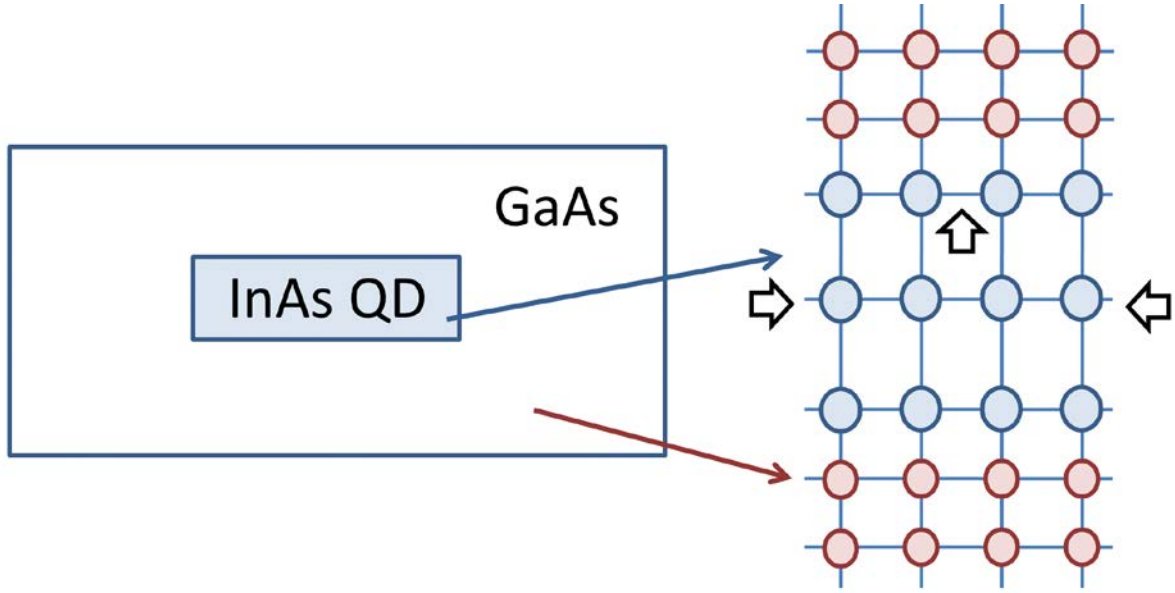


Fig. 51. Schematics of interatomic distance deformation for a self-assembled InAs quantum dot embedded in a GaAs matrix.

Self-assembled quantum dot forms a quite complicated, three-dimensional system. For simplicity, let us first consider a quantitative description of a deformation (strain) for a much less complex one-dimensional case.

Let us start with the example of a metal string with a progressively increasing weight attached. (Fig. 52). According to Hooke's law and our intuition the string deformation degree depends on its kind (e.g. spring wire material and thickness) and is proportional to the attached weight.

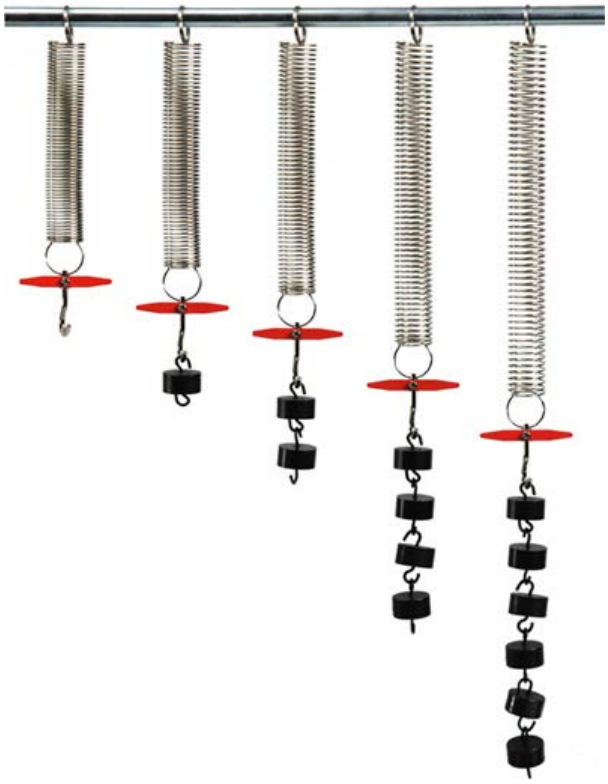
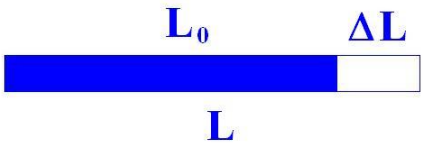


Fig. 52. Metal spring deformation (strain) as a function of increasing weight.

For a sake of our considerations, we are not interested in forces involved or the stresses that are inflicted upon the spring. We only want to measure the occurring deformation. Let us introduce then a definition of strain by a comparison of unstrained and deformed spring lengths (Fig. 53) divided by the unstrained length.



$$\varepsilon = \frac{L - L_0}{L_0} = \frac{\Delta L}{L_0}$$

Fig. 53. Schematics and a definition of strain for one-dimensional system (from the Internet).

The advantage of such definition is that for an unstrained system ε is exactly zero. The sign of ε on the other hand describes the character of strain. ε>0 corresponds to the tensile strain, while ε<0 corresponds to the compressive strain. Additionally the division by the unstrained length allows to introduce a unified strain measure of objects of different lengths. For example, one meter long steel wire stretched by 1 centimeter is under the same (1%) strain, as one kilometer long wire stretched by 10 meters.

The definition of ε is fully satisfactory a deformation of a one dimensional character. The problem occurs however for a more complicated spatial deformation (Fig. 53b).

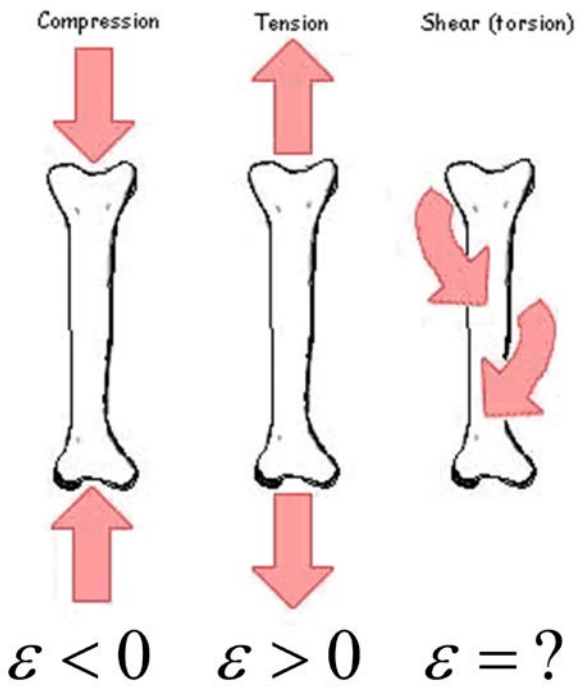


Fig. 53b. Schematics of a strain sign for a one-dimensional case deformation and a failure of a simple 1D description for a more complicated shear deformation.

In general case, one needs to account for the existence of other two dimensions (degrees of freedom). Fig. 54 shows many (often complicated) ways in which a unit cube can be deformed. It is thus possible to squeeze or stretch along different cube edges or to apply a shear or torsion-like deformation.

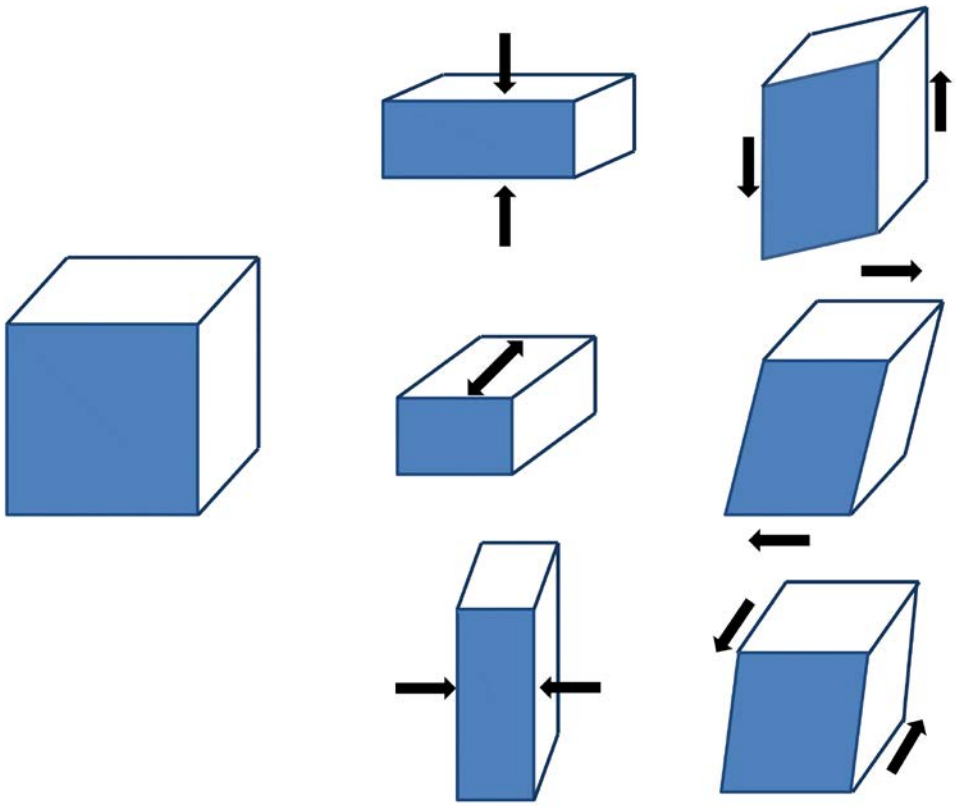


Fig. 54. Schematics of different unit cube deformations.

All possible unit cube deformation can be described by a single, yet tensor (matrix) quantity, so called strain tensor:

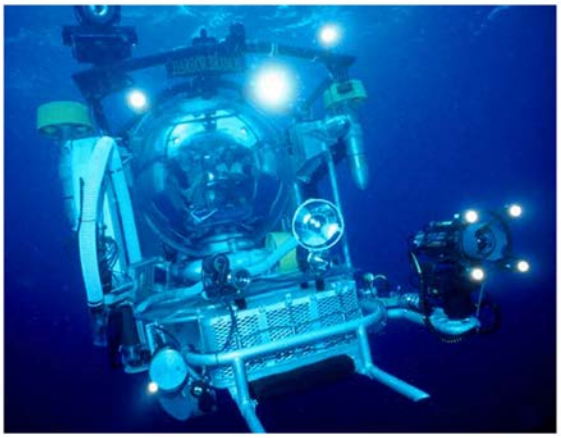
$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix}$$

Strain tensor is symmetric ($\varepsilon_{ij}=\varepsilon_{ji}$) and, what is important, describes unit cube shape deformations, and not its rotations (given in terms of a rotation matrix) or spatial translations.

Types of strain

Without going into mathematical details, let us illustrate (Fig. 55) main types of strain and the corresponding strain tensor form. For example, a hydrostatic deformation (e.g. deformation of a submerged spherical submarine cabin) corresponds to uniform compression ($\varepsilon<0$), with no preferential spatial direction ($\varepsilon_{xx}=\varepsilon_{yy}=\varepsilon_{zz}$). Because under the hydrostatic strain the shape remains unchanged thus off-diagonal strain tensor elements are exactly zero.

