5. Self-assembled quantum dots as an example of strained system

Self-assembled quantum dots

Molecular beam epitaxy (Fig.5.1) 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.5.1 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. 5.2) between the substrate and the deposited layer.
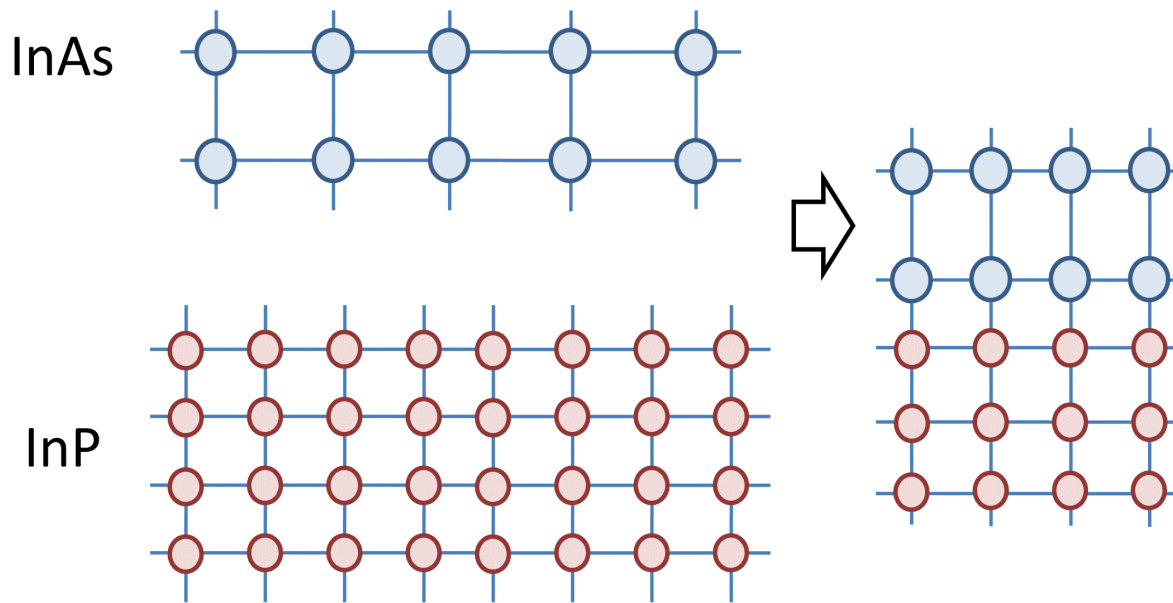
Fig.5.2 Schematics illustrating bond lengths alignment for an epitaxially deposited layer of lattice mismatch materials.

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.5.2, 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. 5.3). 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.
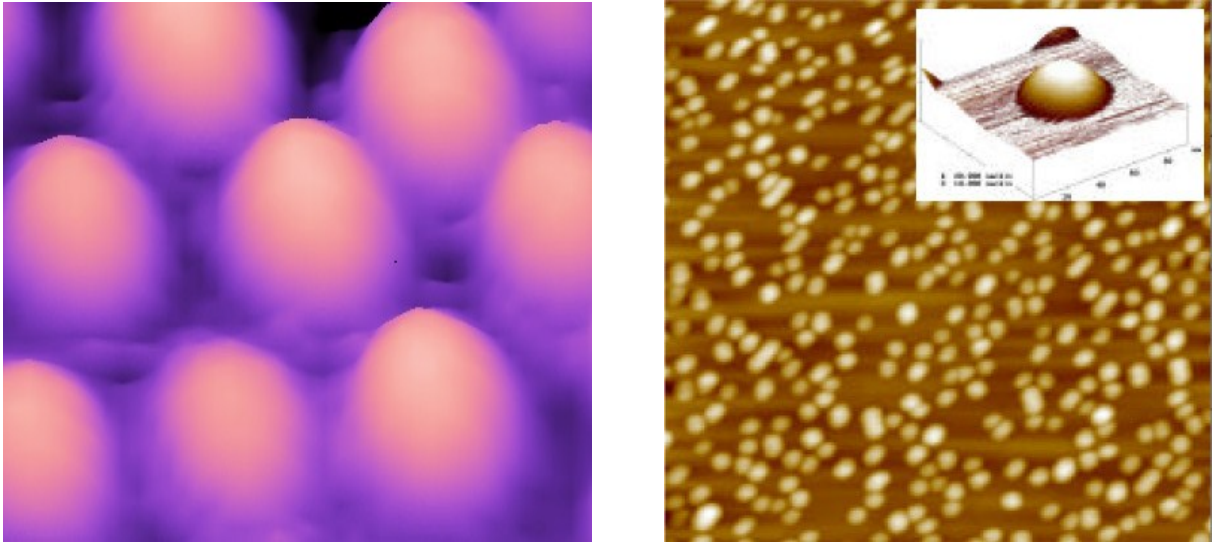
Fig.5.3 Atomic force microscope pictures showing formation of „islands" (quantum dots) on a lattice mismatched substrate (from the Internet).

