EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

New materials and technologies

doc. Eduard Belas, 22.12.2021

eduard.belas@mff.cuni.cz

http://semiconductors.mff.cuni.cz

Contents:

1. Introduction, Classification of Solid-state matter
2. Crystal growth methods of semiconductors
    - Growth from Melt, Liquid phase epitaxy, Vapor phase epitaxy, MBE, MOCVD
3. Materials.
    - Silicon - Si
        - Production of metallurgical grade silicon
        - Silicon purification (Siemens process)
        - Production of high purity polycrystalline silicon
        - Production of high purity silicon single crystals
        - Si wafer processing
        - Preparation of thin Si layers - CVD
        - Si wafer oxidation, properties of SiO_2
        - Doping (into melt, ion implantation, thermal diffusion
        - Photolithography
        - Porous Si
        - Quartz glass
    - Germanium - Ge
    - Gallium arsenide - GaAs
    - Cadmium telluride, Cadmium zinc telluride - CdTe/CdZnTe
    - Cadmium mercury telluride - CdHgTe
    - Indium phosphide - InP
    - Indium antimonite - InSb
    - Gallium nitride - GaN
    - Sapphire
    - Diamond
    - Silicon carbide - SiC
    - Graphene
    - Fullerenes
    - Perovskites
4. Characterization methods

Lecture materials have been produced within and supported by the project „Increasing the quality of education at Charles University and its relevance to the needs of the labor market" kept under number CZ.02.2.69/0.0/0.0/16_015/0002362.

Abbreviations

2D	2-dimensional
DC, AC	Direct, alternating electric current
ACRT	Accelerated crucible rotation technique
AFM	Atomic force microscopy
BGO	Bismuth germanium oxide
BLIP	Background limited infrared photodetectors
CCD	Charge coupled device
CMT	Cadmium mercury telluride
CVD	Chemical vapor deposition
CVT	Chemical vapor transport
CZ	Czochralski method
CZT	Cadmium zinc telluride

References

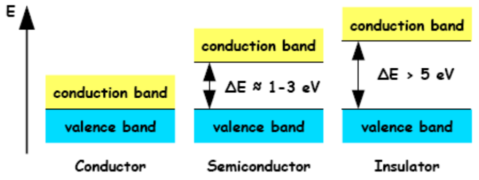
Handbook of Crystal growth, Second Edition, 2015, Ed. P. Rudolph, Elsevier, ISBN 978-0-444-63303-3, DOI: https://doi.org/10.1016/C2013-0-09791-5

Bulk Crystal Growth of Electronic, Optical and Optoelectronic Materials, 2005, Ed. P. Capper, Wiley, ISBN 0-470-85142-2

A. Owens, A. Peacock, Nucl. Instr. Meth. A 531, 18 (2004).

1. Introduction.

1.1. Classification of Solid-state matter



In Solids, there are many electronic bands that can be occupied by electrons. These bands may overlap each other or there is a gap between them where no possible electronic states exist. This gap is called **forbidden gap** (E_g). Electrons in the matter fill the electronic bands from the lowest energetic states (preferably). The last electronic band occupied by electrons is called **valence band** in that it consists of valence electrons from individual atoms. The first unoccupied electron band is called **conduction band** because the electrons in the fully filled valence band cannot contribute to the electrical conductivity of the material.

According to the width of the forbidden gap E_g.

Conductors (metals) ($Eg \leq 0$, bands overlap), Semiconductors ($0<Eg<5eV$), Insulators ($Eg>5eV$) (see Fig.1).

1.2. Classification of semiconductors:

a) According to the position of elements in the periodic table of elements: elementary - IV, Compounds III-V, II-VI (Fig.2).

b) According to the width of the forbidden gap E_g: Narrow gap ($0<Eg<1eV$), Wide gap ($1<Eg<5eV$) (Fig.2)

c) According to the shape of valence and conduction bands: direct (maximum of the valence band lies in the k-space below the minimum of the conduction band), indirect (maximum of the valence band is shifted in the k-space out of the minimum of the conduction band) (Fig.3)

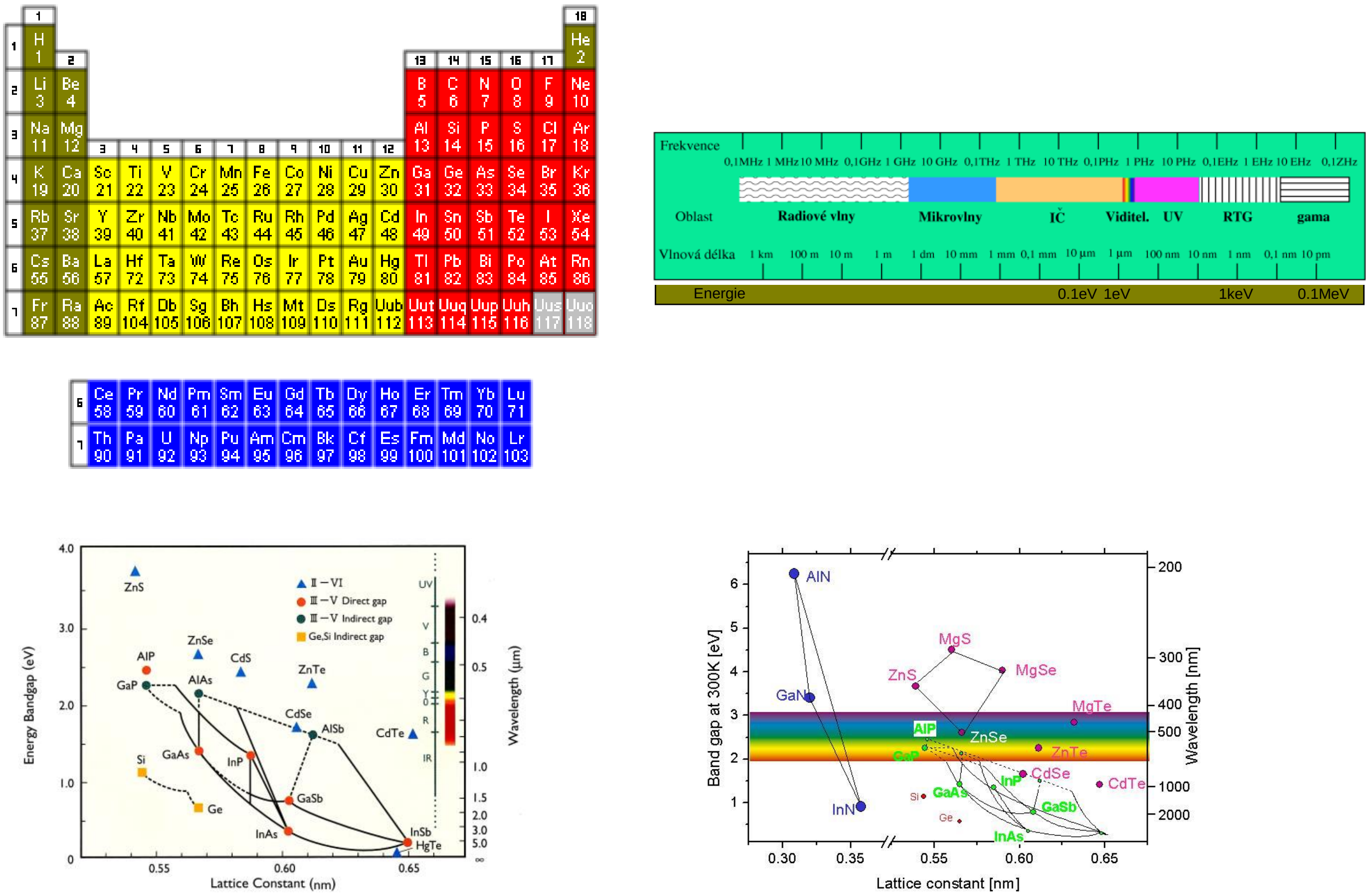
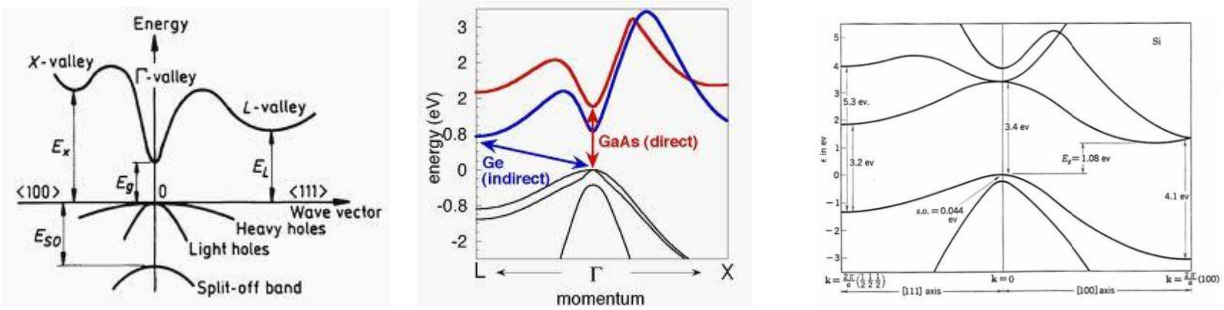


Fig.2. Classification of semiconductors according to the position in the periodic table, to the width of the forbidden gap and to the lattice constant, respectively.



GaAs GaAs, Ge Si

Fig.3. Classification of semiconductors according to the relative position of the valence and conduction band.

2. Single crystal growth methods of semiconductor materials

2.1 Growth from Melt

2.1.1. Czochralski method, CZ

In 1916, Czochralski demonstrated that the single crystals can be grown from the melt by a pulling mechanism. Three decades later, the first silicon (Si) single crystals were grown using the CZ technique. In case of Si, the polycrystalline silicon feed material is loaded into a cylindrical, bowl-shaped highly purified silica crucible and melted at approximately 1413 oC an inert gas (6N argon) atmosphere, see Fig.4.a. Crystal growth is initiated by dipping an oriented silicon seed crystal (which is mounted on a moveable pull rod) into the free surface of the silicon melt. The rotational movement of the seed in the range of 0.5–20 rpm according to a predefined program (mostly counter rotation to the crystal rotation) is used to create certain convective melt flows and the temperature of the melt is also very carefully monitored and controlled. The seed is then slowly withdrawn from the melt, which causes the crystallization of silicon atoms at the melt–crystal interface of the seed by forming a new crystal portion. The diameter of the growing crystal is first increased until a nominal diameter is achieved. This diameter is kept constant to provide a cylindrical shape of the crystal body throughout the total growth process. Finally, when most of the melt is consumed, the crystal diameter is decreased again to form the shape of a cone and the crystal is removed from the residual melt. After cooling down, the as-grown crystal is removed from the growth facility. The resulting product (ingot), which consists of a single crystal, can reach up to 450 mm in diameter and up to 3 m in length.