On the other hand, for an axial (uniaxial, biaxial) strain, a compression in one direction (e.g. $\varepsilon_{zz}<0$) is assisted by an expansion in the other two directions ($\varepsilon_{xx}=\varepsilon_{yy}>0$). By choosing coordinate system axes along deformation directions we again obtain a diagonal form of strain tensor, however with different diagonal matrix elements (e.g. $\varepsilon_{xx}= \varepsilon_{yy}\neq \varepsilon_{zz}$).



Hydrostatic

$$\begin{pmatrix} \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{pmatrix}$$

Biaxial (uniaxial)

$$\begin{pmatrix} \varepsilon_{\leftrightarrow} & 0 & 0 \\ 0 & \varepsilon_{\leftrightarrow} & 0 \\ 0 & 0 & \varepsilon_{\updownarrow} \end{pmatrix}$$

Fig. 55. Examples of a hydrostatic and axial strain and the corresponding strain tensor form (pictures from Internet).

Yet another is a shear strain, as shown on Fig. 5.56, in this case the structure linear dimensions are not changed, and a shape deformation is created by applying a force parallel to the surface. It should be pointed thought that all three kinds of strain can occur simultaneously. For example, in case of „typical" self-assembled quantum dots and dominant role is played by both hydrostatic and biaxial strain, while shear strain is less important.

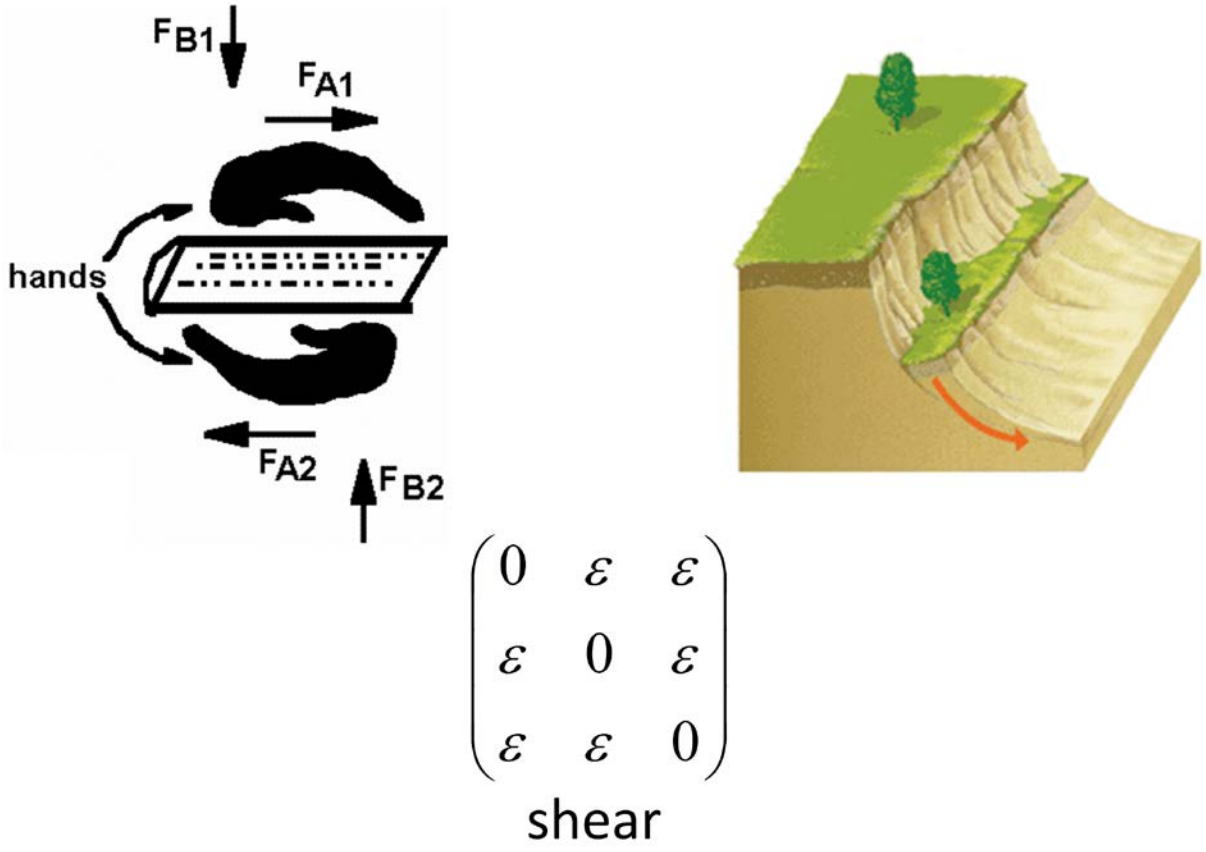


Fig. 56. Schematics showing shear strain (from the Internet).

Using strain tensor elements one can construct quantities useful to categorized different sorts of strain. In particular, the trace of the strain tensor (sum of diagonal matrix elements) is a quantitative measure of hydrostatic strain:

$$Tr(\varepsilon) = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}$$

For a „small" deformation Tr(ε) describes the unit cube volume change:

$$V \approx V_0\left[1 + Tr(\varepsilon)\right]$$

Negative value of Tr(ε) describes thus a compressive strain and positive trace value corresponds to a tensile strain case. The Tr(ε) is further even more useful as it is invariant under coordinate system transformation.

One can also measure axial strain, e.g. using following definition:

$$B(\varepsilon) = \sqrt{\left(\varepsilon_{xx} - \varepsilon_{yy}\right)^2 + \left(\varepsilon_{yy} - \varepsilon_{zz}\right)^2 + \left(\varepsilon_{zz} - \varepsilon_{xx}\right)^2}$$

For a purely hydrostatic strain ($\varepsilon_{xx}=\varepsilon_{yy}=\varepsilon_{zz}$) we have thus B(ε)=0, as all linear dimensions are deformed in the same way. On the other hand pure axial deformation results in Tr(ε)=0, while B(ε)≠0, which corresponds to volume conserving strain. Let us reiterate that general deformations can be complicated combinations of all basic kinds of strain.

Stress and strain

In our discussions so far we have described one example of a strained system: a deformation due to the lattice mismatch of two semiconductor compounds. Strain however occurs in many daily life systems, in particular in building constructions or mechanical devices etc. We should also specify that we discuss elastic(reversible) deformations and we are not here interested in plastic (irreversible) deformations characteristic for systems such as chewing gum or clay.

We also need to distinguish strain from stress. The strain describes a deformation, by a comparison with an initial case, and ε is a dimensionless quantity. The stress (denoted as σ) is on the other hand as certain measure of the forces involved, and its unit is Pascal (N/m^2, force unit per surface unit). For example, our intuition has tells us that 1% strain of a $1mm^2$ cross-section steel wire demands application of a much large force the 1% deformation of the same cross-section rubber band.

The relation between the stress and strain is nothing else but a Hooke law:

$$\sigma = E\varepsilon ,$$

where the proportionality factor is a material parameter known as the Young's modulus also known as the tensile modulus. For example, steel has the Young's modulus value close to 210 GPa, while it is only about 0.1 GPa for rubber, what reveals very different elasticity of both materials.

Similarly to strain, stress in general case (for anisotropic materials) is a tensor (matrix like) quantity (σ_{ij}) and the relation between stress and strain is given by a generalized version of Hooke's law:

$$\sigma_{kl} = \sum_{ij} c_{klij}\varepsilon_{ij}$$

Because both stress and strain are second-order tensors, the linear relation between them is given by fourth-order elastic tensor, c_{ijkl}, which elements are known as bulk elastic constants. Elastic tensor has 3^4=81 matrix elements, however crystal symmetry significantly reduces the number of non-equivalent component. For example the cubic lattice has only 3 non-zero c_{ijkl} components, customarily denoted as C_{11}, C_{12} as C_{44}, where we used a notation in which single indices 1, 2 and 4 correspond to pairs of c_{ijkl} indices : 11, 22, 23.

Continuum elasticity and atomistic approach.

So far we have discussed lattice deformation of an entire crystal. Yet, in the beginning we have described a self-assembled quantum dot and claimed it is highly strained, while the substrate was almost strain unaffected. Thus, in general case, the deformation and so the strain tensor components, are position dependent. It is shown on Fig. 57, where the quantum dot volume is affected by a pronounced, compressive strain (Tr(ε)≈-0.08), while the substrate 10 lattice constants from the quantum dot remains practically unstrained.

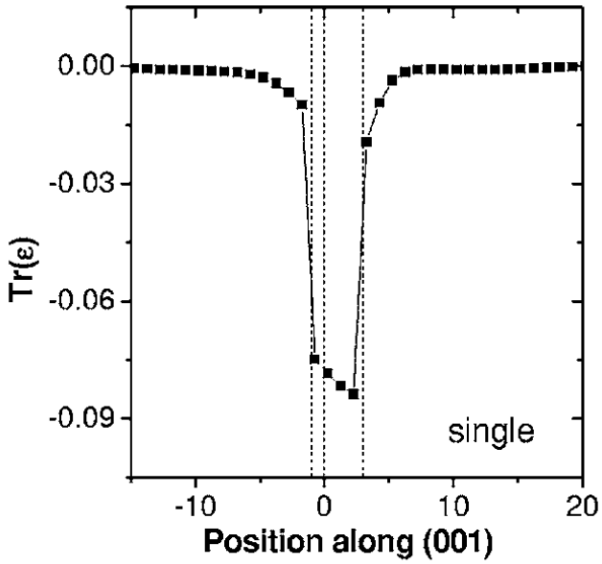


*Fig. 57. The trace of the strain tensor Tr(ε) calculated as a function of position; a cross section along the growth axis through a geometric dot center of a lens type InAs/GaAs quantum dot (from Phys. Rev. B **74**, 195339 (2006)).*

One of the methods allowing for calculation of spatial strain distribution is a continuum elasticity approach, where the atomistic character of crystal is neglected, and the medium is divided into small parts sharing the same properties as the undivided volume.

In our case we define the elastic (strain) energy in every point of discreet three dimensional grid. For a cubic symmetry systems we thus have:

$$E_{CE}(x,y,z)=\frac{V}{2}\left[\begin{array}{l} C_{11}(x,y,z)\left(\varepsilon_{xx}^2+\varepsilon_{yy}^2+\varepsilon_{zz}^2\right)+C_{44}(x,y,z)\left(\varepsilon_{yz}^2+\varepsilon_{zx}^2+\varepsilon_{xy}^2\right)+ \\ 2C_{12}(x,y,z)\left(\varepsilon_{xx}\varepsilon_{yy}+\varepsilon_{zz}\varepsilon_{xx}+\varepsilon_{xx}\varepsilon_{yy}\right) \end{array}\right],$$

where C_{11},C_{12} and C_{44} are position depended elastic constants (different for a quantum dot and a substrate material), V is a grid node volume. ε values at each grid point are calculated so the total strain energy reaches minimum, i.e. sum of E_{CE} over all grid points is minimized.

Under a close inspection, the is a fundamental issue related to this approach. First, the grid discretization (or volume V) must be chosen as small as possible to avoid numerical discretization errors. On the other hand even for a small grid step, comparable with the crystal lattice constant, the computational grid symmetry would be different from the lattice symmetry. In particular, for InAs or InP, it can be quite complicated zinc blende or wurtzite lattice symmetry. Further, we have seen

earlier X-STM quantum dot pictures, showing random fluctuation in atomic position. These fluctuations play a significant role for alloyed quantum dots of mixed chemical composition ($In_xGa_{1-x}As$, Fig 58) and cannot be simply accounted for by a continuum elasticity approach.