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. 5.4 and Fig. 5.5).
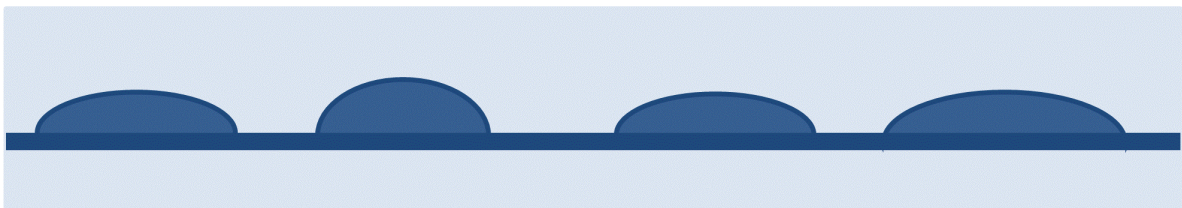

Fig.5.4 Schematic cross-section of quantum dot layer.

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.
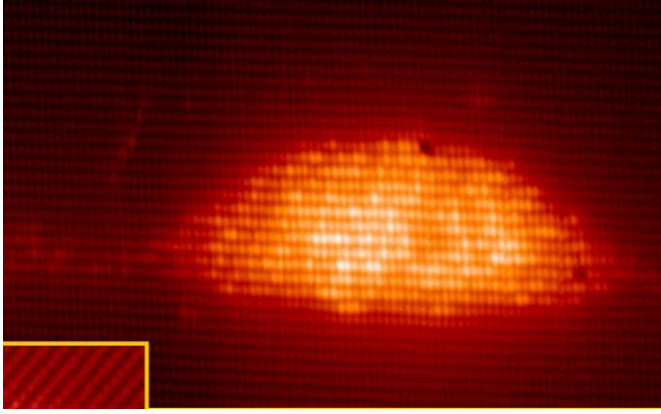
Fig.5.5 Cross-section scanning tunneling microscope picture (X-STM) of self-assembled InAs/GaAs lens type quantum dot (from: papers by P.M. Koenraad et al.).

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. 5.6).
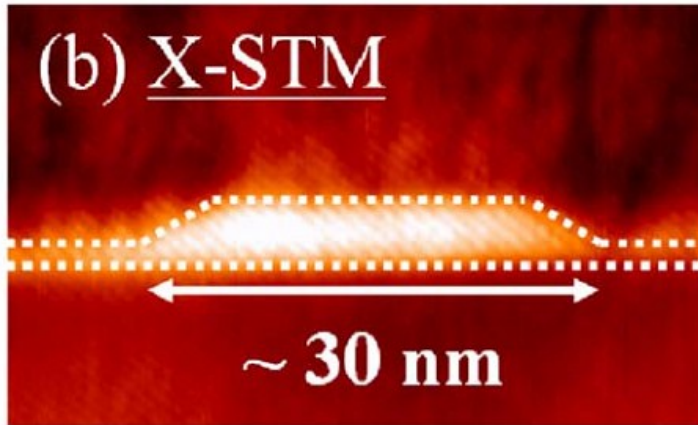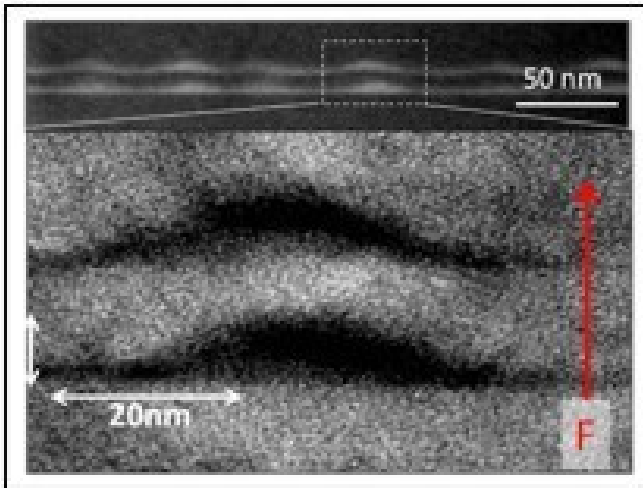

Fig.5.6 Cross-section scanning tunneling microscope picture (X-STM) of self-assembled InAs/GaAs disc type quantum dot (from: papers by P.M. Koenraad et al.).