Modification Fig.4b-c.:

a) Liquid Encapsulated Czochralski - LEC CZ (GaAs, CdTe, InP...)
A boron trioxide (B_2O_3) has turned out to be best suited to fully encapsulate the charge for synthesis and growth of compound semiconductors. Boron oxide lid is used to prevent melt decomposition due to the leak of the more volatile component and to form a separating layer between crucible and charge.
b) Using nonstationary magnetic field for melt mixing – Foucault’s eddy currents

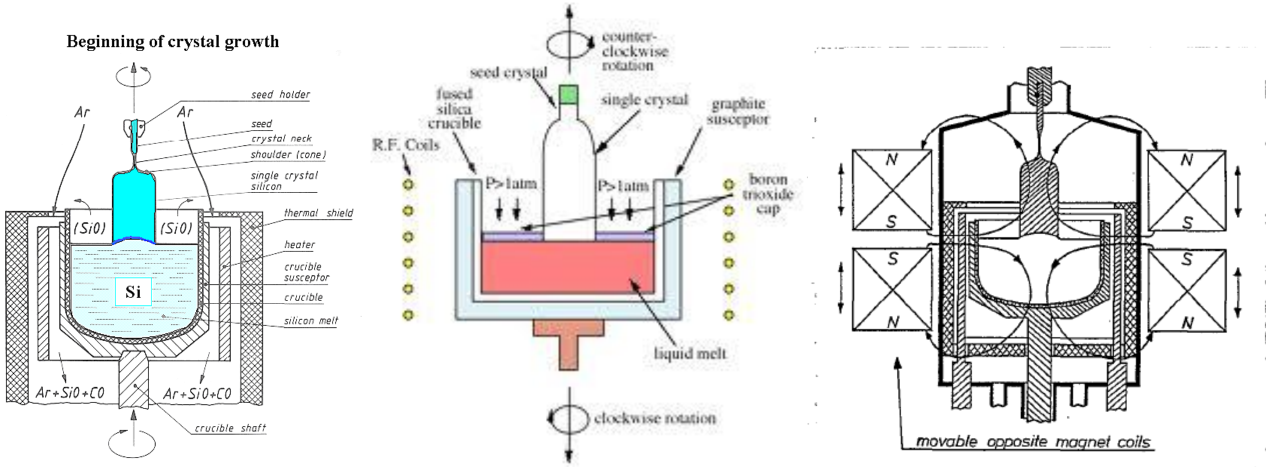


Fig.4. Czochralski growth of Si, LEC CZ growth of GaAs, Mixing of the melt by Eddy currents

2.1.2. Floating zone method (Pfann method, FZ)

Zone melting is a technique that was invented for purifying congruently melting materials in a physical way. Originally, a narrow local induction heating (induction coils) generates a molten zone in the solid substance being in a horizontal boat-like crucible made of a material that must not pollute that substance. The coil generates a high-frequency magnetic field and eddy currents are induced in the conductive object in the middle of the coil, which heats it. Even for small coils, a current of tens A is required. If the molten zone is continuously shifted through the material in the boat by the relative movement between boat and heater, impurities are partly gathered in the molten zone due to their almost higher solubility in the liquid state according to the segregation coefficient $\boldsymbol{k_o = C_S/C_L}$, where C_S and C_L are the solubility in the solid and in the liquid states, respectively. In that way, impurities are traveling with the molten zone to one end of the rod where they are concentrated in the last solidifying material. As a result, most material of the rod becomes purer excluding the end part, which is cutted-off. Multiple zone melting enables to produce high purity grades semiconductors.

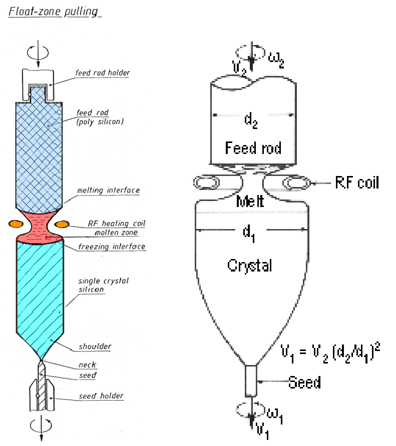
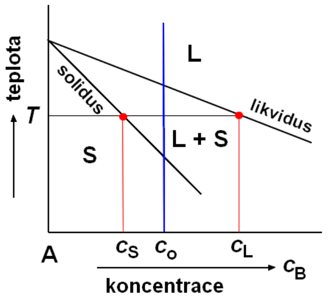


Fig. 5. FZ method

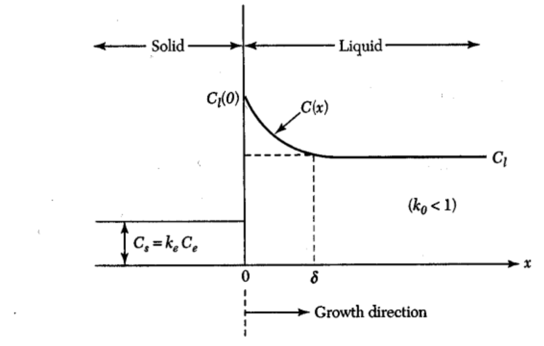
2.1.3. Segregation (distribution) coefficient

Crystal growers are usually interested to the segregation of impurities and dopants occurring in melt-grown semiconductor crystals, which include directional solidification/crystal growth, and purification achieved by zone refining. During solidification (crystallization), the solubility of the impurities is different in the solid and liquid phases, and therefore the melt is enriched or depleted by the impurity during solidification. The different solubility of the impurity for a given temperature T is expressed by an equilibrium segregation (distribution) coefficient $\boldsymbol{k_o = C_S/C_L}$, where C_S and C_L are the soluble impurity concentration in the solid and in the liquid phase, respectively, at the crystal–melt interface. The $\boldsymbol{k_o}$ is a fundamental thermodynamic property of a binary solid–liquid system and not a function of the temperature or hydrodynamic melt properties. In case of $k_o<1$, impurity-enriched melt layer is formed at the crystal–melt interface, since the rate of crystallization is generally faster than the diffusion of the impurity in the melt. In case of $k_o>1$, impurities are preferably absorbed



into the crystalline phase. The former case is more common than the latter, but both situations occur.

Actually solute boundary layer having some thickness δ, develops ahead of the crystal–melt interface. The concentration of impurities, $C_L(x)$, within this boundary layer changes from its initial equilibrium value, $C_L(0) = C_0$ at the interface, to a lower, or higher, value $C_L(\delta) = C_L$ in the bulk melt, depending, of course, on the value of k_o relative to unity. Therefore, an effective segregation coefficient k_{eff} is introduced



Doping distribution near the solid-melt interface.

$$k_{eff} = \frac{C_S}{C_L} = \frac{k_O}{1 - k_O \exp\left(\frac{-R \cdot \delta}{D}\right)}$$

Effective segregation coefficient
R = Crystal growth velocity
δ = Diffusion length, fit. parameter
D = Dif. coeff. of impurity in the melt

If we suggest perfect mixing of the melt, we can calculate the Gulliver–Scheil segregation coefficient, as a function of the fraction of the molten charge solidified g, which is given by $C(g) = k \cdot C_O (1-g)^{k-1}$, where C_o is the soluble impurity concentration in the liquid phase, k is equilibrium or effective segregation coefficient.

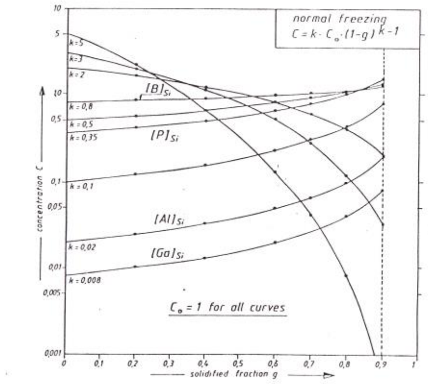


Fig. 7. Distribution of the impurity concentration in silicon.

The standard horizontal floating zone method fails for silicon growth, because any material for the boat including quartz markedly reacts with the silicon melt, and silicon adheres mostly at the boat causing cracks during cooling down due to different thermal contraction. Therefore the special crucible-free vertical FZ method using a seed crystal had to be invented for silicon, where the “floating” molten zone is neatly stabilized by the strong surface tension and is located between the vertically positioned growing crystal at the bottom and the melting silicon feed rod at the top. It is commonly generated by the contactless inductive heating via the RF magnetic field of a one-turn induction coil. The zone “floats” upward due to the relative movement between the heating coil and the silicon rods. The induction coil and the RF generator are fixed and crystal and rod are moved downward while rotating. At a suitably selected speed of rotation and shift of the zone, a monocrystalline ingot is formed. At the beginning of the growth, a high-quality oriented seed without cracks and dislocations is used, where the grown crystal retains the crystallographic orientation of the seed. Silicon single crystals can be grown in a dislocation-free structure after eliminating all dislocations by the thin-neck method. Because the FZ method does not use a crucible, the grown crystal is not in contact with anything, it is only in a vacuum or an inert atmosphere, therefore there are very few impurities in the grown single crystal compared to the CZ method. The main problem of the FZ method is holding of the widest possible molten zone.

2.1.4. Bridgman-Stockbarger method

In the Bridgman method, the polycrystal or individual elements are placed in a crucible and/or in the growth ampule, where they are sealed after evacuation. The crystal growth is performed in a multi-zone growth furnace with a temperature in the separate zones above and below the melting point of the substance to be crystallized. The zones with different temperatures defines the position and shape of the solid/liquid interface and the temperature gradients in the melt and the crystal. The ampoule is placed to the zone, where a temperature higher than the melting point is gradually set. At this temperature, the material is completely melted and, in the case of individual elements, they are compounded at the same time. The ampule is then gradually shifted to the second zone in the furnace with the temperature below the melting point. Single crystal begins to grow from the coldest end of the ampule. The temperature gradient is shifted by a downward or upward movement either of the furnace with the stationary ampule or the ampule with the stationary furnace, respectively. There are both horizontal and vertical arrangements of the furnace. With a horizontal arrangement, the vapor pressure balance above the melt is established more quickly due to the larger area. On the other hand, the resulting ingot has a semicircular cross section. At the beginning of growth, a high-quality oriented seed without cracks and dislocations is often used, where the grown crystal retains the crystallographic orientation of the seed. The growth process is followed by a controlled cooling to room temperature in which the cooling rate is determined such that thermal stress is below the critical shear stress for dislocation multiplication and glide processes.

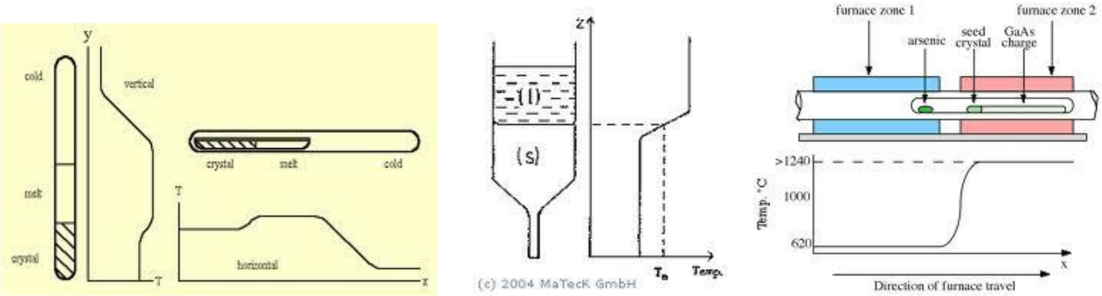


Fig.8. Bridgman method

Modification:

a) Liquid Encapsulation (GaAs, CdTe, InP...) - vertical configuration

Similar to CZ method, a boron trioxide (B_2O_3) has turned out to be best suited to fully encapsulate the charge for synthesis and growth of compound semiconductors. Boron oxide lid is used to prevent melt decomposition due to the leak of the more volatile component and to form a separating layer between crucible and charge.

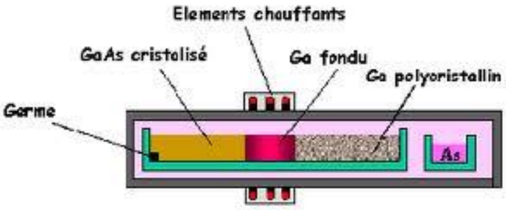
b) Additional source of more volatile component as a source of saturated vapor pressure above the melt. The vapor pressure is the same everywhere in the ampoule according to Pascal's law and depends on the lowest temperature in the ampoule. Increased vapor pressure prevents decomposition of the compound and leakage of volatile components.

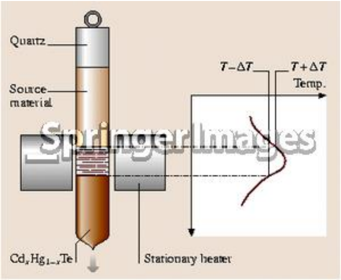
c) High vapor pressure of the inert gas (6N Ar) over the melt also prevents decomposition of the compound and leakage of volatile components.