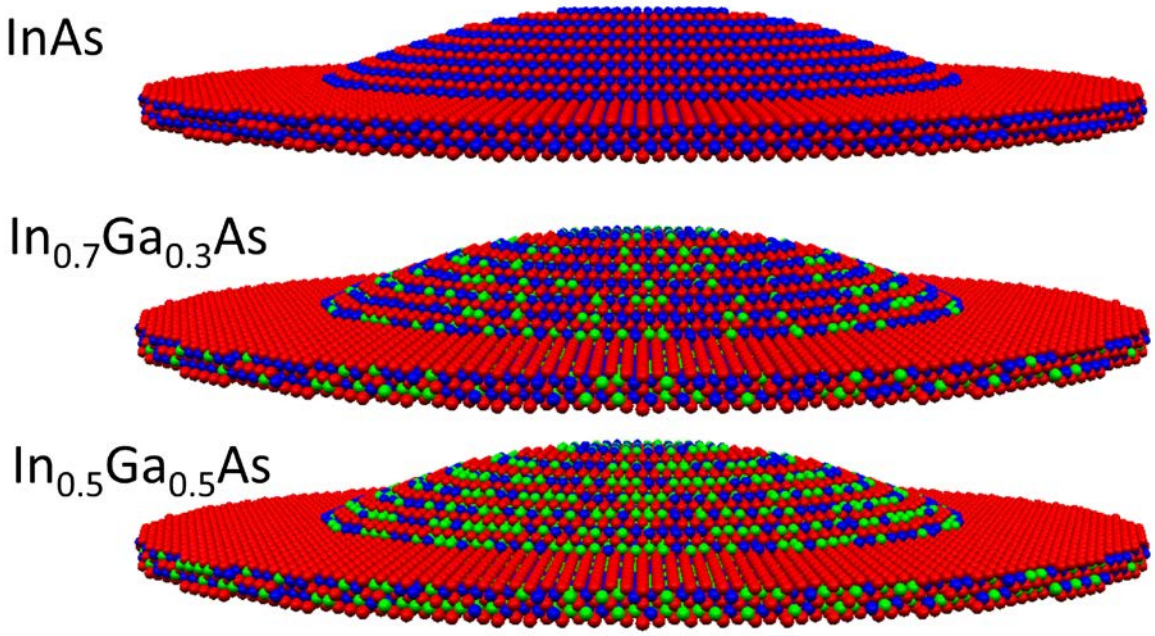


Fig. 58 .Positions of arsenic (red), indium (blue) and gallium (green) atoms for a wetting layer and lens type $In_xGa_{1-x}As$ quantum dot of different chemical composition. The surrounding GaAs matrix is not shown for picture the clarity.

Once way to account for the above issues is to include to the atomistic character of crystal lattice from the very beginning. By a comparison to continuum elasticity such methods are generally called "atomistic" approaches.

One example of an atomistic approach is the valence force field method (VFF), in which the total elastic energy is given in term of atomic positions, that is bond lengths and angles between bonds. For tetrahedral symmetry (such as in InAs/GaAs) we have:

$$E_{VFF} = \frac{1}{2}\sum_{i}^{N}\sum_{j=1}^{4} A_{ij}\left(\left(\mathbf{R}_i - \mathbf{R}_j\right)^2 - \left(d_{ij}^0\right)^2\right)^2 + \sum_{i}^{N}\sum_{j=1}^{3}\sum_{k=j+1}^{4} B_{ijk}\left(\left(\mathbf{R}_j - \mathbf{R}_i\right)\left(\mathbf{R}_k - \mathbf{R}_i\right) - \cos\theta d_{ij}^0 d_{ik}^0\right)^2,$$

There are two summations in the above formula. In the first summation, the *i* goes over all (N) atoms in the system, while *j* goes over (4) nearest neighbors, A_{ij} is a force constant related to *i* and *j* atoms bond length change (squeeze/stretch). The second summation goes over all atoms and atom pairs (*ij*), while B_{ijk} is a force constant related to the (*ij-ki*) bonds angle change when compared to the ideal tetrahedral crystal value ($\cos(\theta)=1/3$). In both summations $\mathbf{R}_i$ is atom *i* position.

Finding strain minimum in the VFF approach correspond to finding atomic positions ($\mathbf{R}_i$) such that E_{vff} reaches minimum. For a case of self-assembled quantum dots the number of atoms in the

computational domain can reach even up to 100 million atoms presenting a significant computation challenge.

Function minima – gradient descent method

There many numerical methods for finding function extrema, one example is the gradient descent approach. In these method the function minimum or maximum is found after a series of iterative steps toward a direction of largest function descent or ascent, depending on which of the extrema are of the interest (Fig. 59).

Fig. 59. The shortest path to the top is usually the steepest one.

The direction of the largest descent at a given point can be associated with the (negative) gradient at this point, more mathematically strict formulation would be the following:

$$x_{k+1} = x_k - \alpha_k \nabla f_k ,$$

where $f : R^N \mapsto R$ is continuous and differentiable (Fig. 60) function of N variables.

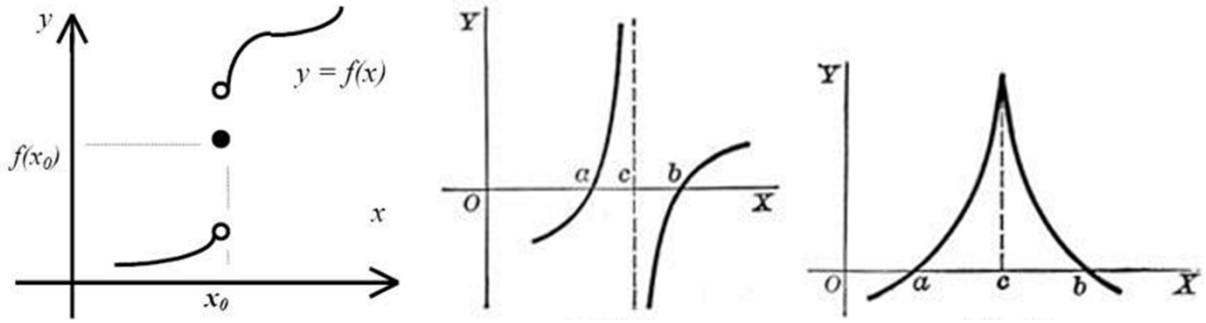


Fig. 60. Examples of discontinuous and nondiffrentable functions.

The iterative process (Fig. 61) starts at x_0, which choice is arbitrary and usually problem dependent. Step size (α_k) should be chosen so the function values are not increasing in the subsequent iteration steps ($f(x_0) > \ldots > f(x_k) > f(x_{k+1})$). Too large step will make the method unstable, however too small step will slow down the calculation. Similarly to the case of x_0,the choice of α_k is to a large degree arbitrary and problem dependent.

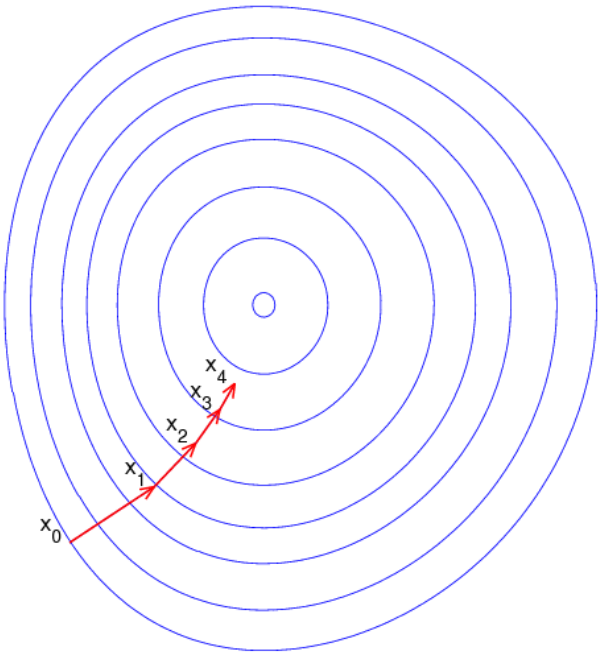


Fig. 61. Several steps of the gradient descent method.

One of the ways to avoid the above problem is to use the steepest descent method, where the step size is not the parameter, but is calculated at every iteration step, so the function reaches minimum along the gradient direction:

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{\alpha>0} f(x_k - \alpha \nabla f(x_k))$$

The advantage of the above approach is that one avoids the ambiguity of the step size α_k choice, on the hand the steepest descent method is both formally and numerically more complex than the gradient descent approach.

Irrespectively from the α_k choice the iterative process is continued until the sufficiently good approximation to the minimum is found. To achieve this goal several convergence criteria can be used, e.g. the gradient length should be smaller than a certain error value ($|| \nabla f(x_k) || < \varepsilon$) or the distance between subsequent iteration points should be smaller than a certain distance ($|| x_{k+1} - x_k || < \delta$). Both criteria can be combined and used together.

Finally, we should point that in both gradient or steepest descent, local and not global function minimum is found (Fig.5. 20). For a function having multiple local minima the output of the gradient algorithm depends on many factors, most importantly the choice of the iteration starting point x_0.

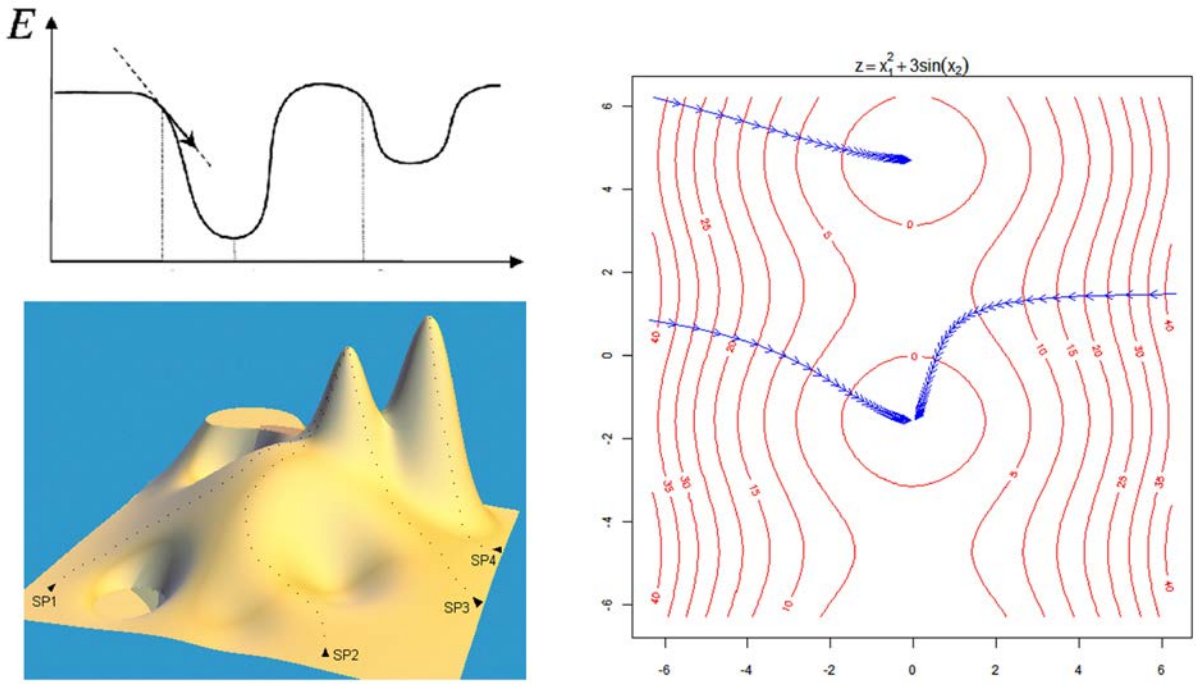


Fig. 62. One- and two-dimensional functions with multiple local minima.