Apart from single quantum it is possible to grow double quantum dots (Fig. 5.7) or even vertically stacked multiple quantum dots systems (Fig. 5.8).

Rys.5.7 Transmission electron microscopy (TEM) cross-section of a double InAs/GaAs quantum dot system (from the Internet).
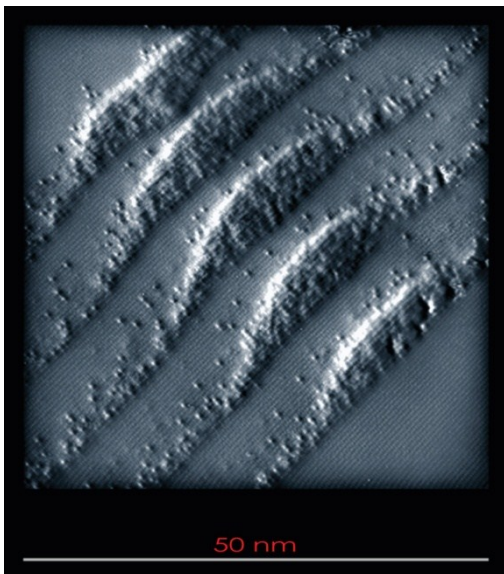
Rys.5.8 X-STM picture of a vertical stack of InAs/GaAs self-assembled quantum dots (from the Internet).

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

<u>Strain tensor</u>

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. 5.9), thus one can suspect that strain effects will play a major role and affect quantum dot spectral properties.
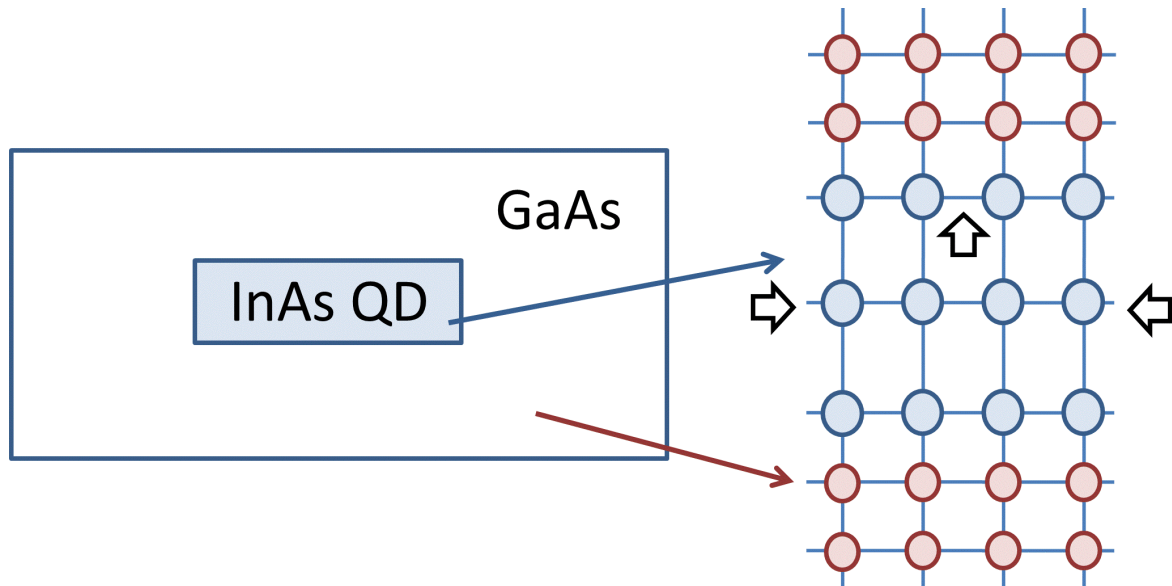

Fig.5.9 Schematics of interatomic distance deformation for a self-assembled InAs quantum dot embedded in a GaAs matrix.