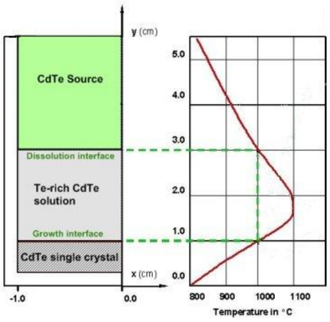
d) Vertical Gradient Freeze method (VGF).

This method is a modification of the Bridgman method, where the temperature gradient is created and controlled electronically in the furnace and the furnace itself and the ampule do not move.

2.1.5. Traveling heater method (THM) (Similar to FZ)







The traveling heater method (THM) has been the most popular form of traveling solvent techniques for the last decades. It combines zone melting and growth from solutions. In THM, a molten zone is made to migrate through a solid homogeneous source material by the slow movement of the ampoule relative to the heater. Different types of solvent zone heaters were used, either with electrical resistance heating (RH) or radio frequency heating (RF) or optical heating. The solvent zone dissolves the source material at its hot upper interface and deposits, in near equilibrium conditions, a material of nearly the same composition at the lower interface, which is colder than the upper interface because of the movement of the heater. Matter transport from the source to the growing crystal occurs by convection and diffusion across the solvent zone under the influence of the temperature gradient resulting from the movement of the ampoule relative to the heater. Significant prerequisite of this method is an appropriate temperature profile of the growth furnace. It has to possess a sharp temperature peak producing a narrow molten zone (roughly of length close to the ampoule diameter) and a sharp temperature gradient at the growth interface to prevent constitutional supercooling. One of two arrangements is usually used as a pulling system: either traveling down the ampoule in a stationary heater or moving up the heater along a stationary ampoule. There are again both horizontal and vertical arrangements of the furnace.

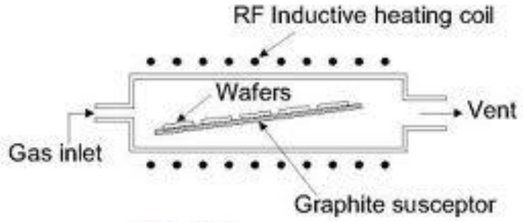
2.2. Epitaxial methods (Thin film preparation)

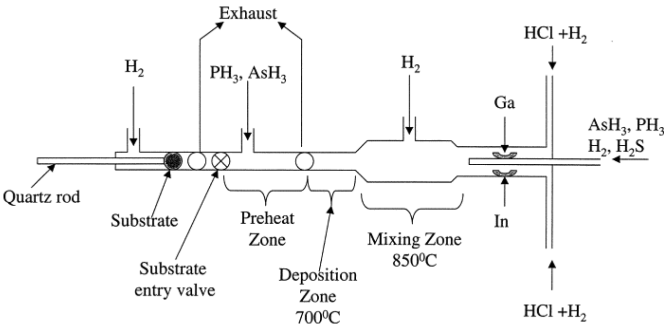
Epitaxy is a process in which a thin crystalline layer grows on the surface of a substrate, in which the crystal lattice of the newly formed layer is directly connected to the crystalline lattice of the substrate. The term epitaxy comes from the Greek words epi "on" and taxis "ordered". Epitaxial layers can be grown from gaseous, liquid or solid precursors (substances from which we want to grow the layer). If the layer is deposited on a substrate with the same composition, we call it homoepitaxy, otherwise heteroepitaxy. In homoepitaxy, a thin layer grows on a substrate of the same material. We used this method if we want to grow a cleaner layer than the substrate or if the new layer should be doped by different elements.

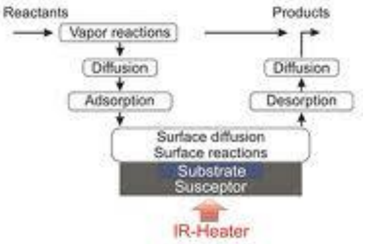
Epitaxy is used in silicon-based manufacturing processes in the manufacture of bipolar transistors and modern CMOS, but is also important for the preparation of compound semiconductors such as silicon germanide (Si-Ge, gallium arsenide (GaAs), indium phosphide (InP), cadmium telluride (CdTe), gallium nitride (GaN)) and others.

2.2.1. Vapour Phase Epitaxy (VPE, MOVPE, MOCVD)

(Vapour Phase Epitaxy, Metal-Organic Vapour Phase Epitaxy, Metal-Organic Chemical Vapour Deposition)

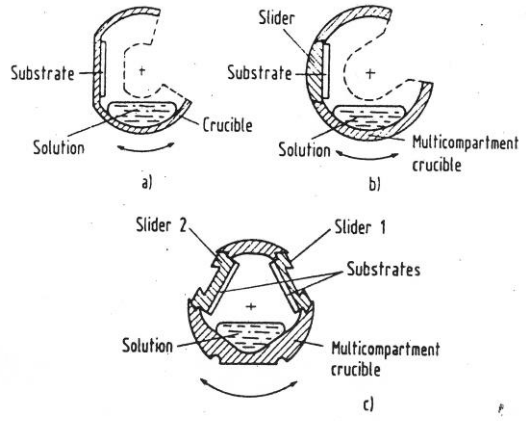






2.2.2. Liquid phase epitaxy (LPE)

Liquid Phase Epitaxy is a method of growing semiconductor crystalline layers on solid substrates from the molten material. The semiconductor is dissolved in a melt of material enriched in one of the components. Under conditions that are close to the equilibrium of the liquid material and the substrate, the growth takes place slowly and evenly. This occurs at temperatures well below the melting point of the deposited semiconductor. LPE forms a monocrystalline layer from the liquid phase, typically at a rate of 0.1 V to 1 µm per minute. Equilibrium conditions are highly dependent on the temperature and concentration of the dissolved semiconductor in the melt. Forced cooling of the molten mixture can control the growth of the layers and the intential doping is performed by adding dopants to the melt.



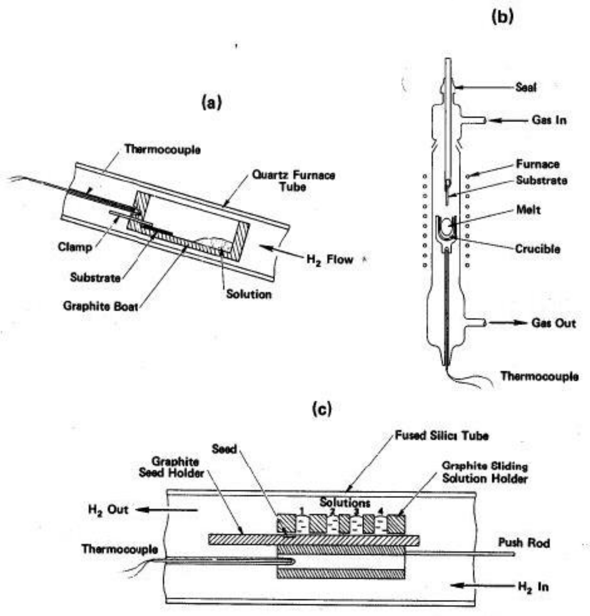


Fig. 5. Drawings of LPE systems: (a) tipper system after Nelson, (1963); (b) dipper system after Rupprecht, (1967) and (c) slider system after Panish, et al. (1969).

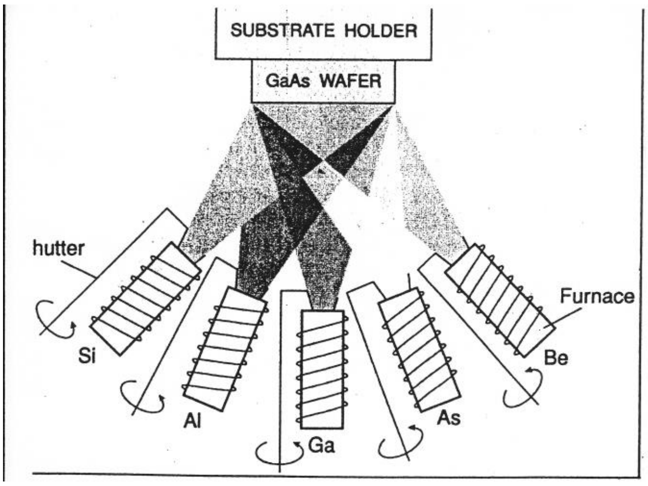
This method is mainly used for the growth of compound semiconductors. It is thus possible to produce very thin, uniform and high-quality layers.

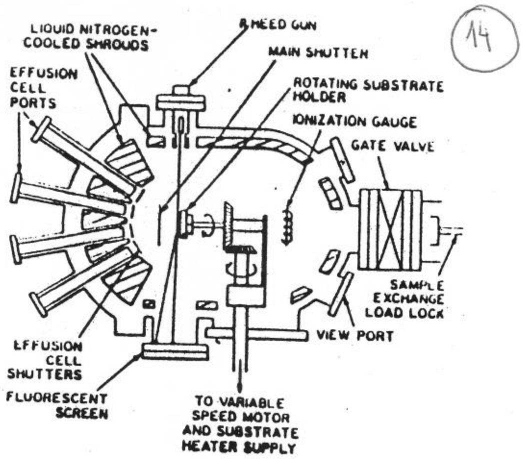
A typical example of the use of this epitaxial method is the growth of ternary and quaternary compounds (groups III-V) on gallium arsenide (GaAs) substrates. In this case, gallium is often used as a solvent. Another frequently used substrate is indium phosphide (InP). The first LEDs and lasers in the world were prepared on GaAs heterostructures using LPE. Other substrates such as glass or ceramics can be used for special purposes. For a good adhesion and to avoid stress in the interface of the grown layers, lattice constant and the thermal expansion factors of the substrate and the layer should be similar.

2.2.3. Molecular beam epitaxy (MBE)

Molecular beam epitaxy takes place in a high vacuum (10^{-8} Pa). The most important feature of MBE is the low material deposition rate, which allows epitaxial growth. However, the low growth rate requires a proportionally higher vacuum to achieve the same levels of impurities as other growth techniques.

For MBE, the source materials are solids. Very pure elements such as gallium or arsenic are heated in the closed Knudsen cells until they begin to sublime slowly. The gaseous material then condenses on the substrate, where different substances can react with each other. For example, single crystal of gallium arsenide (GaAs) is formed from gallium and arsenic atoms.

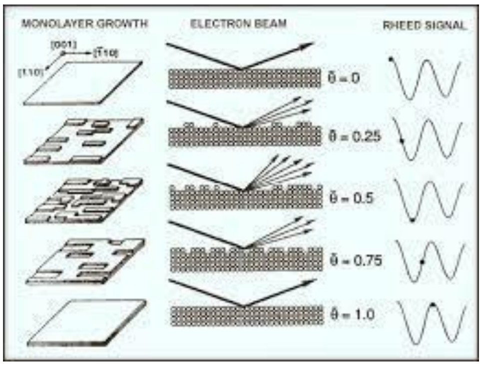


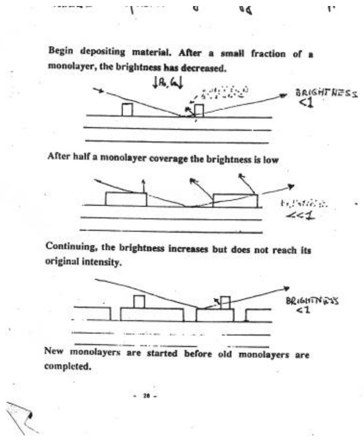


The term beam in the name of the method means that the evaporated particles do not react with each other or with any other gases in the vacuum chamber until they reach the substrate. This is due to the long mean free paths of the atoms due to extremely high vacuum. The high vacuum is created and maintained in the growth chamber by a system of cryogenic pumps and cryopanels which are cooled with liquid nitrogen or cold nitrogen gas to a temperature close to 77 K (-196 ° C). Surfaces with such a low temperature act as a trap of the atoms in a vacuum, so the high degree of vacuum required to deposit the material on the substrate is achieved. At the same time, the substrates on which the layers grow are attached to a rotating substrate, which can be heated to several hundred degrees Celsius during the process.