Modeling of Mechanical and Electrical Properties of Bionanosystems

Introduction to nanotechnology of biological systems. Central Dogma of Molecular Biology.

Let us consider the phenomenon of life. Life lasts continuously some 2 billion years. There are millions of various living species. Biologist classify them into kingdoms, orders, families, etc. Constant evolution modifies these organisms. Changes may be positive (better fitting to environment) or negative. Quite often some species (dinosaurs) extinct, but others (cyanobacteria) last for millions years. The very basic fact of life is reproduction – some individuals don't survive but others have a "progeny". Sometimes a period of hibernation is involved before an organism shows all its capabilities (seeds). Sometimes survival of the progeny critically depends on the performance of parents (mammals). How does information ("life plan") goes from one organism ("parents") to another ("kids"). What chemical mechanisms allow for better fitness and adaptation to the local environment? Why some 60 mln years ago large dinosaurs have became extinct (an impact of large meteorite??), but small mammals have survived? Due to progress in science (biology, chemistry, physics, computer science) we may try to answer these fundamental questions.

Knowing well biological mechanisms and structures we can mimic certain solutions "invented' by the Nature through the millions of years of the evolution. We may try to transfer these solutions into technology – even on nanoscale. This was the motivation for the newly developed of science: bionanotechnology [Ehud Gazit, Plenty of room for biology at the bottom: An introduction to bionanotechnology. Imperial College Press, 2007, a free chapter: http://www.worldscientific.com/doi/suppl/10.1142/p465/suppl_file/p465_chap01.pdf].

People put a lot of effort in "improving the Nature". And here it comes: computer modeling of bionanosystems.

Living organisms are (usually) composed of tissues, and tissues are made of cells. Main metabolic processes, such as intake of energy-rich compounds or structural compounds, building new cell organelles, generation of energy, etc., are localized in the cell. Typical cells are small, the diameter is of the order of one micrometer. They are made of biopolymers, and the interior contains a dens solution of membranes, proteins, signaling or energetic compounds. Structural or enzymatic proteins (and perhaps phospolipides present in membranes) play a major role in cell metabolism. Proteins are synthesized in specially designed regions of cells – in ribosomes. Ribosomes are complex "nanofactories". In each human being there are estimated 1,000,000,000,000,000,000 ribosomes. The 3D structure of the ribosome – extremely complicated macromolecular complex – has been determined at the end of XXth century and for this discovery a Nobel Prize has been awarded (see photo below).

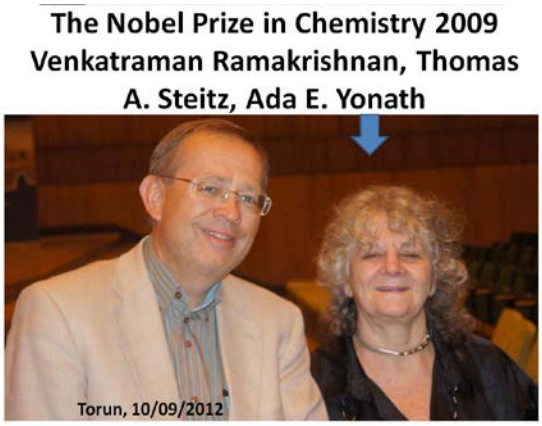


Fig. 63. Prof. Ada Jonath from Israel during her stay at UMK.

Please enjoy viewing the structure of a ribosome here:
http://www.rcsb.org/pdb/101/motm.do?momID=121
or here:
http://www.dnatube.com/video/2425/Translation-Atomic-View).

How this biochemical machine works? How it comes that from such a simple components – amino acids – so complex proteins are made (hemoglobin – governs oxygen transport, actin - muscles,…)? One may answer these questions using results of computer simulations of proteins dynamics. The largest supercomputers are used for the modeling , since it is very CPU intensive job. Movies presenting the action of the ribosome are shown here:
http://www.dnatube.com/video/1228/tRNARibosome-Molecular-Dynamics-Simulation,

Proteins are bioplymers built from amino acids. Amino acids are linked via peptide bond:

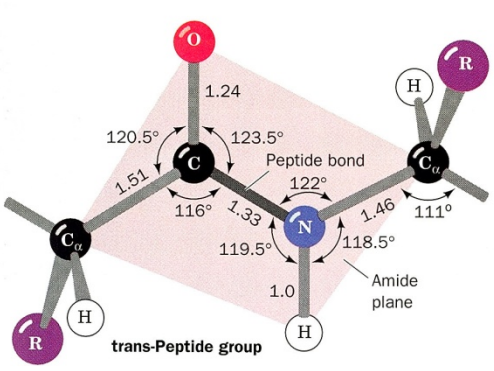


Fig. 64 The peptide Bond (source: D. Voet, J. Voet, Biochemistry, 2n Ed, 1995)

There are 20 basic amino acids (AA) present in natural proteins:

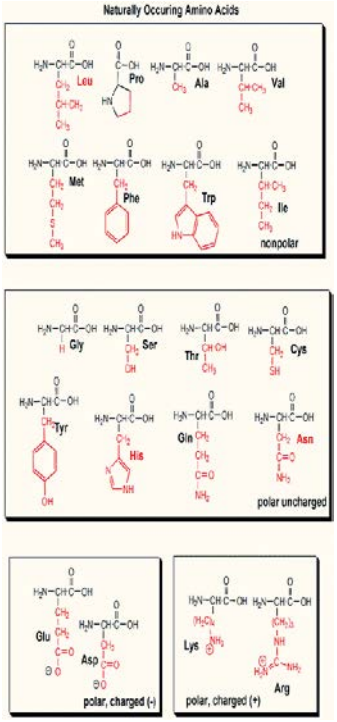


Fig. 65. Structures of basic amino acids (Internet, **http://employees.csbsju.edu/hjakubowski/classes/ch112/proteins/aminoacidprot1.gif** *).*

Practically all proteins are made of these 20 (or 22) building block. Of course, chemists know much more various amino acids (thousands), but only such a small fractions is used in construction of the living matter.

The order of connection of AA into a protein chain is critically important. The particular arrangement affects physical, chemical and eventually biological properties of a protein. This is possible, since the amino acids are different: they have different polarities, charges, hydrophobicity. Nanobiotechnologists may use such diverse building blocks to create new systems. The sequence (a primary structure) is important, but the fold of a protein finally determine its properties. We know nearly 1000 different folds – specific arrangements of a protein's polymer into a 3D functional structure. Should a protein fold spontaneously into a wrong shape during the synthesis – it will be tagged by special enzymes and recycled in the proteosome. Please note, that not all proteins necessary for life have defined 3D fold. There are also intrinsically disordered proteins. In recent years this became a field of vigorous research.

Proteins may be, *inter alia*, enzymes and/or structural proteins. Wool, silk or spider fibers are made (mainly) from proteins.

What is a flow of information into a ribosome showing what protein has to be made? In what order AA should be assembled? The factory receives a precise recipe from DNA via RNA. Cells contain condensed DNA in their nuclei. This fantastic molecule is also classified as a biopolymer. It is made from nucleotides: there are only four types of nucleotides (letters) used.

The nucleotide is composed of DNA (or RNA) base) a five ring sugar (deoxyribose or ribose) and a phosphate core. The chemical bases involved are called purines (two rings) and pirimidynes (one ring). The 3D structure of DNA has been discovered in 50. by J.Watson and F.Crick (a Nobel Prize).

That time they built rather metal then computer models of DNA. A typical D-DNA has alpha-helical structure (twisted stairs):

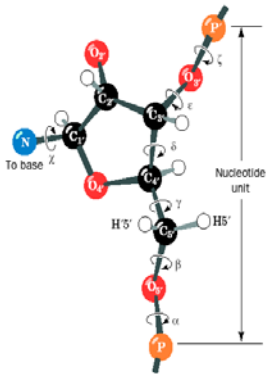


FIGURE 29-7 The conformation of a nucleotide unit is determined by the seven indicated torsion angles.

Fig. 66. A scheme of nucleotide (from: D.Voet, J. Voet, Biochemistry, 2n Ed, 1995).

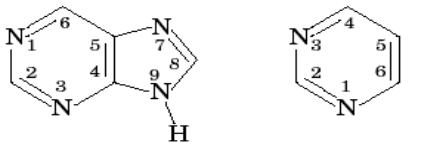


Fig. 67. A scheme of puryne and pirymidyne rings. These structures code information in DNA and RNA.

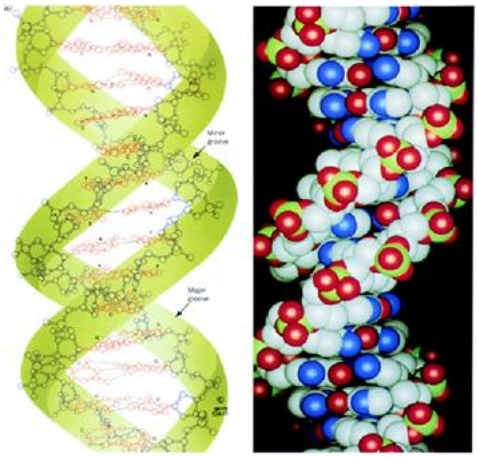


Fig. 68. A structure of B_DNA (D.Voet, J. Voet, Biochemistry, 2n Ed, 1995).

There is only one copy of nuclear DNA in the cell. It is a very long molecule – 1.5 meter, 3, 000, 000, 000 pairs of bases are there. It is spitted into 23 chromosomes. The diameter of DNA jest very small, mere 2-3 nm. There are two complementary chains of DNA polymers. The standard pairing of DNA bases (via 2 or 3 hydrogen bonds) is A-T, C-G.

Certain fragments of DNA contain information how each protein should be synthesize. These fragments are not contiguous, data for one protein (a gene) are usually scattered over a large span of DNA chain. The coding regions are called exons, non-coding ones: introns.

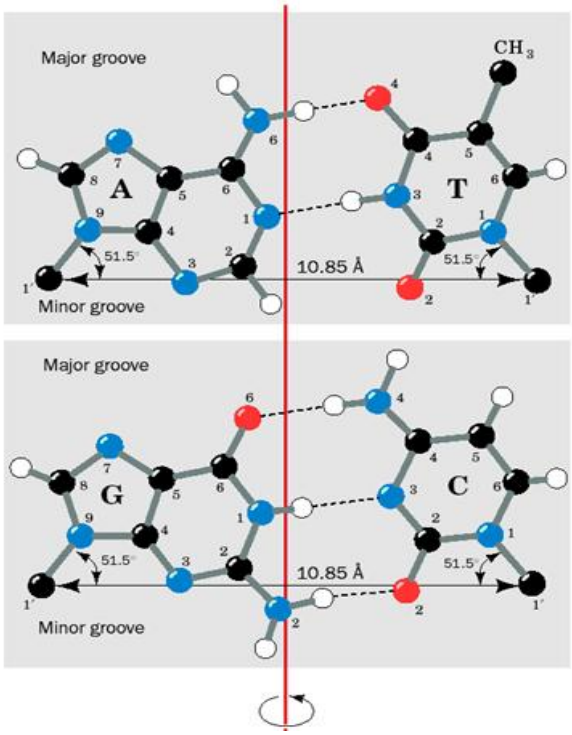


Fig. 69. Standard Watson_Crick (WC) base pairs (bp) (source: D.Voet, J. Voet, Biochemistry, 2n Ed, 1995).