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. 5.10). 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.
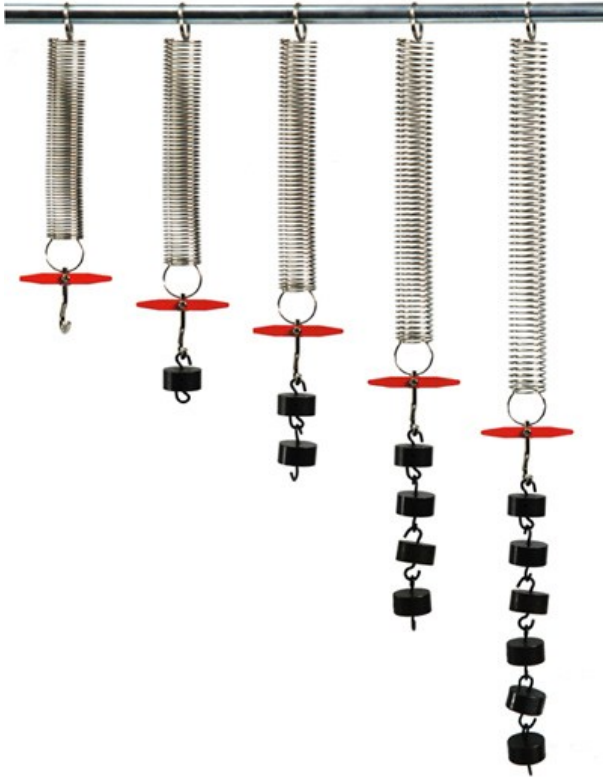
Fig.5.10 Metal spring deformation (strain) as a function of increasing weight.

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. 5.11) divided by the unstrained length.

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

Fig.5.11 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. $\varepsilon > 0$ corresponds to the tensile strain, while $\varepsilon < 0$ 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. 5.11).

Compression Tension Shear (torsion)

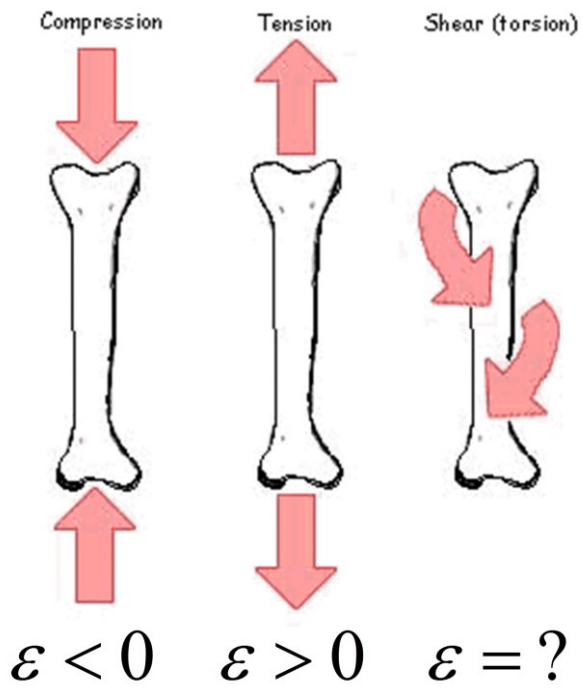$$\varepsilon < 0 \quad \varepsilon > 0 \quad \varepsilon = ?$$