The RHEED (Reflection High Energy Electron Diffraction) method is often used in MBE to monitor the growth of individual crystalline layers. The computer controls the shutters in front of each whole (furnace), which very gently affects the thickness of each layer up to the level of one atomic layer. A RHEED system requires an electron source (gun) and photoluminescent detector screen. We use low-angle electron beam as a probing beam and we measure directly in the growth chamber under high vacuum (long free electron length between electron scattering with other atoms) intensity of the output RHEED signal scattered from the surface. The highest intensity of the

RHEED signal is in case of the smooth surface. If atomic layer starts to grow, intensity decrease and after filling of the half surface intensity starts to increase.

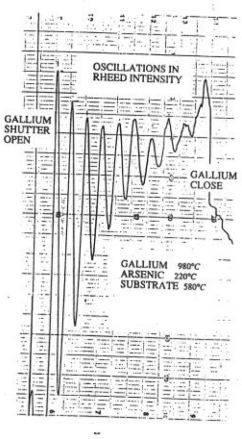



Begin depositing material. After a small fraction of a monolayer, the brightness has decreased.

After half a monolayer coverage the brightness is low

Continuing, the brightness increases but does not reach its original intensity.

New monolayers are started before old monolayers are completed.



A complex heterostructures consist of the layers with different materials can be produced by this method. MBE made it possible to create heterostructures where electrons are enclosed in a small volume - quantum wells or even quantum dots. Such layers are now a critical component of many modern semiconductor devices such as semiconductor lasers and LEDs. Molecular beam epitaxy is also used for the growth of some types of organic semiconductors. In this case, molecules are deposited on the substrate instead of atoms.

Materials produced by MBE:

IV: Si, Ge, stranged Si, Si-Ge

III-V: AlGaAs, AlGaInP, AlGaN, AlGaP, GaAsP, GaAs, GaN, GaP, InAlAs, InSb, InGaN, GaInAlAs, GaInAlN, GaInAsN, GaInAsP, GaInAs, GaInP, InN, InP

II-VI: ZnSe, HgCdTe, CdTe, ZnO, ZnS,

3. Materials

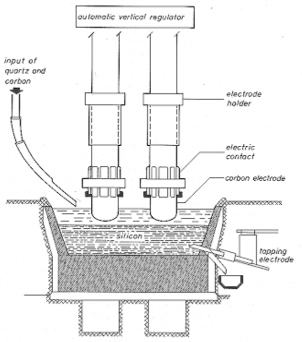
3.1. Silicon – Si

After oxygen, silicon is the second most abundant element in the Earth's crust. According to the latest available data, they make up 26-28% of the earth's crust. Silicon is not found in its pure form and may only be found in compounds. In seawater, its concentration is relatively low: only 3 mg Si/l, in space there are about 30,000 hydrogen atoms per one silicon atom. Silicon is a basic component of the most rocks forming the earth's crust – for example sandstone rocks, clays, granites and especially aluminosilicate rocks based on orthoclase or plagioclase. Silicon has the diamond cubic crystal structure with a lattice parameter of 0.543 nm. The nearest neighbor distance is 0.235 nm. Diamond lattice structure - two Face centered cubic (FCC) structures one shifted by a 1/4 of body diagonal. Si atoms are located in all lattice points. Zink Blende structure (ZnS) - Diamond lattice structure for compounds, where elements occupied different FCC structure.

Why silicon won the competition for the best material for the electronics industry? Highly abundant element, high electron mobility, relatively large forbidden gap width, easy to prepare p and n type materials and p-n junctions, existence of the semi-insulating native oxide.

3.1.1. Production of the metallurgical grade silicon

Silicon properties	***Value***
Atomic number	*14*
Atomic weight	*28.0855* $g \cdot mol^{-1}$
Crystal structure	*Diamond*
Lattice constant	*5.43 Å*
Density	*2,328* $g.cm^{-3}$
Atomic density	$5x10^{22}$ *atoms/*cm^3
Hardness	*6.5*
Electronegativity	*1.90 (Pauling scale)*
Thermal conductivity	*149* $W \cdot m^{-1} \cdot K^{-1}$ *(300K)*
Melting temperature	*1415°C*
Forbidden band width	*1.12 eV (300K)*
Electron mobility	*1500* cm^2/Vs *(300K)*
Hole mobility	*600* cm^2/Vs *(300K)*
Relative permittivity	*11.9*



Arc furnace

The greatest amount of silicon worldwide is consumed in the metallurgic industry, which is followed by solar cells, and only a small part of the production is used in the semiconductor industry. This is the reason why basic silicon is called metallurgical grade silicon (**metallurgical grade Si** or ***MG-Si*** for short***)***. In 2020, global world production reached 8 million metric tons per year, 5.4 million metric tons in China, 0.54 million metric tons in Russia, 0.34 million metric tons in Brazil and Norway. The raw material for the production of metallurgical silicon is mineral silica (a silicon dioxide) with very high purity in the form of silica sand, which can be mined only in Brazil, China and a few other countries. Industrially, *MG-Si* silicon is produced by reduction (deoxidation) of the silica melt in an electric arc furnace on the graphite electrode, where the sand and carbon in the form of coal is continuously added. Electric current of about 10,000 A passes through the electrodes. The supplied carbon and electrodes are burned to carbon monoxide gas according to the following reaction:

$$SiO_2 + 2C \rightarrow Si + 2CO.$$

The reduction process takes place at the temperatures of 1500-2000°C and metallurgical silicon is formed in the purity of about 97% – 99% - so called 2N purity (one foreign atom per 99 Si atoms). For the purposes of electrotechnical industry, however, this purity is entirely insufficient, as even a very low concentration of foreign impurities in silicon significantly affects the quality of the produced electronic components. High purity semiconductor-grade silicon has to be produced by purification of the MG-Si.

3.1.2. Production of polycrystalline semiconductor-grade silicon

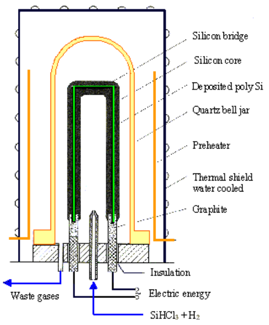
Zonal melting is one of the oldest methods used in the production of high-purity silicon. The material is first adjusted into the shape of a long thin rod, which is then gradually remelted so that the molten zone moves from one end to the other. In this process, the impurities present in the material are concentrated in the molten zone and they gradually move to the end of the rod, which is finally removed by cutting. By repeating this process several times material of high purity is created.

Currently, chemical methods are used to prepare extremely pure semiconductor-grade polycrystalline silicon (SG-Si). In the so-called Siemens method, invented in the early 1960s, the powdered MG-Si reacts with hydrogen chloride gas (HCl) at temperature 300°C to form trichlorsilane ($SiHCl_3$), a highly volatile liquid, as intermediate product. Considerable part of impurities in MG-Si are removed by forming chlorides (BCl_3, $FeCl_3$ ..) and high purity $SiHCl_3$ is obtained. For further purification, rectification (multiple distillation) of $SiHCl_3$ is performed (boiling point of $SiHCl_3$ is 31.8°C). The purified $SiHCl_3$ reacts with high-purity hydrogen gas and decomposes at temperature over 1100°C back to Si, which form the polycrystalline silicon in the form of the U-rod (CVD - chemical vapor deposition). The reaction of trichlorosilane is illustrated by the following equations:

$$Si + 3\,HCl \rightarrow SiHCl_3 + H_2$$

$$SiHCl_3 + H_2 \rightarrow Si + 3\,HCl$$

This process finally produces extremely purified polycrystalline Si with the purity so called 10N-11N, which fully meets the requirements for the production of semiconductor components. In such purified silicon the number of impurities is about $5x10^{12}$ in $1cm^3$. For comparison, the number of Si atoms in the Si crystal is $5x10^{22}$ in $1cm^3$.

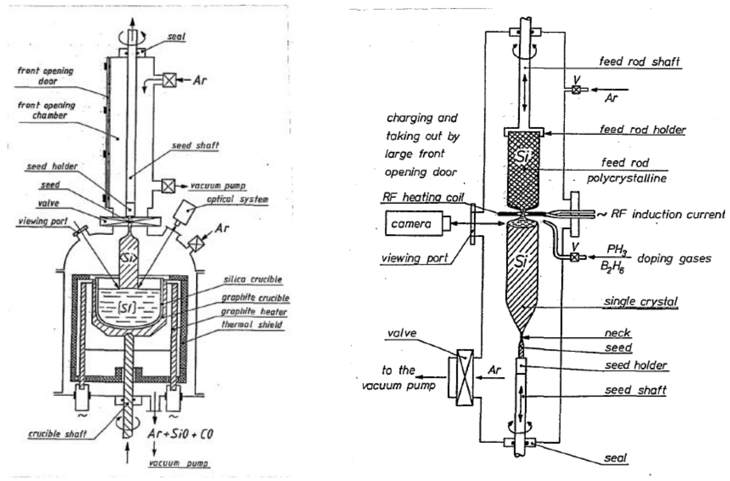


Poly-Si production

Industrial polysilicon plants in China

3.1.3. Production of Si single crystals

3.1.3.1 Czochralski method, CZ



Czochralski method, FZ method

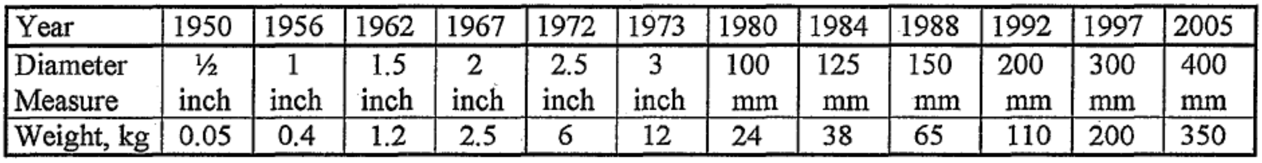
Polycrystalline silicon is unusable for the production of semiconductor devices due to the disorder of the crystal lattice (low carrier mobility). Therefore, single crystalline silicon has to be produced from polycrystalline one, which has a regularly arranged lattice over a long distance. The dominant method for the production of Si single crystals is controlled crystallization from the melt using the Czochralski method. Pure Si polycrystal is inserted into a quartz crucible, which is placed in a protective graphite crucible. The polycrystal is melted in a crucible, and then a high-purity oriented single crystalline silicon seed is introduced into the silicon melt. Both the seed crystal and the crucible rotate in the opposite direction according to a predefined program, while the melt temperature is also very carefully monitored and controlled. Then, pulling of the seed crystal from the melt starts, and new layers of extraordinarily pure monocrystalline silicon begin to grow at the end of the seed crystal. The whole process takes place in a crucible made form a very pure quartz in inert argon atmosphere of the purity of 6N. The melted Si also reacts with the quartz crucible according to the reaction

$$Si + SiO_2 \xrightarrow{T > 1420\ °C} 2\,SiO$$

Silicon monoxide is a highly volatile substance. About 80% of SiO evaporates and is carried away into waste by the argon stream. The rest, i.e., about 20% of SiO remains in the Si melt and passes into the growing single crystal. In it, the oxygen from SiO deposits between Si atoms into an interstitial position. Hence the name interstitial oxygen with the designation O_i. The content of O_i is usually 3×10^{-3} %.

The final silicon ingot, which is made up by only one single crystal, may reaches up to 400 mm (16") in diameter and the length of 2 m. After cooling, the produced ingot is cut into thin wafers (typically 0.5 mm), polished and used as a substrate for the production of semiconductor devices.

Tab.1 Development of the size of Si ingots produced by the CZ method.