The median length of an exon is 12 bp, on average a gene contains 8 exons (up to 350), the average length of a coding sequence is 447 AA. Thus in 3 bln letters of human DNA there are more non-coding regions than coding ones. In 2000 the human genome has been red out, it came out theta the number of human genes is smaller than initially expected: 20 687 (as of year 2012).

When a new protein has to be synthesized, for example for formation of the progenitor cell, a fragment of the DNA polymer winds up. This process of getting access to a tightly packed polymer is catalyzed by appropriate enzymes. Then on a single, locally „isolated" DNA strand (for example, 3000 bp) a synthesis of a new, complementary RNA chain occurs. His short RNA chain is called messenger RNA (mRNA). RNA is chemically quite similar to DNA, instead of T we have uracil (U) here, and deoxyribose is replaced by a similar sugar (ribose as well) – ribose. The mRNA leaves nucleus and goes to ribosome. There information encoded in the mRNA sequence (the same or modified by a process of splicing) is red (3 letters at once) like a magnetic tape. It has been found that the universal genetic code is used in all living organisms (humans, animals, plants, ..). A given sequence of 3 RNA letters denotes an amino acid. The code is shown in the table. Note, that the code is redundant (there are 64 combinations of 3 element string composed of 4letters, but only 20 AA), this helps to

protect critical protein regions against spontaneous (or chemically induced – cigarettes smoke!) mutations which destroy precious genetic information.

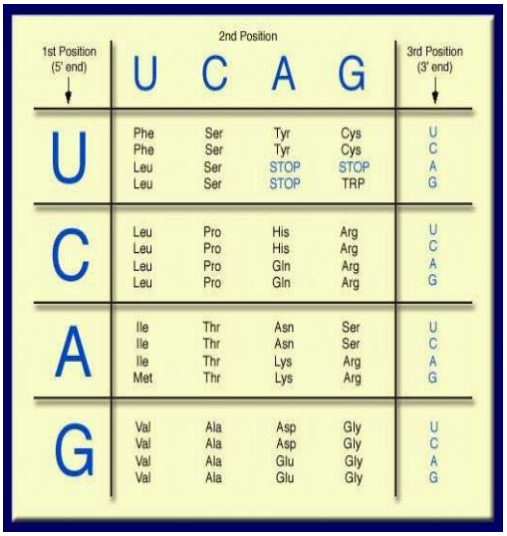


Fig. 70. The genetic code (Internet)

Redundancy helps to protect DNA. The ribosome check 3 letters (i.e. AUU) and allows for creation of the new peptide bond if and only if t-RNA brings into the synthesis place the corresponding amino acids coded by this sequence (Ile). All other possibilities (wrong amino acids statistically delivered to the site) are rejected. This process is very fast, it is controlled by complex enzymes.

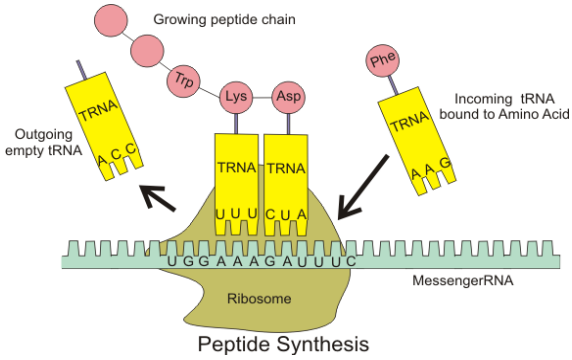
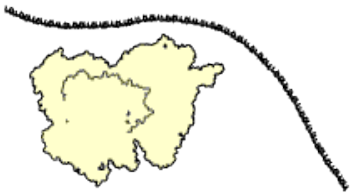


Fig. 71. Protein synthesis in the ribosome "nanomashine" (Internet) .

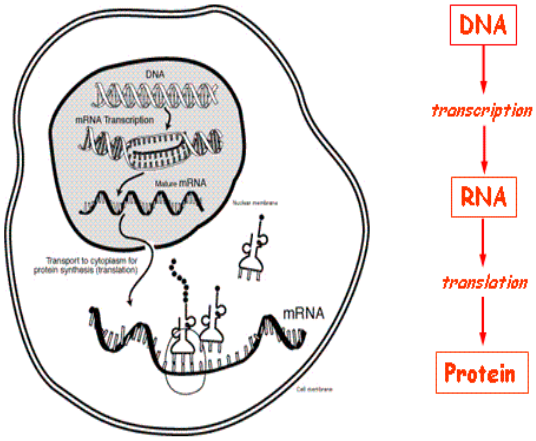
It is recommended to watch the following computer animation of biotechnological process and a nano-machine in action (it is not a MD trajectory!)



http://en.wikipedia.org/wiki/Translation_%28biology%29

Fig. 72. An animation showing synthesis of a protein. It is not MD trajectory but just computer graphics animation!

Imporant: The transfer of genetic information from DNA via RNA to protein is called a Central Dogma of Molecular Biology



It is worth to know that after synthesis a protein chain may be further post-translationally modified, for example, serines may get phosphorylated.

The Central Dogma of Molecular Biology was the topic of our first lecture. Now we understand that changes in genes (DNA sequence) may lead to production of altered protein products, which in turn can result in disturbance of cell metabolism, often manifesting as a disease. Although lethal changes result in death of the cell concerned, there are also mutations that render the cell immortal, ie. able to multiply indefinitely. This mechanism is involved in neoplasm formation (oncogenesis). Thus, there are questions that remain unanswered - how can we prevent mutagenesis, what means should we

use to stop uncontrollable growth of cells? How can we protect organisms from „falsification" of genetic information by environmental factors, such as ultraviolet radiation (eg. sunbathing, which a risk factor for melanoma) or tobacco smoke (lung cancer)?

Health is a sought-after good, which explains why rich societies spend considerable sums on research into molecular basis of diseases and new treatments. Because biological molecules (such as proteins and DNA) are so small that nanometers are used to describe their size, the new discipline of practical research concerning their properties is called bionanotechnology. It employs not only physical methods, such as X-ray diffraction, NMR or AFM (atomic force microscope), but also computer simulations.

It is worth noticing that modern methods of molecular biology enable us to make a range of organisms (bacteria, yeast) produce virtually any protein. We are also able to use enzymes to edit the DNA sequences as we like. Automated machines let us describe the composition of biopolymers and bioinformatics analysis helps us in optimizing their physicochemical properties. Within bionanotechnological framework scientists try to merge biopolymers with other materials or objects, such as quantum dots, in hope that this will lead to new advances, eg. materials efficiently transforming light into electrical current. In the following lectures methods of computing analysis of structural, electric and mechanical properties of such nanomaterials will be presented.

WN-W2 (2h)

Methods of modeling – a formal basis.

Introduction

The whole idea of computer simulations is rather simple: We construct a mathematical model of a biopolymer (protein, DNA, phospholipids, a hormone, a drug, etc.). The model has to be based on some empirical data: dimensions of the molecule, masses of atoms, charges, presence of the environment having some counter-ions, temperature, pressure and a definite pH. One of the main goals of modeling is a search for optimal structures of these biopolymers. Quite often we are interested in dynamics: how fast and how often these biomolecule conformation change one into another. It is possible to monitor how a protein interacts with the other one (proteomics). Models are useful in testing effects of a substitution of one amino acid for another in an enzyme: what happens to the electric field in the catalytic clef – is it stronger, weaker or the same after mutation?
One can dock to such a model small chemicals (drugs) and can check what is the interaction energy. It is believed that strong binders are good candidates for inhibitors (new drugs). One may even tailor molecules having desired biochemical properties, thus having predefined physiological effects (rational drug design). A number o research group tests putative mechanisms of enzymatic reactions, for example, those involved in biotechnological production of beer or widely used chemical acrylamide.

Computer modeling of biopolymers is recognized research method. It allows us to obtain relatively quickly new, unique information in a rather inexpensive way. Computer modeling has already been used in drug design processes, several compounds (such as anti-virial drugs) are marketed. It has also been useful in design of innovative diagnostics tests and in green chemistry (biotechnology). Using computer modeling we gather new information on molecular biology of cell, breeding processes, evolution, metabolism, photosynthesis, some pathological process. There is a continuous quest for new materials, for example hybrid materials. To this end biomimetic approach exploits “experience” gathered through millions of years of evolution.

The modeling requires solid physical background, a lot of computer science skills (informatics) and a good deal of math. Commercial software suites may cost hundred thousand dollars. Software houses may employ up to 200 persons with PhD degrees to write such specialized software. Computer specialists are also hired in this high-tech field.

Currently some companies dominate computer modeling commercial market, but there is a lot of academic software, or freeware, which may do a very good job for us. We will present fundamentals of some computational methods and will learn (during tutorials) how to use the **ArgusLab** code (semi-empirical quantum chemistry) and the NAMD+VMD (classical molecular dynamics) suite.

A map of molecular modeling methods - quantum vs classical approach

In computer modeling of chemical systems (molecules or their complexes) and/or biopolymers we are interested mainly in two questions:

(a) What is their molecular structure?
(b) What interactions with the environment are involved?

We answer these questions using calculations, not measurements (these are difficult to perform and expensive). Quantum mechanics is the physical theory which should be applied here. However, biomolecules are too large to be studied using a full quantum mechanical approach. They simply contain too many electrons. Thus, we adopt more approximate and simplified description and models. An atom is the smallest piece of matter analyzed separately in the classical approach. Sometimes (coarse grained models) it is even a group of atoms ("a bead"). In the technical and scientific literature there are numerous acronyms referring to the level of approximation used in the studies. In molecular quantum mechanics (quantum chemistry) a crude map of the methods is as follows:

Methods of molecular quantum mechanics:

- „*ab initio*" – based on the variational principle [the Hatree-Fock method (HF), the configuration interaction method (CI),] or the perturbation theory [Many Body Perturbation theory, (MBPT), the Coupled Cluster method (CC, CC SD), MP2]
- DFT – Density Functional Theory (DFT B3LYP, LSDA, TD DFT , etc.)
- Semi-empirical methods (the Huckel method, the PPP method, the Tight Binding method, the Austin Model 1 (AM1) method, Parametric Method 3 (PM3, PM6), ZINDO/S method).

These methods have different levels of approximations adopted, however, each has its own advantages and disadvantages. The positive aspects of semi-empirical methods that we plan to use during tutorials (ArgusLab code and AM1, PM3, MNDO) is their efficiency, speed, good quality of optimized geometries, applicability to very large systems, up to hundreds of atoms.

The following packages for quantum chemical calculations are perhaps the most popular nowadays: Gaussian09, Gamess, NWCHEM, VASP, Jaguar, Hyperchem. The license fees (one year) for some of them are higher than a thousand of US dollars.

A 3D structure of a biomolecule, important for assessment of its biological or technological role, corresponds to the minimum of potential energy surface (PES). We will introduce this concept formally, since this a basis of the whole bionanotechnological modeling.

Potential Energy Surface (PES)

Quantum mechanics predicts that the whole information about a molecule is present in its wavefunction in Dirac's notation $|\Phi\rangle$. The wavefunction one can obtain solving the Schroedinger equation:

$$\hat{H}|\Phi\rangle = \varepsilon|\Phi\rangle$$

The operator of energy, Hamiltonian $\widehat{H}$, depends on both electronic (*i,j*) and nuclear degrees of freedom (A,B), in atomic units we have:

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{i>j}^{M}\frac{1}{r_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}$$

Atomic units (a.u.):

$\hbar = 1$

"$c = 1$" (137.)