Fig.5.11 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. 5.12 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.
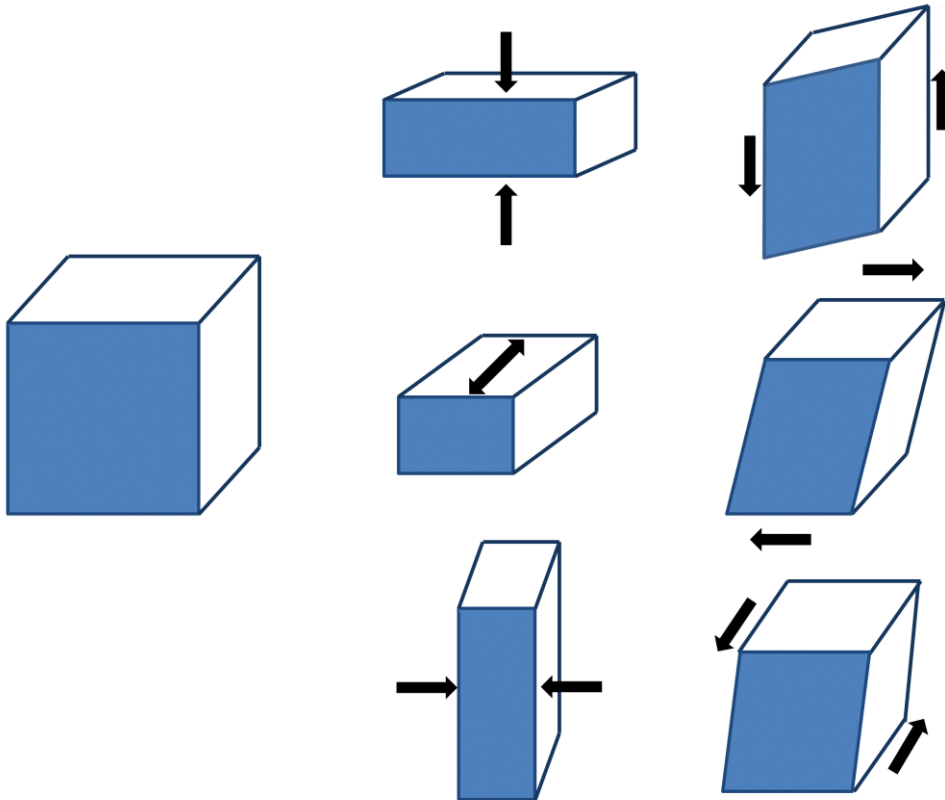
Rys.5.12 Schematics of different unit cube deformations.

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. 5.13) 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. ε_{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. $\varepsilon_{xx}=\varepsilon_{yy}\neq\varepsilon_{zz}$).
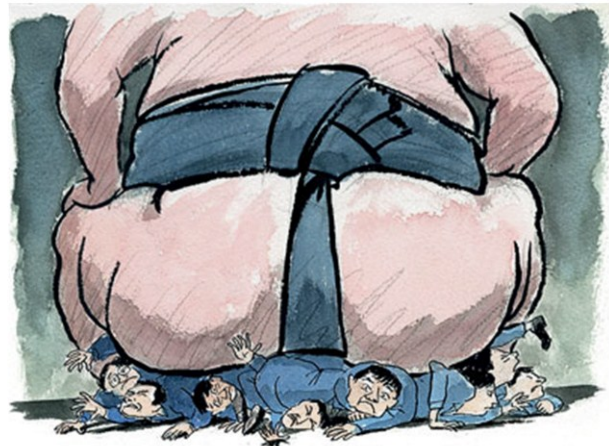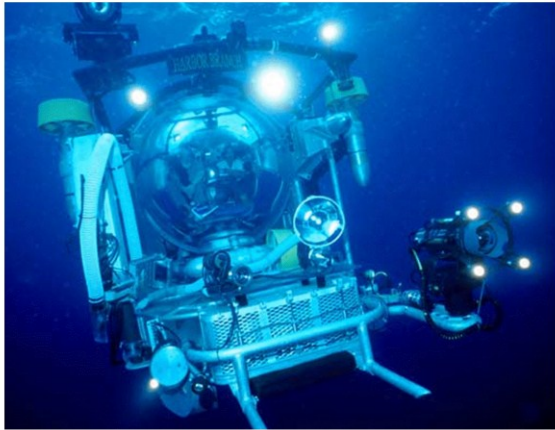
Hydrostatic

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

Biaxial (uniaxial)

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

Fig.5.13 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. 5.14, 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.