Year	1950	1956	1962	1967	1972	1973	1980	1984	1988	1992	1997	2005
Diameter Measure	½ inch	1 inch	1.5 inch	2 inch	2.5 inch	3 inch	100 mm	125 mm	150 mm	200 mm	300 mm	400 mm
Weight, kg	0.05	0.4	1.2	2.5	6	12	24	38	65	110	200	350

3.1.3.2 Floating zone method (Pfann method, FZ,)

The method was developed for the purification of polycrystalline Si, which uses a small segregation coefficient of many impurities. The purified material is first shaped into a long thin rod, which is then gradually remelted in a special oven so that the molten zone moves from one end to the other. In this process, the impurities presented in the starting material are concentrated in the molten zone and gradually reach the end of the rod, which is finally removed by cutting off. Highly pure material is created by multiple repetition of this process. If we choice a suitable speed of rotation and movement of the melted zone, a single crystalline ingot is formed at the same time. As no crucible is used in the FZ method, the grown crystal is not in contact with anything, it is only in a vacuum or an inert atmosphere. Therefore there are very few impurities in the grown single crystal compared to the CZ method, where oxygen and carbon are dominant impurities. The main problem of the FZ method is to maintain the largest possible molten zone. At present, crystals with a diameter of up to 150 mm, very rarely up to 200 mm, are produced by this method. FZ crystals are therefore used only in applications where a very low concentration of impurities is required (PV photovoltaics, radiation detectors and others).

Only Si single crystals grown by Czochralski method are suitable for the production of integrated circuits, because:

- They contains a certain amount of dissolved oxygen (about 0.003% atomic) because of the contact of the Si melt (> 1420°C) with the crucible material (SiO2). This so-called interstitial oxygen in silicon strengthens the Si crystal lattice, which becomes more resistant to the thermal shocks during high-temperature operations of Si production. Interstitial oxygen is also used to get unwanted heavy metals in Si single crystals.
- Si single crystals with a diameter greater than 150 mm (6") cannot be grown by the FZ method.

For PV applications, there are two major advantages of FZ-Si.
1) The long lifetime of the carriers and better microcrystalline quality. These parameters increase the efficiency of solar cells by 10-20%.
2) Lower cost of production (no crucible, faster growth).

The main disadvantage is the necessity of the high quality homogeneous seed without cracks, which is one of the main financial costs. In Si PV technology, approximately 30% of the cost is crystal production, 20% is cutting, 20% is cell production and 30% is module production. FZ-Si is therefore used for the production of high-quality solar cells and CZ-Si for the production of large-area cells of average quality. High-speed multi-wire saws are used in cutting, which are able to cut around 500 wafers / hour opposite to the saw with an internal cut (25 wafers / hour). They also make it possible to cut thinner wafers with a less damaged surface.

Parameters of the CZ a FZ growth methods for Si single crystal production

Characteristic	CZ	FZ
Growth Speed (mm/min)	1 to 2	3 to 5
Dislocation-Free?	Yes	Yes
Crucible?	Yes	No
Consumable Material Cost	High	Low
Heat-Up/Cool-Down Times	Long	Short
Axial Resistivity Uniformity	Poor	Good
Electrical power requirements	60kWh/kg	30 kWh/kg
Oxygen Content (atoms/cm^3)	$>1\times10^{18}$	$<1\times10^{16}$
Carbon Content (atoms/cm^3)	$>1\times10^{17}$	$<1\times10^{16}$
Metallic Impurity Content	Higher	Lower
Bulk Minority Charge Carrier Lifetime (□s)	5-100	1,000-20,000
Mechanical Strengthening	10^{17} Oxygen	10^{15} Nitrogen
Production Diameter (mm)	150-300 (400)	100-150 (200)
Operator Skill	Less	More
Polycrystalline Si Feed Form	Any	Crack-free rod

3.1.4. Si wafer processing

The wafer processing after growth of the Si single crystal has roughly the next steps:

Step number	**Operation**	**Purpose of the operation**
STEP 1	Ingot CUTTING	Shoulder and tail parts are removed using a sawing machine
STEP 2	Ingot GRINDING	Achieving the exact the same diameter of the whole ingot
STEP 3	Facets GRINDING	1 to 4 longitudinal facets for the purpose of crystallographic identification (X-ray diffraction) and production of IC.
STEP 4	Ingot SLICING into slices	Making of a raw Si slices using saw with inner diameter blade or multi-wire saw.
STEP 5	LAPPING of Si wafers	Lapping of Si wafers on both sides to remove the surface layer damaged while cutting.
STEP 6	LAPPING of Si wafer edges	Rounding of wafer edges to remove places on the circumference of Si wafers prone to chipping
STEP 7	ETCHING of Si wafers	Removal of damaged surface and impurities from the surface of Si wafers

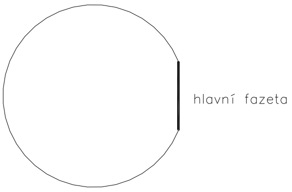
		by etching in HNO_3 + HF
STEP 8	POLISHING of Si wafers	One-sided polishing to mirror shine
STEP 9	WASHING of Si wafers	Removal of remain parts of organic substances, dust particles and heavy metals from the surface in alkali-acid baths.

STEP 2 – Grinding of Si ingot to the desired diameter

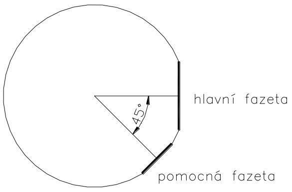
The ingot of Si single crystal never has an exact diameter throughout its whole length. This is why it has to be grinded to a desired diameter and the surface damage from the grinding operation has to be removed by etching. The original ingot diameter therefore exceeds the desired diameter by a few mm.

STEP 3 – Grinding of facets on the Si ingot

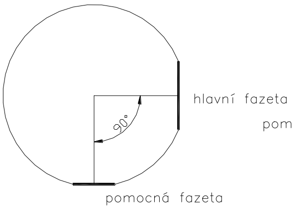
1 to 4 longitudinal facets are grinded on the surface of the ingot. These facets serve to the purpose of the orientation of Si wafers during positioning in technical devices (the main facet) and for the identification of the crystallographic orientation and conductivity type of Si wafers (the ancillary facet). The facets look on Si wafers (example of 1 to 2 facets) are shown in following figure. The main facet is always longer than the ancillary one. Also the 3rd or 4th facet may be present on Si wafers – depending on customers' requirements.



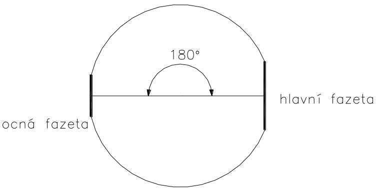
(111) - P typ



(111) - N typ



(100) - P typ



(100) - N typ

STEP 4 – Slicing of facetted Si ingot

The slicing is done using a diamond saw with the inner diameter blade with the rotation velocity of about 1 mm/s. The slice with the diameter of 100 mm is cut in about 100 s. Recently, the slicing has been done with a multi-wire saw.

STEP 5 – Lapping of Si slices

The surface of Si slices is very rough and damaged from the slicing operation. This is the reason why slices are grinded on both sides. The purpose is to achieve a finer surface, flatness of the wafer, the best possible planarity of both sides and desired thickness.

STEP 6 – Lapping the edges of Si slices

The edges of Si slices are rounded to remove the sharp edges as places prone to chipping in IC production.

STEP 7 – Cleaning of the Si surface by etching

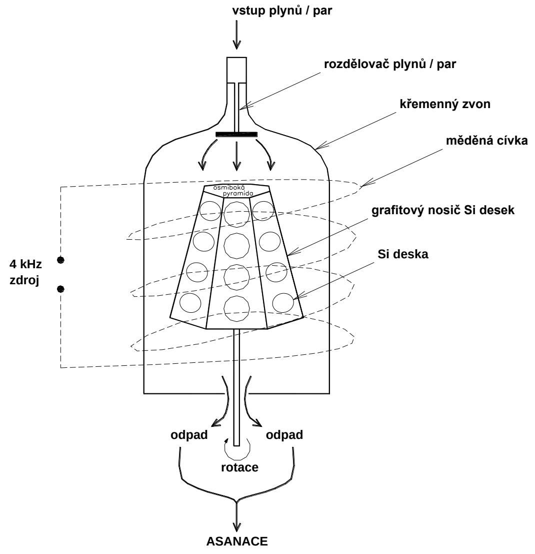
Si slice defects and impurities on the surface are removed by etching in a mixture of nitric acid (HNO_3) and hydrofluoric acid (HF).

STEP 8 – Polishing of Si slices
Polishing is done only on 1 side of the Si wafer. The goal is to achieve a mirror shine on the surface, without scratches and other surface defects. The procedure uses a polishing paste on the basis of a very fine aluminum oxide. The result is a silicon wafer.

STEP 9– Final washing of Si wafers
The goal is to remove the remains of organic substances, dust particles and h heavy metals from the surface of the entire Si wafer. The operation is performed in baths based on

a) hydrogen peroxide H_20_2 and hydrochloric acid HCl
b) hydrogen peroxide H_20_2 and ammonium hydroxide NH_4OH

3.1.5. Production of the Si thin layers (**Chemical Vapor Deposition-CVD**)



Most of the integration circuits, i.e. transistors, diodes, resistors, capacitors, etc. are fabricated in the thin epitaxial Si layers produced on Si wafers. The silicon layer is most often formed by decomposition of silicon tetrachloride $SiCl_4$ in pure hydrogen gas at a temperature of about 1200 °C using phosphine PH_3 as a dopant:

$$SiCl_{4(g)} + 2H_{2(g)} \leftrightarrow Si_{(s)} + 4HCl_{(g)}$$

$$2PH_{3(g)} \leftrightarrow 2P_{(s)} + 3H_{2\,(g)}$$

The heart of the epitaxial reactor is the graphite holder of Si wafers in the shape of a polygonal pyramid, which is placed in a quartz bell. The holder is made from very pure graphite (99.9995%), covered with a 100 µm thick protective layer of silicon carbide SiC (99.9999%). A copper coil located around the bell mediates the heating of graphite holder and thereby also Si wafers by induction currents with a frequency of 4000 Hz. The reacting substances i.e. H_2, H_2 + HCl, $H_2 + SiCl_4 + PH_3$, H_2 flow separately into the bell from above with help of the very pure carrier gas H_2 and finally inert N_2 gas for cleaning.

Unreacted waste gases and vapors are replaced from hydrogen by spraying with water, which removes > 99.9% of $SiCl_4$ and HCl. The growth rate of the layers depends on the ratio of the two dominant source gases, $SiCl_4$ and HCl. Growth rates above 2 micrometers per minute produce polycrystalline silicon. A negative growth rate (etching) can occur if too much HCl is present as a second product of the reaction in the reactor. Sometimes HCl is added intentionally to etch the material.

Vapor phase epitaxy of silicon can also use silane SiH_4, dichlorosilane SiH_2Cl_2 or trichlorosilane $SiHCl_3$ as source gases. For example, the reaction with silane proceeds at temperature of 650 ° C as follows:

$$SiH_4 \rightarrow Si + 2H_2$$

This reaction does not etch the material and occurs at lower temperatures than the deposition from silicon tetrachloride. However, if the reaction is not strictly monitored, a polycrystalline layer will form and the oxidation of reaction products that can contaminate the epitaxial layer with undesired compounds such as silica occurs.

3.1.6. Si wafers oxidation

Layer of amorphous silica (SiO_2) is commonly used as a separating insulating layer in the manufacture of integrated circuits and also as a protective layer in ion implantation, thermal diffusion and surface passivation. SiO_2 layer is created on the surface of Si wafers by thermal oxidation. This layer is formed both on the front mirror-like side and on the back matt side of the Si wafer, respectively. The oxidation process takes place at high temperatures (900 °C – 1200 °C) and can be very well controlled. Thickness of SiO_2 layers may range between 0.01 µm to 1 µm. The SiO_2 layer has a number of unique properties that allow for production of large numbers of IC chips on the Si wafer at the same time. One of the important parameters is the layer's good stability against various acids and bases. SiO_2 dissolves only in the solutions of hydrofluoric acid (HF). Therefore, HF is used for etching of SiO_2 layers on Si wafers. Si wafer itself is inert against diluted HF solutions. The properties of the Si-SiO_2 interface is among the best described and measured among solid substance in general.