$m = 1$

Usually so called Born-Oppenheimer approximation is adopted. W assume that the atomic nuclei are infinitively heavy, thus they don't move. In that way a separation of electronic and nuclear degrees of freedom is possible:

$$\hat{H}_{el}\Phi_{el} = \varepsilon_{el}\Phi_{el}$$

Here the electronic Hamiltonian reads:

$$\hat{H}_{el} = -\sum_{i=1}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{i>j}^{M}\frac{1}{r_{ij}}$$

and the electron wavefuction Φ_{el} depends on positions of atoms $\{\vec{R}_A\}$ in a parametric way (semicolon denotes such a dependance):

$$\Phi_{el} = \Phi_{el}\left(\{\vec{r}_i\};\{\vec{R}_A\}\right)$$
$$\varepsilon_{el} = \varepsilon_{el}\left(\{\vec{R}_A\}\right)$$

We see that the electronic energy of the biomolecule studied ε_{el} depends on positions of atoms. For each fixed individual position of the nuclei we have distinct wavefunction Φ_{el} (it depends on electronic degrees of freedom $\vec{r}_i$). Once the interatomic distances R_{AB} are set, the total enerrgy of the biomolecule may be obtained:

$$\varepsilon_{tot}^{fix\,N} = \varepsilon_{el} + \sum_{A}^{N}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}$$

Nuclear Hamiltonian $\hat{H}_{nucl}$, on the other hand, reads:

$$\hat{H}_{nucl} = -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \left\langle \hat{H}_{nucl}\right\rangle_{\Phi_{el}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} =$$
$$-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \underbrace{\varepsilon_{el}\{R_A\} + \sum_{A}^{N}\sum_{B}^{M}\frac{Z_A Z_B}{R_{AB}}} =$$
$$-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 \quad + \quad \varepsilon_{tot}^{fix\,N}\{R_A\}.$$

In this expression we see a term $\varepsilon_{tot}^{fix\ N}$ that is a potential in which atomic nuclei move. This term is called potential energy surface (PES). In classical molecular dynamics the quantum character of PES is negleced and this surface is approximated by analytical potentials (force fields). In numerous laboratories force fields have been developed. They are used to model different nanosystems, for example, proteins or crystals. Of course each type of nanosystem requires a dedicated, special force field.

Let us notice that solving (using QM) the nuclear Schroedinger equation:

$$\underbrace{\hat{H}_{nucl}\Phi_{nucl} = \varepsilon_{nucl}\Phi_{nucl}}$$

one can obtain information on molecular oscillations, rotations, and translations. We can obtain the same data, not refering to the QM, but just to mere calssical mechanics, provided that a realistic formula for PES is known. Such surfaces (force fields) are subject of both commercial and academic developement based on experiments and quantum chemical calculations.

Force Fields (FF)

Applications of force fields in nanosystems modeling are based on an underlying assumption that these parametrized analytical formulas are good approximations to real PES. There is yet another important assumption, that the parameters of FF describing a group of atoms (amino acid) in one biomolecule may be transferred without modifications to another one**.** The general formula for FF is the following (note that there are even more complex formulas):

$$V = \sum_{bonds\ i} V_i^B + \sum_{bond\ angles\ j} V_j^A + \sum_{torsional\ angles\ k} V_k^D + \sum_{improper\ angles\ l} V_l^E$$

$$+ \sum_{atom\ pairs\ (r,s)} (V_{r,s}^C + V_{r,s}^P + V_{r,s}^{VdW})$$

The physical meaning of individual energy terms are described in the figure caption.

The main terms correspond to energies of chemical bonds (V_i^B), energies of bond angles deformations, (V_j^A, V_k^D, V_l^E) and two-body interactions between all pairs of atoms present in the system. The nonbonding interactions contain Coulomb energy terms (electrostatics, partial atomic charges) (V^C) and Van der Waals (V^P+V^{VdW}) interactions.

Let us repeat, that in computer modeling not only “all atom” force fields are used, but crude coarse-grained models (a group of atoms is a centre of potential) show growing popularity.

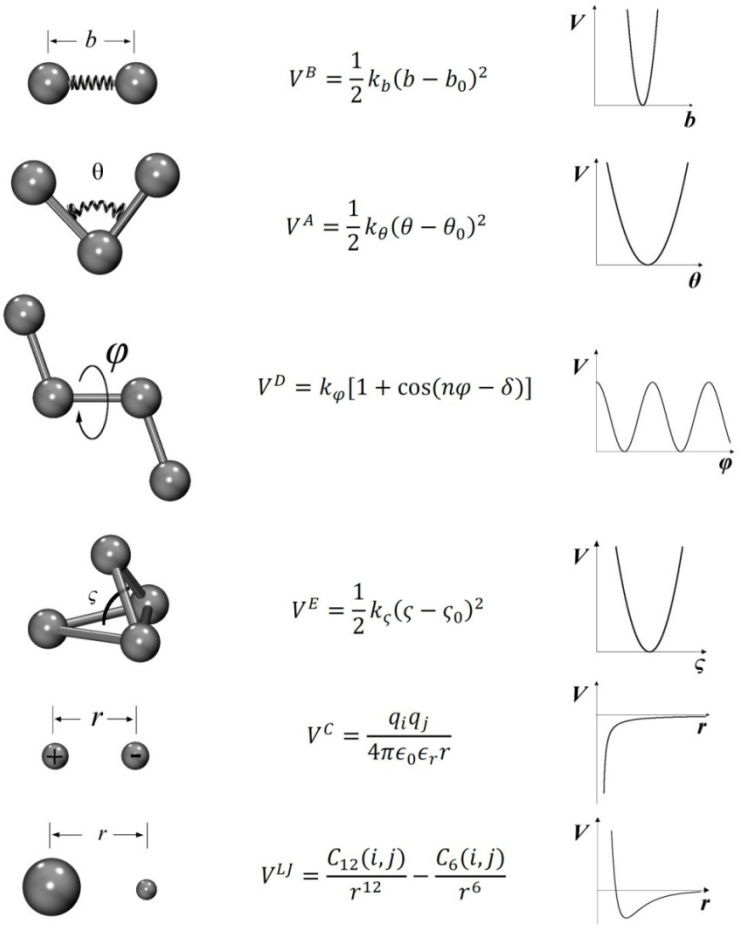


Fig. 73. (based on H.Grubmueller, "Proteins as Molecular Machines: Force Probe Simulations", published in Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes, Norbert Attig, Kurt Binder, Helmut Grubmueller, Kurt Kremer (Eds.), John von Neumann Institute for Computing, Julich, Germany, NIC Series, Vol. 23, ISBN 3-00-012641-4, pp. 401-422, 2004. © 2004 by John von Neumann Institute for Computing), Interaction contributions to a typical force field. Bond stretch vibrations are described by a harmonic potential VB, the minimum of which is at the equilibrium distance b0 between the two atoms connected by chemical bond i (the indices i, j etc. are not shown in the Figure). Bond angles and out-of-plane (improper) angles are also described by harmonic potential terms, VA and VE, where Θ0 and ζ0 denote the respective equilibrium angles. Dihedral twists (torsional angles) are subjected to a periodic potential VD; the respective force constants are denoted by k's with appropriate indices. Non-bonded forces are described by Coulomb interactions, VC, and Lennard-Jones potentials, VLJ = VP + VVdW, where the latter includes the Pauli repulsion, VP ~ r^{-12} and the van der Waals interaction, VVdW ~ $-\mathrm{r}^{-6}$, respectively.

In protein and DNA modeling the most popular force fields are: CHARMM, AMBER, GROMOS, OPLS. References for their full formal description may be found in Wikipedia. We will use CHARMM v.27 in our tutorial exercises.

Geometry optimization (=energy minimalization)

According to Anfinsen hypothesis native proteins adopt conformations corresponding to minimum of their free energy F. The free energy function, defined as F=E-TS, takes into account a classical potential energy E (taken from the FF) and entropic factors (S) and a temperature (T) of a system. Quite often we neglect entropic factors. Then finding the minimum of F is equivalent to finding the corresponding minimum of E. Thus, we use computers to answer the fundamental question: for what atomic positions the energy E of a biomolecule is the lowest. Since E depends on positions of so many atoms, and we need take into account water as biological environment, the task is not trivial. We need to minimize a function of many, many variables: a rough estimate for hemoglobin (20 000 atoms) gives 60 000 degrees of freedom. All this has to be optimized! Fortunately it is not a hopeless problem: despite lack of a general algorithm for finding the global minimum of such multidimensional function, there are methods for finding some local minima.

„Brute force" minimization

When the number of degrees of freedom is not large, the systematic scanning of the whole conformational space may lead to good results. For biomolecules it is a dead end, since the number of conformers exponentially grows with the number of atoms.

Steepest descent methods

We will present the steepest descent method (SD) for one dimensional function f(x).

1. Select the starting point x_0
2. For point $x_k = x_0$ calculate the negative gradient:

$$-\nabla f(x_k)$$

3. The next point is determined from the formula:

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

4. The step size (in a given direction) is determined by linear minimization in this direction, we select this factor γ_k, that the following condition is fulfilled:

$$f\big(x_k - \gamma_k \nabla f(x_k)\big) = \min_{\gamma>0} f\big(x_k - \gamma_k \nabla f(x_k)\big)$$

5. Iterate points 2-4 until stop criterion is met:

$$\nabla f(x_k) \leq \varepsilon \text{ and } f(x_k) - f(x_{k-1}) \leq \varepsilon'$$

The scheme may be generalized into multidimensional functions.

The Conjugated Gradients (CG) method offers better efficiency than SD. Here the direction of the minimization is determined not form the pure gradient, but form a "mixture" of the current gradient and the previous direction of the minimization.

1. Find the value of the steepest descent (sd)

2. If sd < ε, end the search
3. If it is a step no. 1 – remember the direction of the descent (for step no 1 it will be a direction of the gradient)
4. For a step no. K, K>1, determine the direction of the descent. It will be a combination of the current gradient and the previous step direction of the descent (conjugate directions). Find a minimum in that direction. Go to this minimum.
5. Go to point no. 1.

There are numerous popular variants of the conjugate gradient methods: Fletcher-Reeves, Polak-Ribiere, Hestenes-Steifel (see:

http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher_reeves.html

The CG algorithm requires moderate memory size, is relatively fast converging, but does not work well for structures far from the minimum. In such a case the SD algorithm is better.

The main drawback of these algorithms (SD CG and similar) applied to molecular mechanic problems is the need for calculations of energy function gradients. In typical force Fields such a function has to be numerically differentiated. This is a computationally intensive task. Back in 1970s another minimization algorithm has been introduced, similar to CG, but in this method there is no need for grad E calculations. It is called Powell algorithm, since it is widely used in MD simulations, it is recommended to study details of this approach:

(See Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007). "Section 10.7. Direction Set (Powell's) Methods in Multidimensions". *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). New York: Cambridge University Press. or http://people.equars.com/marco/poli/phd/node55.html)

2nd order methods (Newton-Raphson, N-R)

Much faster convergence may be achieved (in 1 step for quadratic surfaces!) when the information of second derivatives of the optimized function (PES) may be included.