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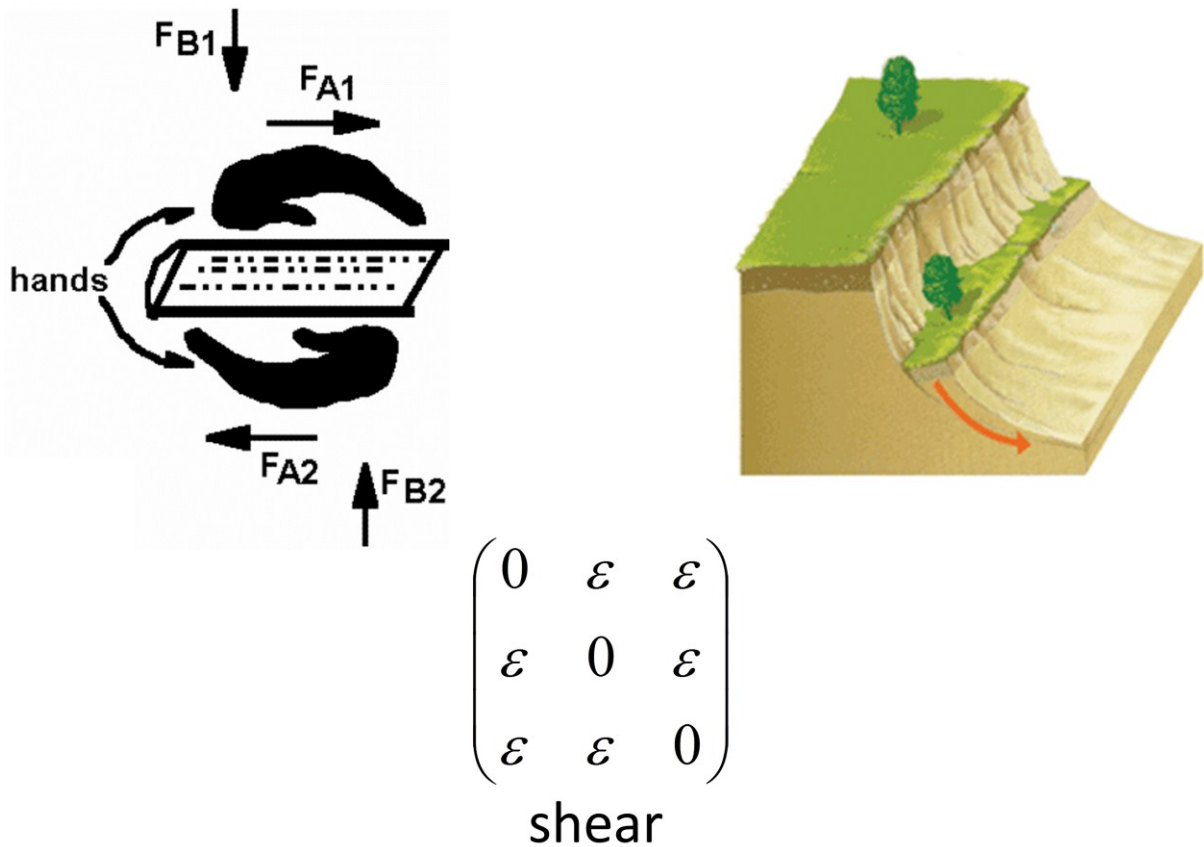
shear

Fig.5.14 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 $B(\varepsilon)=0$, as all linear dimensions are deformed in the same way. On the other hand pure axial deformation results in $Tr(\varepsilon)=0$, while $B(\varepsilon)\neq0$, which corresponds to volume conserving strain. Let us reiterate that general deformations can be complicated combinations of all basic kinds of strain.

<u>Stress and strain</u>

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. 5.15, 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.
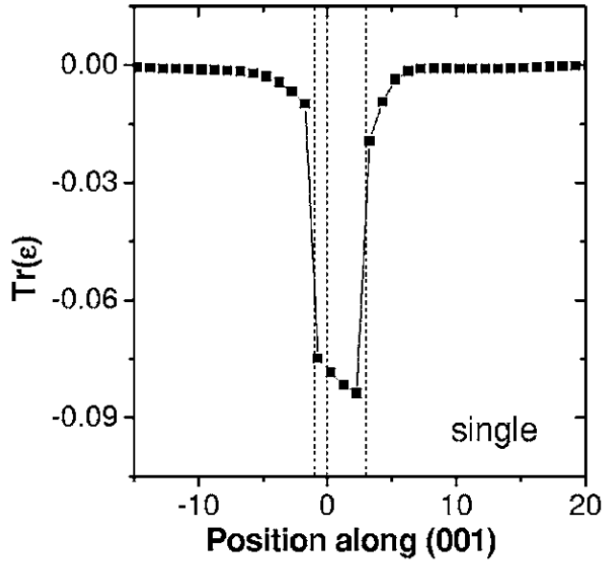

*Fig.5.15 The trace of the strain tensor Tr(ε) calculated as a function of position; a cross section along the growth axis through a geometric dot center of a lens type InAs/GaAs quantum dot (from Phys. Rev. B **74**, 195339 (2006)).*