Tab.1. Some properties of SiO_2 in comparison with pure Si

Characteristic	SiO_2	Si
Specific weight	2,25 g/cm^3	2,33 g/cm^3
Resistivity	> 10^{15} Ωcm	2,3 × 10^5 Ωcm
Melting point	~ 1700 °C	1420 °C

Layers of amorphous SiO_2 on Si wafers are used for the following purposes:
- as a protection mask against the ion implantation and diffusion of elements into Si
- for passivation of Si surface (protection of Si surface from environmental influences)

In principle, oxidation of Si wafers occurs in 2 ways:
a) oxidation by oxygen

$$Si + O_2 \longrightarrow SiO_2$$

b) oxidation using water steam

$$Si + 2H_2O \longrightarrow SiO_2 + 2H_2$$

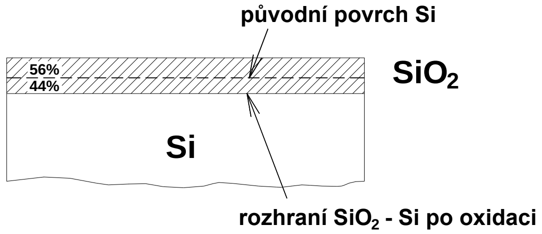
Water steam is obtained by burning of H_2 in O_2 before entering the oxidation furnace.

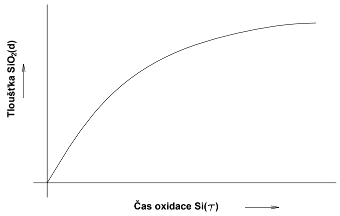
Reaction a) is relatively slow. It is used mainly for preparation of very thin and precise SiO_2 layers.

Reaction b) is several times faster than reaction a). It is used mainly for the preparation of thick SiO_2 layers.

Regardless the specific reaction type, both types of oxidation consume Si from the surface of the Si wafer. Let say, from a layer of SiO_2 with 0.1 µm thickness, 44% of the Si wafer thickness is consumed; i.e., 0.044 µm. The thickness of SiO_2 increases with the root of time (slows down gradually).

It may be illustrated as follows:





The processes of preparation of SiO_2 layers on Si wafers take place in the furnace with the tubes from very pure fused quartz, which chemically is pure SiO_2. If oxidation is combined with diffusion, tubes from SiC are used. Both types of tubes are heated externally by resistance heating made from kanthal wires. Kanthal is an alloy of Fe – Cr – Al – Co (iron – chrome – aluminum – cobalt). Kanthal heating wires are wound on a ceramic tube made from aluminum oxide. Heating wires, however, are also a source of contamination by heavy metals; the ceramic unit is a source of Na.

3.1.7. Doping

3.1.7.1. Doping during the growth

Silicone belongs to the 4th group of the periodic table. For the preparation of Si with a defined specific electrical resistivity and type of conductivity (n or p-type), doping of Si by elements of the 3rd and 5th group is used. These elements are characterized by the fact that they significantly affect the electrical conductivity of Si even in very small quantities. Let's imagine a hypothetical situation when the arrangement of Si atoms in the single crystal is absolutely perfect without any crystallographic defects, and the single crystal contains no foreign atoms (impurities). Then, at the temperature of 20°C, the electrical resistivity of Si will be $\rho = 2.3 \times 10^5\ \Omega$cm. At the same temperature, ρ of metals (conductors) is $> 10^{-6}\ \Omega$cm, and ρ of insulators $> 10^{15}\ \Omega$cm. For the production of semiconductor devices, however, it is necessary to obtain Si with specific electrical resistance in the range of $10^2\ \Omega$cm to $10^{-3}\ \Omega$cm. If we, for example, want to obtain Si with $\rho = 1\ \Omega$cm, we have to insert doping atoms (dopants) into Si with the volume concentration in Si of 10^{-5} %. This means that in every 10 million atoms, there will be 1 doping atom. This single atom in 10 million atoms of Si may decrease electric resistivity by more than 100 000 times! This is why polycrystalline and monocrystalline Si has to be so pure. In addition, the dopant in Si not only decreases the electrical resistivity of Si but also change the temperature dependence of the resistivity. In ultrapure Si, electrical resistivity drops with increasing temperature. In doped Si, on the other hand, electrical resistivity increases with temperature due to decreasing of the mobility. The lower doping concentration, the steeper the change.

High-resistivity Si with a low concentration of electrically active impurities may only be produced using the FZ method. In the Czochralski method, oxygen originating from the crucible is built into the lattice and behaves like a donor. According to customers' requirements, crystals may be intentionally doped by an electrically active impurities (dopants). Arsenic (As), antimony (Sb) and phosphorus (P) are used for donor n-type doping. Boron (B) or gallium (Ga) is a typical acceptor p-type dopants. Dopant atoms are added to the melted silicon before the crystal is pulled (CZ) or directly blown into the molten zone (FZ).

N-type (electron conductivity) : arsenic (As), phosphorus (P) or antimony (Sb)
P-type (hole conductivity) : boron (B)

For the production of bipolar integrated circuits, P-type Si wafers with the electrical resistivity of 7-12 Ωcm are used. This value of resistivity is equivalent to a boron concentration of 4×10^{-6} %. For each 1 billion Si atoms, there is therefore only 40 atoms of boron.

3.1.7.2. Post grown doping - Ion implantation

Ion implantation is a very precise and reproducible process of incorporation impurities in the form of elements at shallow depths of Si wafers. It is a process when electrically charged atoms (B^+, P^+ As^+, Sb^+) are being accelerated in applied electric field toward the Si wafer. The energy of the positive ions is very high, typically in the range of 40 keV – 180 keV. 1 eV (electron volt) is the energy that an electron obtains when passing through a voltage drop of 1 volt. When the voltage drops is 100 kV in ion implantation device, the electron energy will be 100 keV. At this energy, the electron has the speed of 187 000 km/s. At the voltage drop of 3 MV, the electron's speed would be close to the speed of light in a vacuum, i.e. to 300 000 km/s. If we, for example, apply a 100 kV voltage drop on As^+ ions, its speed will be 500 km/s because the speed of the charged particles in the electric field is in an inverse proportion to the square root of their weight.

As mentioned above, B^+, P^+ As^+ and Sb^+ ions have high energies when being shot into Si wafers, and thereby also high speeds – hundreds of km/s. When these ions hit the surface of Si, they penetrate its surface layers. It is obvious that the depth of penetration of implanted ions will be strongly dependent on their energy and the type of ions. The same applies to the penetration of implanted ions into the masking layer of SiO_2.

3.1.7.3. Post grown doping - Diffusion of dopants at high temperature

In the production of the integrated circuit chips, we need to get electrically active dopants (B, P, As, Sb) to a certain and precise depth at the selected parts of the Si wafer with a specific profile of the doping concentration. We use high temperature diffusion of dopants into Si wafers at temperatures range between 1100°C to 1200°C. For incorporation of impurities into the Si wafer, we need high temperatures, as the diffusion of impurities in Si at the room temperature is extremely slow.

The diffusion length of impurities in Si depends on:
- Diffusion coefficient of the impurity
- Temperature of the Si wafer
- Time of the diffusion
- O_2 in the diffusion atmosphere ($O_2 + N_2$)

An amorphous SiO_2 layer application on the surface of Si wafers to protect the diffusion of impurities into Si wafers is a key attribute of Si IC production technology. The speed of the diffusion of B, P, As and Sb impurities in SiO_2 is 10 to 10,000 × lower than in the case of Si.

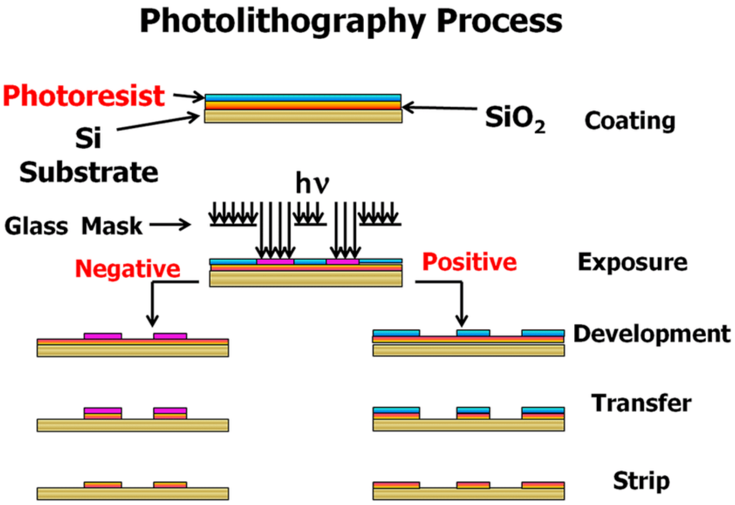
Essentially, dopants are added into Si wafers in 2 forms:
- as elements (B, P, As, Sb) or
- as oxides (B_2O_3, P_2O_5)

While elements as B, P, As, and Sb may be directly diffused into Si after being applied to an Si wafer, dopants in the form of oxides such as B_2O_3 (boron trioxide) or P_2O_5 (phosphorus pentoxide) are first need to reduce to the respective elements by silicone:

3.1.8. Photolithography

Photolithography, also called optical lithography, is a process used to create a small pattern on a thin film or the bulk substrate (wafer). This method can create extremely small patterns, down to a few tens of nanometers in size. Photolithography shares some fundamental principles with photography. Photolithography uses a photosensitive chemical photoresist called photoresist and the photomask (optical mask) with a geometric pattern. First step is the application of the photoresist layer on the surface by spin coating process. After layer drying we will align the photomask to the layer and expose photoresist through the photomask by visible or UV light. The pattern on the mask

protect underneath photoresist to be expose by the light. We have two types of photoresist - Positive and Negative. In case of positive type, exposed part of the photoresist decomposed and it can be dissolved in the developer. In case of negative photoresist the situation is opposite, exposed part of the photoresist is protected for dissolving in developer. Exposured pattern in the photoresist is used for selective etching or metallization using lift-off technique. In complex integrated circuits, a CMOS wafer may go through the photolithographic cycle as many as 50 times.
We have also electron lithography, where electron beam created by electron gun is focused directly to the photoresist and irradiate the surface based on the scheme.
Using photolithography we usually etched the protected layer of SiO_2 prepared on silicon wafer. Prepared pattern is used for selective surface etching, diffusion or ion implantation.



3.1.9. Porous silicon – por-Si

Porous silicon (**por-Si**) is a nanocrystalline material, which is composed of thin silicone columns with the diameter of less than 3 nm, which have a crystalline structure and are surrounded by a layer of SiO_2. It was first created in 1956 in Bell's laboratories in the USA during research into electric polishing of Si. However, it was not before 1990 when Leigh Canham found out that when exposed to radiation, por-Si effectively emits light in the visible spectrum, unlike the common crystalline silicone, which emits no light at all. Unfortunately, its electroluminescence is very weak. The main advantage of por-Si is its relatively easy production; its disadvantages are, on the other hand, low chemical stability, fast photo-degradation, large dispersion of diameter of nanocrystals and poor chemical controllability.