Les us put:

$\mathbf{x}_k$ – a point in configuration space in step k

$\mathbf{d}_k$ – direction of minimization in step k,

-gradient:

$$-\nabla f(\boldsymbol{x}_k) = -\boldsymbol{g}_k$$

ω, λ – scalar coefficients

H –matrix of second derivatives (Hessian matrix)

V – an approximation (matrix) to the Hessian reciprocal matrix

According to the N-R method we have:

$$x_{min} = x_0 - \boldsymbol{H}_0^{-1} \nabla f(\boldsymbol{x}_0)$$

***

An example (1 –dim)

$f(x) = a + bx + cx^2$

$f'(x) = b + 2cx$

$f''(x) = 2c$

In minimum we have $f'(x_{min}) = 0$ thus $b+2cx_{min}= 0$ and $x_{min} = -b/2c$

Therefore:

$x_{min} = (-f'(x)+2cx)/2c= x - f'(x)/f''(x)$.

***

Calculations of second derivatives for biopolymers are highly impractical, since very large RAM spaces are required. There are, however, numerous approximate N-R methods: quasi-Newton (Davidon-Fletcher-Powell or Broyden-Fletcher-Goldfarb-Fano). These methods are used in the ArgusLab code (semi-empirical quantum chemistry) to optimize geometry of medium-sized molecules.

Integration of equations of motion (Newton, Langevin)

In the classical molecular dynamics we assume that time evolution of the molecular biosystem is given by Newton equations of motion:

$$\boldsymbol{a}_i = \boldsymbol{F}_i \, / m_i$$

In order to determine the position x_i of an atom *i* in time point *t* we have to know its mass m_i , the force acting on the atom $\mathbf{F}_i$ (t) and integrate the equation twice with respect to time. We need to repeat this procedure for each atom one by one, sometimes 20 000 or 200 000 times (N). If we know the force field (and we know it , indeed) we may calculate the gradient of E, in a point of the configurational space that we study in the current moment. Thus we can obtain required forces $\mathbf{F}_i$ acting on atoms:

$$\boldsymbol{F}_i = \, -\nabla E(\boldsymbol{x})$$

We need to integrate a differential equation for the acceleration $\mathbf{a}_i$ twice, in order to find a trajectory for a given atom x(t). In the numerical analysis methods we search for the most effective way of performing this integration (using a computer, of course). Among many methods the most popular one is perhaps an algorithm developed by Verlet. We present a few variants of this algorithm.

In this approach time is not continuous anymore: it is divided (discretized) into equidistant points. We solve the problem on a mesh. The time step depends on mechanical problem studied. Too small time step is expensive and redundant, to large leads to a crash of simulations. Usually we require that the time step is not smaller than the fraction 1/10th of the period of the fastest periodic motion in the system under study. In practice, for proteins at 300K, the fast test are C-H oscillations, and we use 1 fs time step (sometimes 2 fs will do as well). During practical exercises it is worth checking how

big a time step may be before your system crashes. If one sees that the energy of the system suddenly rises up, and MD simulation stops before reaching assigned number of steps, then it is recommended to repeat this job with the time step reduced by a factor of 2.

We want to simulate motions of biomolecular atoms. Atomic motions are related to equilibrium temperature of the system and surroundings. It is safe to use the statistical concept of the temperature, since our system contains usually thousands of atoms. We initiate MD calculation by selecting of initial velocities of atoms. We require conservation of the momentum:

$$\mathbf{P} = \sum_{i=1}^{N} m_i \mathbf{v}_i = 0$$

Directions of the velocity vectors are random. Velocity components are generated from the Maxwell-Boltzmann distribution:

$$p(v_{xi}) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left[-\frac{m_i v_{xi}^2}{2 k_B T} \right]$$

$$T = \frac{1}{3N} \sum_{i=1}^{N} \frac{|p_i|}{2m_i}$$

The „positional" Verlet algorithm uses information on atomic positions in time points t and $t - \delta t$ in order to determine new positions in time point $t + \delta t$. One can easily generate this formula using the following Taylor expansions of r(t):

$$\left.\begin{aligned} r(t+\delta t) &= r(t) + v(t)\delta t + \frac{1}{2} a(t) \delta t^2 \\ r(t-\delta t) &= r(t) - v(t)\delta t + \frac{1}{2} a(t) \delta t^2 \end{aligned}\right| -$$

Limiting the expansions to the second order terms, and subtracting both series one has the recipe for trajectory propagation:

$$r(t+\delta t) = r(t) - r(t-\delta t) + \frac{1}{2} a(t) \delta t^2$$

The „Leap-frog" algorithm

This is another popular variant of Verlet algorithm. Like in a childrens game „leap-frog" in this algorithm we calculate velocities for time points $t + \delta t/2$, then positions are determined for $t + \delta t$.

$$\begin{cases} v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t \\ r(t+\delta t) = r(t) - v\left(t + \frac{1}{2}\delta t\right)\delta t \end{cases}$$

The advantage of LP variant of Verlet is the possibility of storing velocities (we don't work in the configurational space but in the phase space). Note, however, that the positions and velocities are shifted in time, often we use average for velocities:

$$v(t)=\frac{1}{2}\left[v\left(t-\frac{1}{2}\delta t\right)+v\left(t+\frac{1}{2}\delta t\right)\right]$$

Sometimes „ velocity" variant of Verlet algorithm is used:

$$\begin{cases} r(t+\delta t)=r(t)+v(t)\delta t+\frac{1}{2}a(t)\delta t^2 \\ v(t+\delta t)=v(t)+\frac{1}{2}[a(t)+a(t+\delta t)]\delta t \end{cases}$$

WN-W3 (2h)

Practical aspects of molecular dynamics simulations. A study on a mechanical strength of nanocomposite biomaterial.

Temperature in SMD simulations

While integrating equations of motions we need to conserve the assumed statistical properties of the modeled system. We may want to keep constant the total energy (canonical ensemble), or temperature (microcanonical ensemble, a contact with thermal bath is assumed). During our simulations we will keep temperature constant (Langevin method). The simple method of velocity scaling also helps to maintain temperature during MD simulations. We use energy equipartion theorem – in each degree of freedom, in equilibrium, an energy of ½ kT is stored :

$$E_{kin} = \frac{3}{2}\,\mathrm{NkT} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{m_i^1 v_i^2}}{2}$$

One can immediately see that the proper temperature may be maintained through atomic velocity scaling.

Simulation of an environment

In bionanotechnology the main solvent is water. Physicists have developed more than 20 computers models of water, of which all have some advantages but none in perfect (http://en.wikipedia.org/wiki/Water_model , Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, Pekka Mark and Lennart Nilsson , *J. Phys. Chem. A*, 2001, 105 (43), pp 9954–9960). The water molecule is particularly difficult to describe. We will use TIP3P model, similar to TIP4P presented below in a figure (fig. from the Internet: u-tokyo/~murayama)):

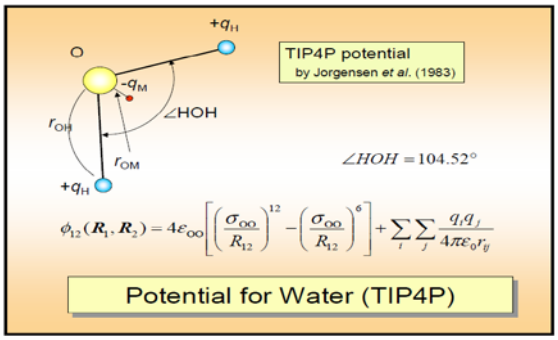


Fig. 74. TIP4P computer model of water molecule.

Sometimes we prefer to embed our studied protein in a droplet of water. This is good for SMD. More popular, especially for the system equilibration phase, is an approach in which Periodic Boundary

Conditions (PBC) are imposed. The space is divided into 27 cubes, the central cube contains modeled system, the 26 cubes have a copy (an image) of the primary box. If an atom leaves from the central cube, its image enters back to the primary box from the appropriate image box. In that way a number of atoms is kept constant and the central model "feels" an environment similar to that in an infinite crystal.

The main technical obstacles in practical MD simulations are very long computations times. Existing codes use up to 70% of CPU time for calculations of nonbonded interactions. The cost growths as N^2 with the size of the system. Since VdW or electrostatic interactions decay rather fast with a distance, these forces are simply neglected above certain cut-off distance. In practice a cut-off of 12-16 Ang. is sufficient for a protein modeling. Of course, one should carefully modify border regions of potentials, in order to maintain the proper asymptotic behavior.

A general scheme of MD simulations

1. Select a model (what do you want to study, what for, what is a source of your atomic coordinates)
   (a) A good source of structures is pdb
   (b) Protein should be "tuned" using special codes (VMD may help), missing atoms or missing residues should be added, immerse protein into a box of computer water (VMD again) TIP3P is a popular model
   (c) Remove waters located too close to protein atoms (? 2.3 Ang.)
   (d) Check waters inside the protein model. Keep that molecules which have at least 3 h-bonds
   (e) Determine the charge of charged amino acids: Glu, Asp, Arg, Lys, His – it is related to pH
   (f) Check protonation of His residues, check protonation of all charged residues with respect to assumed pH
2. Select the force field. Check that you have all necessary parameters.
3. Select cut-offs, size of the PBC box, frequency of non-bonded list regeneration.
4. Impose constrains on a protein. Equilibrate water shell/box for some 0.5-1 ns (up to 300K)
5. Heat gradually the whole system protein+water
6. Check the structure on a graphics (VMD)
7. Equilibrate the whole system without any constrains, in desired temperature, for at least 1 ns (discard these trajectories – useful for diagnostics only)
8. Generate the proper trajectory for the required time (1 ns – 10 ns , 1 fs time step is typical) Save snapshots (frames) from the trajectory each 100, or 1000 steps. Control energy, temperature etc. "on-the-fly".
9. Analyze collected trajectories (sometime up to 20 Gb of data) using your favorite graphical desktop and code, for example, VMD, Maestro, PyMol, Yasara, etc.
10. Repeat calculations using different code/ force field to reduce possible artefacts.

A study on mechanical strength of nanocompsite biomaterial – a caddisfly fiber

Some insects, such as the silkworm, spiders and caddisfly, can produce very strong thread. It serves to catch prey (spider web) and for protection. The composition of these materials remains partially unknown. They contain a big proportion of protein, together with mineral ingredients. In general, such silk is a composite material characterized by interesting mechanical properties. The toughness

of spider silk may surpass that of Kevlar (bullet resistant vests). Bionanotechnology often employs biomimetic approach, which is based on observing the nature. Our aim here is to understand why natural materials have mechanical properties that are so perfect.

We may use classical modeling MD variant called steered molecular dynamics. The idea is very simple: to some of the atoms of a protein a harmonic potential is added, holding them still. To other atom, or atoms (eg. on the other side of the molecule), a force is applied through a spring having a spring constant k (it links the towed atom to an additional atom to which the force is applied). When the towing atom moves away from the towed atom, the spring's length increases and, in accordance with Hooke's law and the third Newton's law of motion, spring acts on the towed atom with increasing force. The force leads to unwinding of the protein, which we observe on a computer screen. This way we can study mechanical properties of different proteins or protein domains. Such a study will be a topic of our exercises.