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$_x$Ga$_{1-x}$As, Fig. 5.16) and cannot be simply accounted for by a continuum elasticity approach.
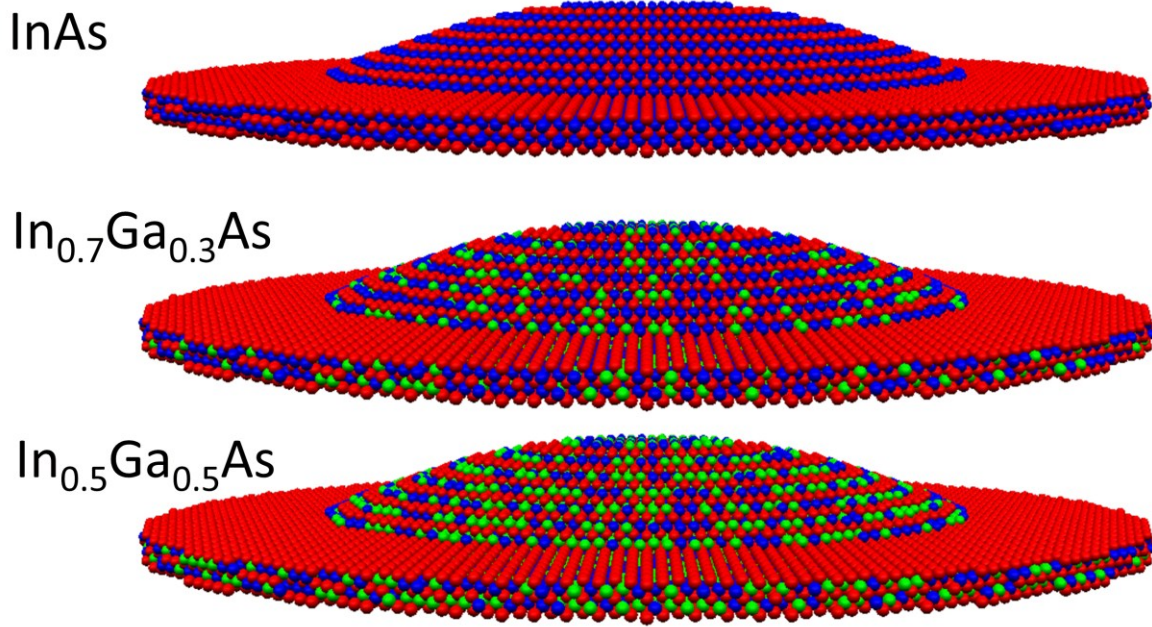

InAs

In$_{0.7}$Ga$_{0.3}$As

In$_{0.5}$Ga$_{0.5}$As

Fig.5.16 Positions of arsenic (red), indium (blue) and gallium (green) atoms for a wetting layer and lens type In$_x$Ga$_{1-x}$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:

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

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(θ)=1/3). In both summations **R**$_i$ is atom *i* position.

Finding strain minimum in the VFF approach correspond to finding atomic positions (**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. 5.17).
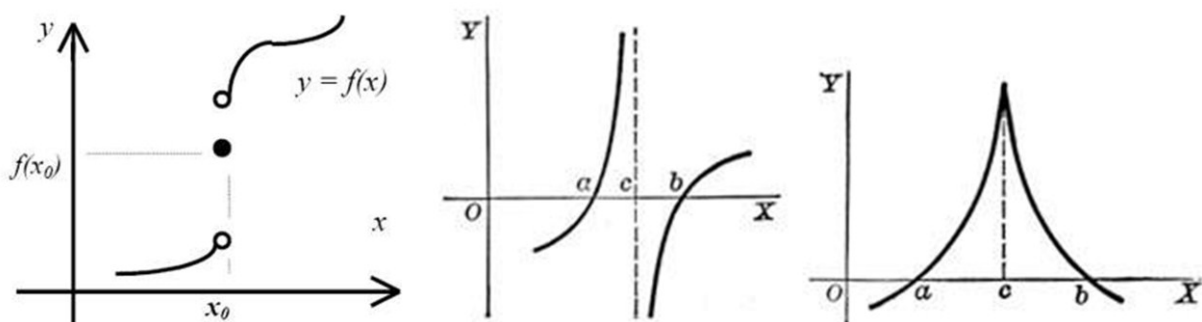

Fig.5.17 The shortest path to the top is usually the steepest one.

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. 5.17) function of N variables.