Porous Si is currently still in the research stage. Its application in the area of sensors seems the most promising, as por-Si has, as every nanocrystalline material, large surface area and is therefore sensitive to the presence of various substances in its environment. The perspective application of por-Si is combination of Si substrate with porous surface for epitaxial growth of materials with different lattice constants. Porous surface is very elastic and accommodates to the epitaxial layer, where no dislocations occur. Other potential use is the gradual release of medical drugs into a patient's body. Human bodies are able to effectively metabolize nanocrystalline silicone. If the pores of por-Si are filled with a medical drug, the drug is released into the body by gradual degradation of nanocrystals, and it is therefore present in the body in small concentration over a long time. Porous silicone is a material emitting light, and there are ongoing efforts to create, e.g., LED diodes using it, which are as of now still manufactured using significantly more expensive materials. Silicone LED diodes could also be easily incorporated into the current microelectronics which is entirely based on silicone.

Production of porous Si

Porous Si is prepared by electrochemical etching of Si in a HF (hydrofluoric acid) solution. Upon passing the interface Si-electrolyte, electric current changes its nature from an electron to the ionic one. In the connection, where "holes are directed into the electrolyte", decomposition of Si takes place. At a low current density, a system of deep pores emerges in silicone, which penetrates deep into the material. The creation of pores may be described by the following equation:

$$\mathrm{Si} + 6\ \mathrm{HF} \rightarrow \mathrm{H_2SiF_6} + \mathrm{H_2} + 2\ \mathrm{H+} + 2\ \mathrm{e-}$$

If high current density is applied, electrical polishing occurs, and new flat smooth surface of Si is created. This mode is used to create the so-called *self-bearing layers* of porous Si, where short-term increase of potential at the end of etching leads to a separation of porous Si from the substrate. Porosity (P) is the main parameter of porous Si, which can be established, e.g., by the weight method from the weight of the silicon wafer before etching m_1, weight right after etching m_2 and weight m_3 after the full decomposition of the porous Si layer (e.g. in KOH or NaOH):

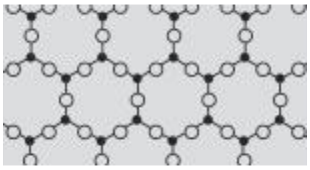
$$P = (m_1 - m_2) / (m_1 - m_3)$$

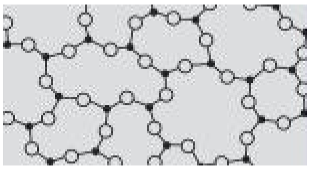
Porosity can be effectively controlled by changing the production parameters. The most important parameters are:

- current density
- electrolyte chemical composition
- etching time
- time and way of oxidation

3.1.10. Quartz glass – production

Quartz (SiO_2) is a very stable crystalline material, which is very widespread in the Earth's crust. Quartz glass (commonly only quartz) contains pure SiO_2 in an amorphous phase. It is producing by melting of highly pure quartz sands in the electric arc furnace at the temperature above 1700°C. After the cooling down of the melt, an amorphous modification of quartz is created – quartz glass. Quartz glass has the highest softening point (1200°C) of all glasses and has a considerably higher melting point (1610°C). General glass also contains SiO_2 but different additives such as soda (sodium carbonate Na_2CO_3) or potassium carbonate (K_2CO_3), which form eutectic mixture with quartz glass are added for lowering of the melting point down to the temperatures around 600°C. Quartz glass is usually produced in clear form, but also exists in matt form or with different colouring (optical filters).





Crystalline SiO_2 Quartz glass – amorphous SiO_2

Quartz glass has some typical chemical properties:

- It is made from pure quartz – high content of SiO_2 > 99.5 %

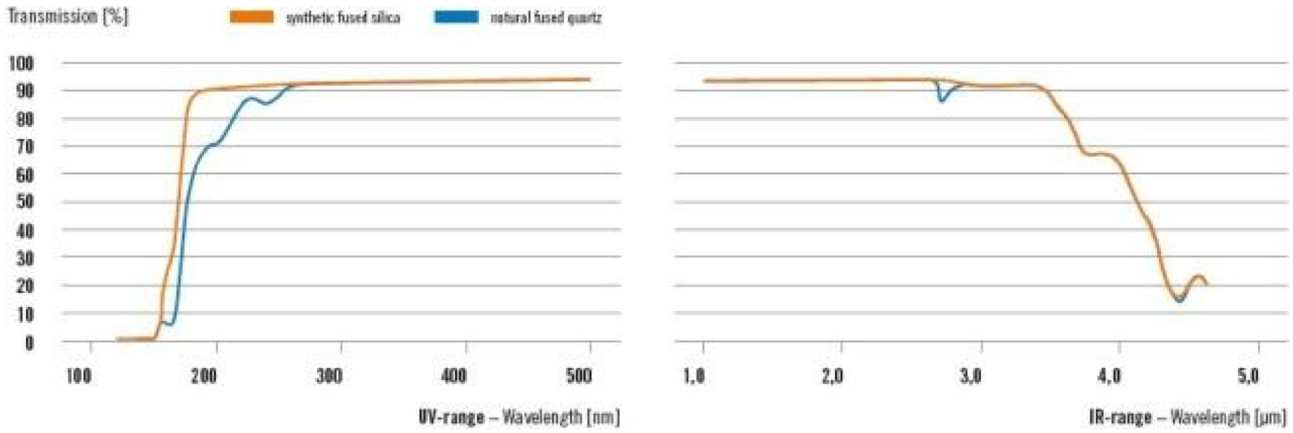
- Very low coefficient of thermal expansion: 20°C - 1000°C = 5.5 x 10^{-7} m/ K
- Extreme resistance against thermal shock and ability to sustain temperatures of up to 1300°C.
- High thermal conductivity: 1.46 W/m K
- Specific weight: 2.2 x 10^3 kg/m^3
- Dielectric constant $\varepsilon = 38$
- Resistant against most acids and bases, (suitability for chemical industry).
- Refractive index of quartz glass is 1.544. Different additives usually increase the refractive index of quartz glasses, which then usually ranges between 1.5 and 2.0.

Quartz glass transmits a broader spectrum of radiation than additive-based glasses, especially on the ultraviolet side of the spectrum (< 350 nm). In additive-based glasses, transmittance ends around the wavelength of 400 nm; therefore, these types of glass are impermeable for UV light. In the infrared part of the spectrum, pure quartz glass has such low attenuation that light is able to pass through it from the distance of tens or hundreds of kilometers. The areas of wavelength transmittance are called quartz windows. Quartz glass has several such windows, with the most important ones being between the wavelengths of 650-750 nm, 850-1000 nm and 1300-1500 nm.

Applications:

Chemical devices and apparatuses – High resistance against most acids and bases and other aggressive chemicals. Semiconductor industry, chemical industry.

UV radiation transmittance of quartz glass is used in the production of bulbs of ultraviolet gas-discharge lamps and tubes and other accessories of UV light sources, but also for the production of greenhouse windows so that the plants in greenhouses are able to accept the whole natural sunlight spectrum. In the area of UV technology, quartz glass plates are used mainly to screen out the space between the UV lamp and the radiated area, serving mainly to separate different air flows or as a protection against pollution.



The low attenuation in infrared parts of the spectrum is used in the production of optical fibers. For this purpose, quartz is highly cleaned by similar processes as those used for the cleaning of semiconductors for the production of semiconductor devices. As, however, ultrapure quartz has a low refraction index and high refraction index, on the other hand, is needed in the cores of optical fibers, ultrapure germanium glass (GeO_2) is usually added to ultrapure quartz glasses, which moves the refraction index towards higher values. Such optical fibers have the lowest attenuation and dispersion for the wavelength window between 1.3µm and 1.55µm and are optimal for production of optical fibers for communication.

3.2. Germanium - Ge

Germanium was predicted by Newlands in 1864 as the missing element in the triad between silicon and tin. Mendeleev predicted its properties in 1871 based on its position in the periodic table and called it “ekasilicia”. In 1886, Winkler isolated germanium from the mineral argyrodite and named it to honor his homeland, Germany (Germania in Latin). Inorganic germanium is a necessary element, similarly to iron or calcium in nature. It is contained in small amount in many plants and in Earth’s crust, where it is a rather rare element. The average content only amounts to 5 - 7 ppm (mg/kg). In seawater, the concentration of germanium is extraordinarily low, only 0.07 micrograms/l. It is estimated that in the universe, there are 30 million atoms of hydrogen for one atom of germanium. In minerals, germanium is present only as an additive in zinc or silver ores, but it is also contained as a trace additive in numerous coal bearing. Germanium’s properties are important for the electrotechnical industry. In 1948, the first transistor and integrated circuit were made of germanium, and later entire computer chip. The current studies showed that germanium is important for the human organism. Doctors and researchers are trying to verify that germanium has similar effects as selenium and may be of significance as a nutritional supplement.

3.2.1. Physical properties of germanium

Germanium is a fragile, dark grey substance crystalizing in diamond structure like silicon (FCC double structure). In the solid state, Ge behaves as a semiconductor, both in the crystalline and amorphous phase. In the liquid state, on the other hand, germanium is a metal. Ge is more reactive than Si. It creates compounds in the following valency: Ge^{-4}, Ge^{+2} a Ge^{+4}. At room temperature, it easily reacts with chlorine and fluor, with sulphur vapors creates GeS. After slight heating in Cl_2 and Br_2, germanium tetrachloride ($GeCl_4$) and germanium tetrabromide ($GeBr_4$) emerge. Germanium does not create carbides, which is why it is able to melt in graphite crucibles. If no oxidation agent is present, it does not react with water or diluted acids and alkali hydroxides. When exposed to air, it oxidizes into germanium dioxide (GeO_2) with the emergence of red heat. Ge dissolves in hot concentrated sulfuric acid and nitric acid:

$Ge + 2H_2SO_4 \rightarrow GeO_2 + 2SO_2 + 2H_2O$
$Ge + 4HNO_3 \rightarrow GeO_2 + 4NO_2 + 2H_2O$

3.2.2. Production of germanium

The only industrially exploited germanium bearing is in Congo (Kinshasa), which is among the world's richest sources of germanium. However, the main sources for Ge production are ashes emerging in metallurgical processing of zinc ores in the zinc production and ashes after combustion of certain types of coal. For instance, coal from the Pilsen basin contain 20 to 70 g of Ge per 1 ton. The ashes are dissolve din sulphuric acid, and from the obtained solutions, it coagulates by zinc dust with the simultaneous effect of aqueous solution of sodium hydroxide (NaOH). The coagulate is again dissolved in sulphuric acid, and the process is repeated. The second coagulate contains about 7% Ge in the form of GeO_2. From this concentrate, germanium in the form of volatile germanium tetrachloride ($GeCl_4$) is obtained due to the effect of the mixture of hydrochloric acid and chlorine.

$GeO_2 + 4\ HCl \rightarrow GeCl_4 + 2\ H_2O$
$GeO_2 + 2\ Cl_2 \rightarrow GeCl_4 + O_2$

Atomic number	32
Atomic weight	72.63 g·mol^{-1}
Crystal structure	Diamond
Lattice constant	5.65 Å
Density	5.323 g.cm^{-3}
Atomic density	4.4x10^{22} atoms/cm^{3}
Hardness	6.0
Electronegativity	2.0 (Pauling scale)
Heat conductivity	60.2 W·m^{-1}·K^{-1} (300K)
Melting point	937°C
Forbidden band width	0.67 eV (300K)
Electron mobility	3900 cm^{2}/Vs (300K)
Hole mobility	1900 cm^{2}/Vs (300K)
Relative permittivity	16

Pure $GeCl_4$ hydrolyses to $GeO_2.xH_2O$, which after filtration at the temperature of 150 to 200°C gets rid of water and passes into waterless germanium dioxide (GeO_2). Metallurgical Ge is made from GeO_2 by reduction in an electric furnace with the presence of carbon

$GeO_2 + C \rightarrow Ge + CO_2$

Global world production of Ge in 2020 was 130 metric tons per year, China was the world's largest producer, whose refinery production was 86 metric tons.