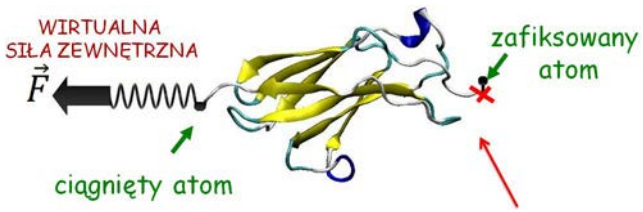


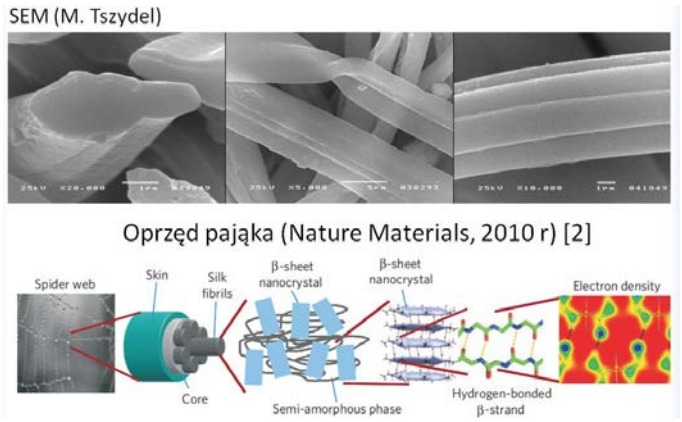
Fig. 75. A principle of Steered MD.

We will study a fragment of caddisfly protein (Trichoptera). These insect make a glue working in water. We hope that better understanding of this material may lead to new types of surgical tissue adhesives. The caddisfly fibers were studied experimentally and computationally in the Department of Biophysics and Medical Physics, NCU in Torun, Poland:

(J. W. Strzelecki, J. Strzelecka, K. Mikulska, M. Trzydel, A. Balter, W. Nowak, *Nanomechanics of new materials – AFM and computer modelling studies of Trichoptera silk*, Cent. Eur. J. Phys. 9 (2011) 482 - 491.)

Fig. 76. Trichoptera make a fiber which glue stones.

Let us have a look at a spider silk fiber structure:



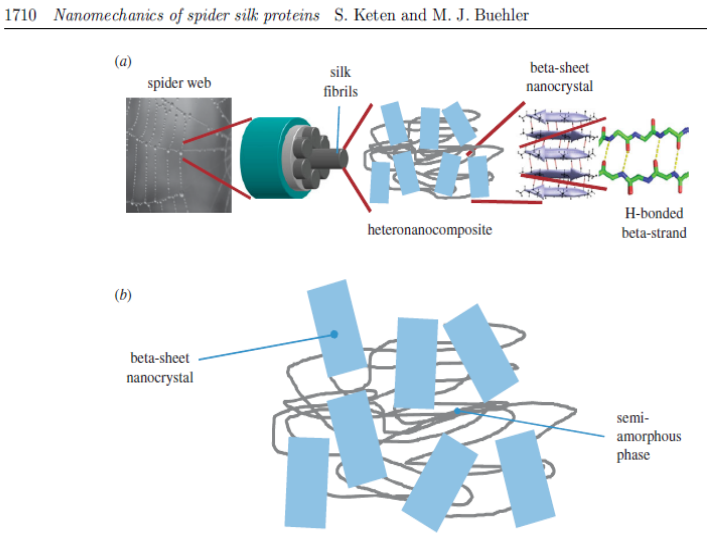


Fig. 77. The complex structure of caddisfly fiber. It is nanocompopsite biomaterial.

For simulations we select only a small, but characteristic, fragment of fibroin. It is a main protein in the fiber:

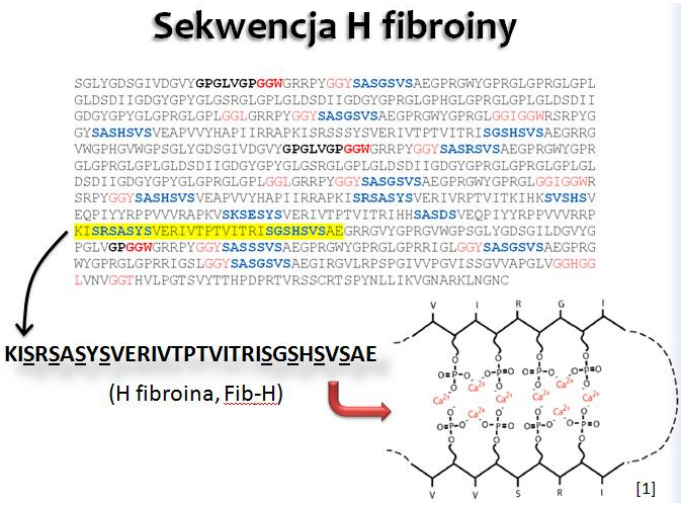


Fig. 78. A fragment of H- fibroin (H-f) protein sequence will be used in computer simulations (KM).

Note that the H-f protein fragment contains 8 serine residues (S, Ser). This amino acid contains the OH group in its side chain. It is prone to enzymatic phosphorylation. The PO^{4-} groups (negatively charged) are added. If Ca^{+2} ions for some reasons are present in model protein, they may "glue" the negatively charged phosphorylated beta strands. W expect that this phenomenon should increase the mechanical strength of the fiber. We will test this hypothesis computationally. We will use the SMD method and will attempt to rupture the protein in at least two cases:

(1) when serines are phosphorylated and contain calcium ions
(2) when serines are phosphorylated but there is no calcium

Of course one may expect that the two remaining cases of SMD simulations are also informative, ie. when serines are not phosphorylated and there is calcium (3) or calcium is absent (4). All these cases were studied in the past, during practical exercises we will limit studies to cases no. (1) and (2).

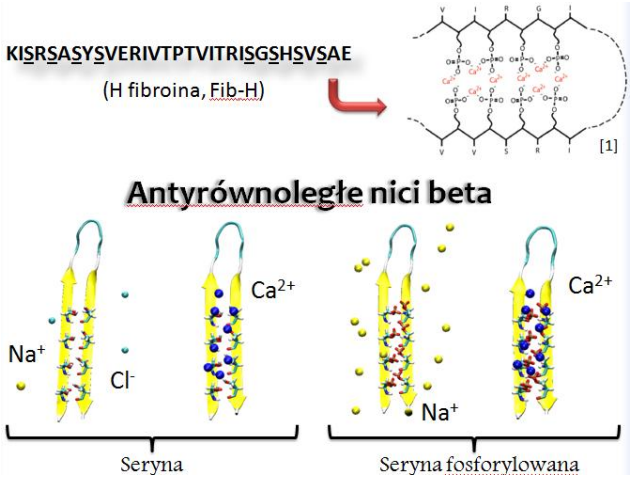


Fig. 79. Four models of beta H-fibroin fragments, used to study on the role of phosphorylation and Ca+2 in fiber nanomechnics (fig. KM and WN).

We need to select SMD parameters, some hints are presented in instructions to the exercises:

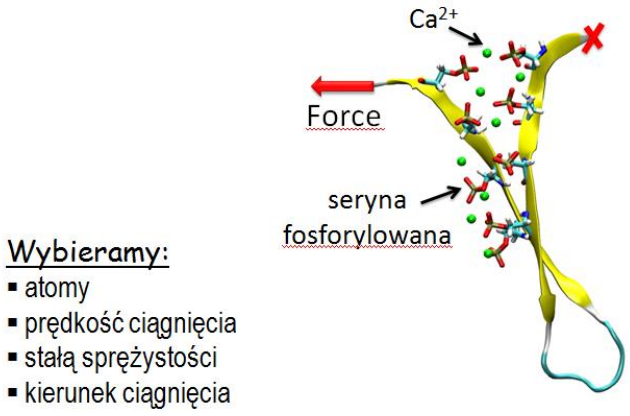


Fig. 80. Suggested set-up for simulations (fig. KM).

It is always recommended that one should know in advance possible outcomes of computer experiments. Here are our earlier results of mechanical unfolding of fibroin fragments:

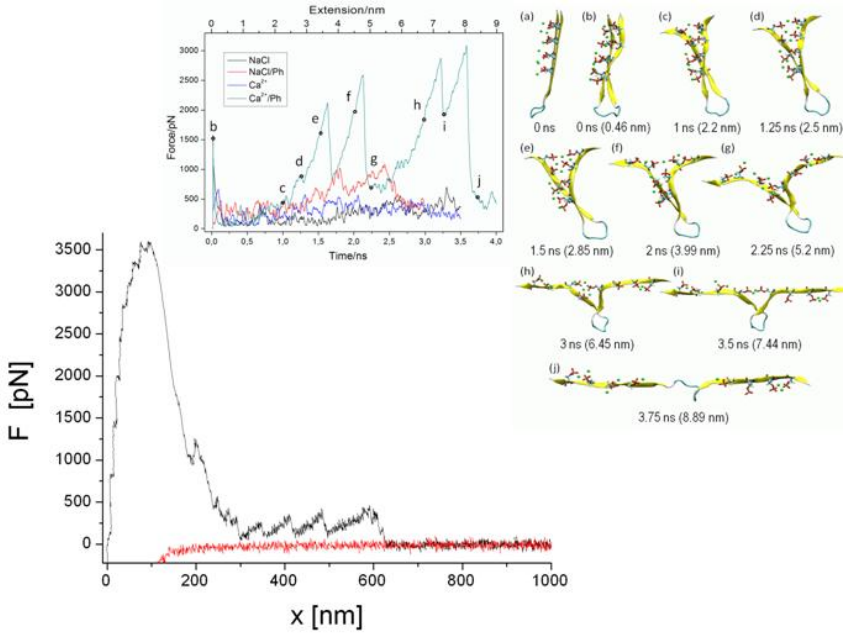


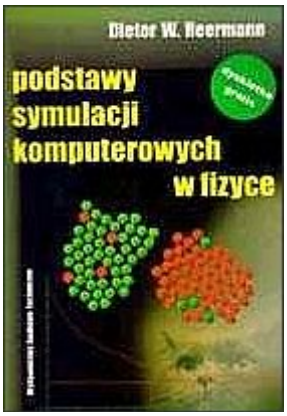
Fig. 81. Selected results of earlier SMD simulations of H-F fragments and AFM measurements of a single fiber (J. Strzelecka). Note how phosphorylation and a presence/absence of calcium ions affect the Maxim force. How the calculated force compares with the AFM curve?

Our conclusions were as follows: phosphorylation alone increases mechanical strength of a fiber by 100%. The phosphorylation in the presence of Ca^{+2} in the model increases a mechanical resistance by 800%-1100%. It would be interesting to know to what extent these results are sensitive to details of computational protocol. Should we accept these finding or reject? Good luck in computational invesitigations ☺.

Literature :

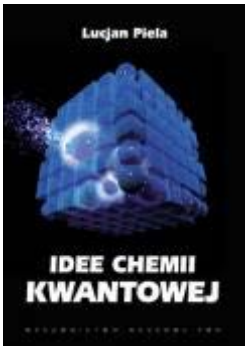
1. Podstawy symulacji komputerowych w fizyce 14.00zł

   Autor: Dieter W. Heermann
   ISBN: 83-204-2047-4
   Ilość stron: 173
   Data wydania: 1997



2. Manuale VMD i NAMDA:
   http://www.ks.uiuc.edu/Research/vmd/current/ug/
   http://www.ks.uiuc.edu/Research/namd/2.7/ug/

3. Handouts
4. L. Piela , Idee chemii kwantowej. (fragm.)
   Wydawnictwo Naukowe PWN 2005, ok. 55 zł, ISBN: 83-01-14000-3



5. Wiesław Nowak,
   “Applications of computational methods to simulations of proteins dynamics.” ,
   in “Handbook of Computational Chemistry”, Springer, 2012 – a book chapter, pp.1129-1114, ISBN 978-94-007-0712-2