Rys.5.18 Przykłady funkcji nieciągłych i nieróżniczkowalnych.

The iterative process (Fig. 5.19) 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 ($f(x_0) > ... > f(x_k) > f(x_{k+1})$). 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.
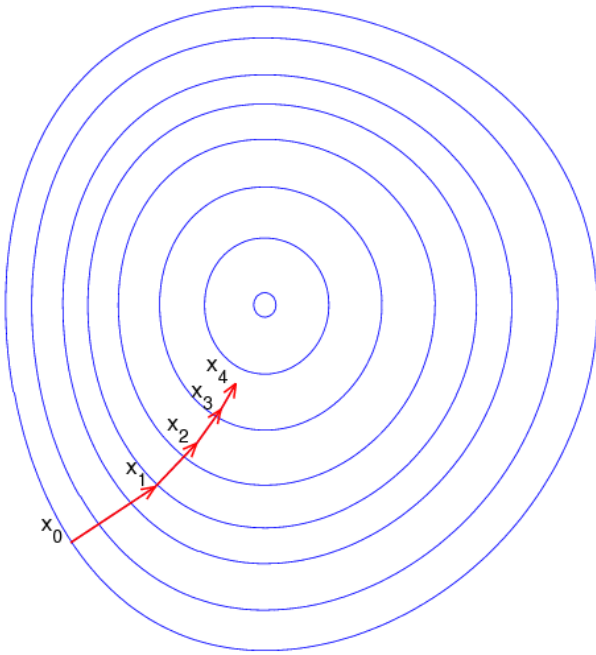
Fig.5.19 Several steps of the gradient descent method.

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.
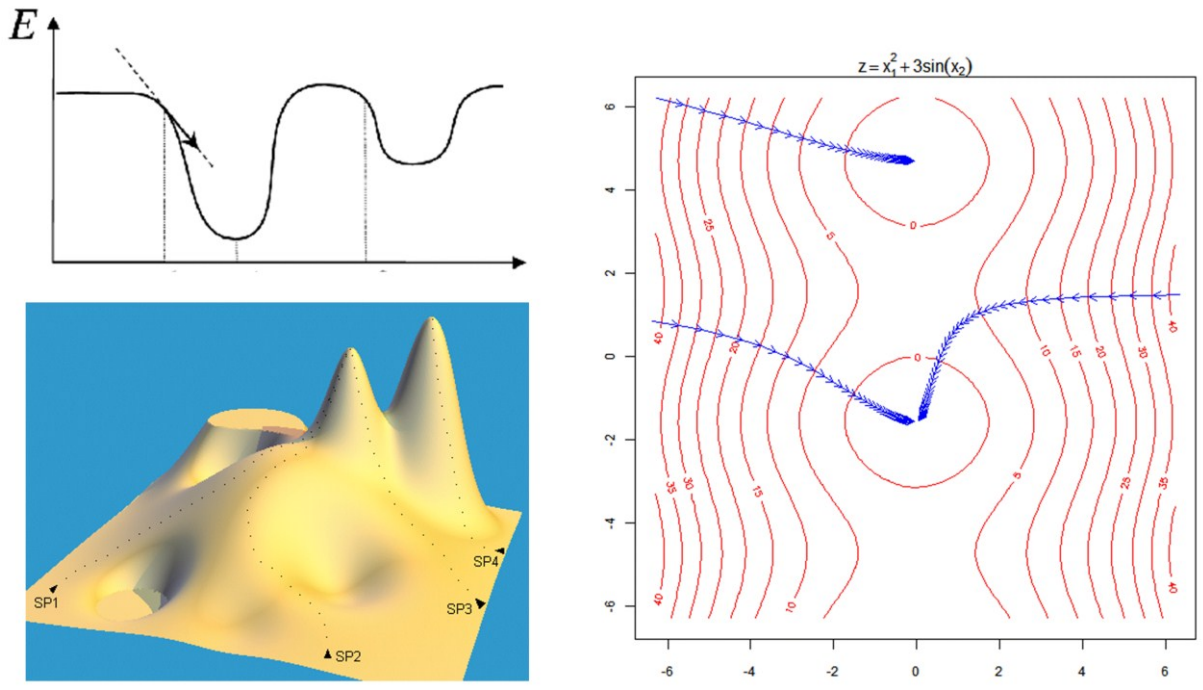
Fig.5.20 One- and two-dimensional functions with multiple local minima.