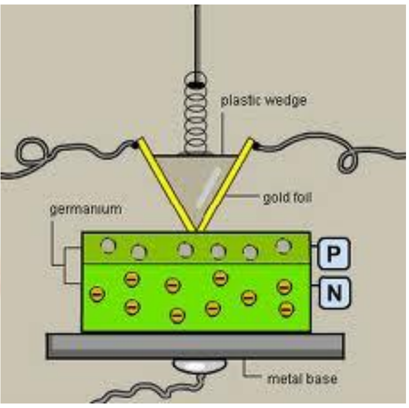
Prepared germanium contains 5-10% impurities and is unsuitable for the purposes of semiconductor production and production of germanium glass. For this reason, Ge is further cleaned using zonal purification. By the repetition of zonal refinement (6 to 10 times), Ge purity reaches extreme values of 12N (for one atom of impurity there are 10^{12} atoms of Ge). This is in present one of the purest industrially produced materials.

Fraction distillation of highly volatile germanium tetrachloride $GeCl_4$ is another method of obtaining highly pure Ge. Before hydrolysis, $GeCl_4$ is distilled, when impurities (such as $AsCl_3$) are removed from the distillate by fraction distillation in the refraction column. After hydrolysis, GeO_2 is reduced by hydrogen in an electric furnace (without any oxygen content) at the temperature of approximately 600°C:

$GeO_2 + 2H_2 \rightarrow Ge + 2H_2O$

3.2.3. Application:

- The interest in germanium started in the 1950s, when the **first transistors** and other electronic components on the basis of pure germanium were invented. During two other decades, germanium was replaced by silicone, which is found in nature in much greater quantity. Germanium is, however, still used for special semiconductor diodes.



The transistor effect was discovered and the transistor was invented on 16 December 1947 in Bell's laboratories. In 1956, .B.Shockley, J.Bardeen and W.H.Brattain were awarded the Nobel prize for the research into semiconductors and the discovery of the transistor effect.

- Currently, germanium is used in industrial production of semiconductors mainly as Si-Ge compound (SiGe) for the production of integrated circuits with high speed of signal transmission. It is also part of circuits, which react to electromagnetic waves in the infrared part of the spectrum. It is thus used in radar technology. Currently, this type of use is decreasing in favor of application in optics, where SiGe started to replace GaAs in devices for wireless communication.

- High purity germanium (HPGe-high purity Ge) is used as a detector of X-ray and gamma-ray radiation.

- Ge has an important application in the production of **light-distributing optics**, as its presence in glass and other materials for optical fibers significantly increases the material's refraction index (similarly also GeO_2). Ge is also transparent for infrared radiation. These properties are useful in the production of special optical components as lenses for wide-angle cameras or optics for the processing of signal in the infrared part of the spectrum (such as in devices for night vision). Due to the high refraction index (4.0), however, all optical devices have to be equipped with anti-reflexive diamond-based layers.

- GeO_2 is of great significance as a catalyst in polymer (plastic) production, but it can also be replaced by titanium.

- Ge epitaxial layers are a part of the heterostructures produced for the most effective solar cells in the world.

- Ge generators change heat into electricity.

- Ge alloys have interesting properties – the alloy with gold (the so-called jeweler's solder) expands when cooling down. The alloy which copper and gold is suitable in dental medicine.

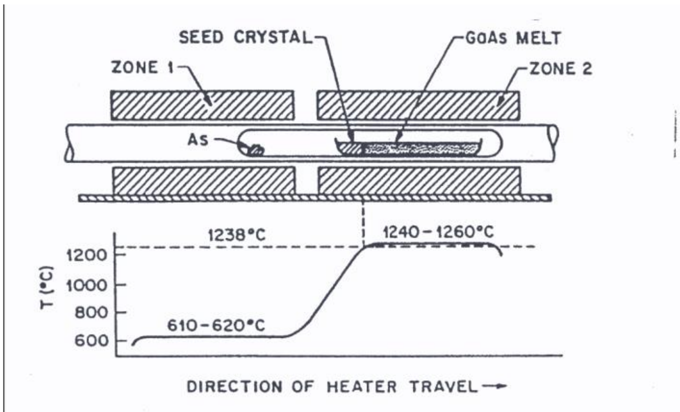
- Gadolinium gallium garnets (GGG) are used in laser technology.

In 2007, 35% of Ge production was used in fiber optics, 30% in infrared optics, 15% was used as polymerization catalysts, and 15% was used for electronic and photovoltaic purposes. The remaining 5% was used in metallurgy and medicine.

3.3. Gallium arsenide – GaAs

GaAs single crystals may be prepared by a direct synthesis from elements through several methods:

1. Liquid Encapsulated Czochralski Method (LEC method)
2. Bridgman method in a horizontal furnace with As additive source (Bridgman-Stockbarger method).



Bridgman technique for gallium arsenide with temperature profile of the furnace.

Alternative methods of GaAs production include:

3. Reaction of gaseous gallium and arsenic trichloride, when chloride decomposes and gallium and arsenic vapors react together and deposit single crystalline GaAs layer on the surface in the cooler part of the furnace.

   $2\,Ga + 2\,AsCl_3 \rightarrow 2\,GaAs + 3\,Cl_2$

4. Reaction of trimethylgallium and arsine (MOCVD –Metalorganic chemical vapor deposition)

   $Ga(CH_3)_3 + AsH_3 \rightarrow GaAs + CH_4$

5. MBE

In comparison with Si, GaAs has a number of unique characteristics. It has a higher saturation electron speed and a higher electron mobility, which makes this material suitable for the production of high-frequency transistors (up to 250 GHz). Thanks to the larger width of the forbidden band, it is less sensitive to the thermal noise. Further, components have lower noise at higher frequencies thanks to greater electron mobility. This is the reason why these components are used mainly in the production of mobile phones, in satellite communication, microwave remote communication and in high-frequency radar systems. Gunn's diodes as microwave generators are produced based on GaAs, as well. Thanks to its greater width of the forbidden band, GaAs is more resistant against radiation damage, which is why it is widely used in satellite and rocket electronics and as optical windows for the high-power applications. The direct forbidden band gap is an advantage for the optoelectronic applications.

Disadvantages compared to Si: Silicon is one of the most abundant elements in the world and its production is very well mastered, and we can grow it on large areas and in great volumes. Silicon is both temperature-wise and mechanically very stable. Si is a basic, simple element, whereas GaAs is a mixed compound. It is possible to grow Si almost without defects and with a low concentration of foreign defects. Si allows extreme integration (currently down to 20 nm). Integration of 500 nm is typical for GaAs. The existence of SiO_2, an excellent insulator, is an extreme advantage of Si. GaAs has no stable oxide, nor does it have another proprietary compound. The third advantage concerns much higher conductivity, which enables the production of fast FET transistors, which are required in CMOS logic structures. CMOS structures based on GaAs have much higher consumption.

Material	**Density (g/cm)**	**Atomic number**	**Forbidden gap width (eV)**	**Lattice constant (A)**	**Electron mobility (cm^2/Vs)**
Si	2.328	14	1.12	5.43	1 500
Ge	5.323	32	0.67	5.658	3 900
GaAs	5.316	31+33	1.45	5.65	8 500
CdTe	5.85	48+52	1.45	6.482	1 000
InP	4.81	49+15	1.344	5.869	54 000
InSb	5.775	49+51	0.17	6.479	77 000
SiC	3.21	14+6	2.3-3.2	4.35	900

3.3.1. GaAs Applications

Solar cells

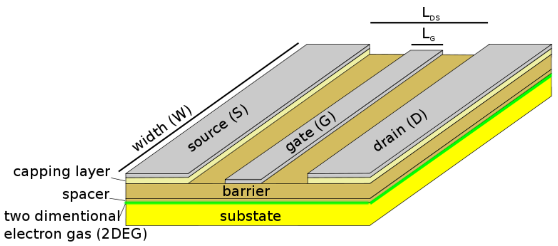
The first solar cells based on GaAs heterostructures were produced by Alferov in 1970 in the Soviet Union. Already in 1980, their efficiency exceeded the efficiency of Si solar cells, and these solar cells are the most effective in the world. Heterostructures on the basis of GaAs, Ge and InGaP reach efficiency of more than 40 % today. They are mostly used in satellite applications.

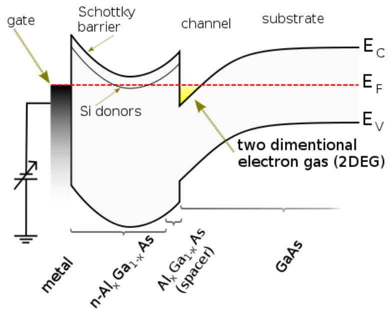
Microprocessors and transistors

The first microprocessor based on GaAs was developed in 1980 as a part of the “Space War” project in the USA. Other microprocessors were used in most powerful CRAY supercomputers.

High electron mobility transistors (HEMT) and other quantum structures

.





Detectors

Highly sensitive detectors for the middle part of the infrared radiation (QWIP-quantum well infrared photodetector) are produced based on $Al_xGa_{1-x}As$ /GaAs heterostructures. GaAs diodes are used for the detection of X-ray radiation.

Light sources

In optoelectronic applications, a direct forbidden band is an advantage. GaAs is used for the production of laser diodes for the area of close infrared and visible radiation. LED and lasers for the visible area are made from the AlxGa1-xAs compound material using the MBE or MOCVD technique (AlAs and GaAs have practically the same lattice constant).

GaMnAs - spintronics

3.4. Cadmium telluride –- CdTe

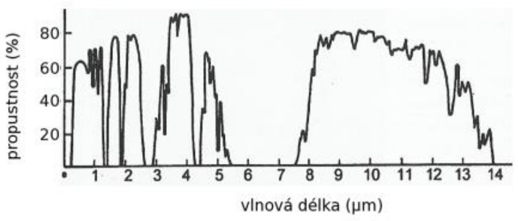
Cadmium telluride single crystals are produced using several methods:

1. - Bridgman method in both horizontal and vertical furnace (the Bridgman-Stockbarger method) both with or without the seed.
2. – Vertical Gradient Freeze Method (VGFM).
3. - Liquid Encapsulated Czochralski Method (LEC method).
4. - Traveling heater Method (THM).
5. - MOCVD
6. – MBE

3.4.1. CdTe Applications

(Mercury cadmium telluride - HgCdTe, Cadmium zink telluride – CdZnTe)

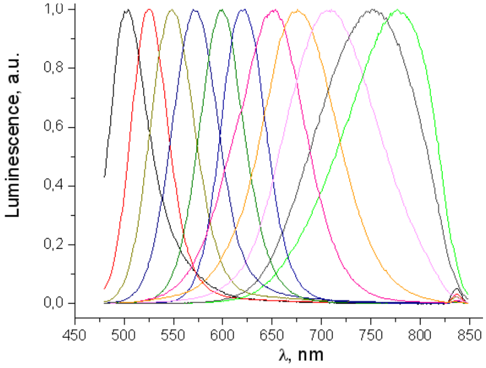
CdTe is an important material for the preparation of detectors of infrared, x-ray and gamma-ray radiation. Thanks to its wide forbidden band width (1.5eV), high atomic number (48, 52), high electron mobility (~1100 cm2/Vs) and relatively long lifetime of free carriers, high-resistance CdTe doped by Cl or In is used for the production of radiation detectors of x-ray and gamma-ray radiation and alpha and beta particles. These detectors work at the room temperature, which allows for the production of compact, portable detector applications. CdTe may be mixed with Zn, and **$Cd_xZn_{1-x}Te$** is created, which is used for the detection of x-ray and gamma radiation and as a substrate for epitaxial growth of narrow bang gap $Hg_xCd_{1-x}Te$. By addition of Hg and precise tuning of its content in CdTe, the compound material **$Hg_xCd_{1-x}Te$** is obtained, which covers three transmission windows in the air for infrared radiation. In the 8-12 um area, it is the most sensitive detector.



Obr. The wavelength interval of the air transmission for infrared radiation

Thanks to the absorption coefficient, CdTe is an effective material for the production of thin-film solar cells, which quality is comparable with thin polycrystalline layers prepared from Si.

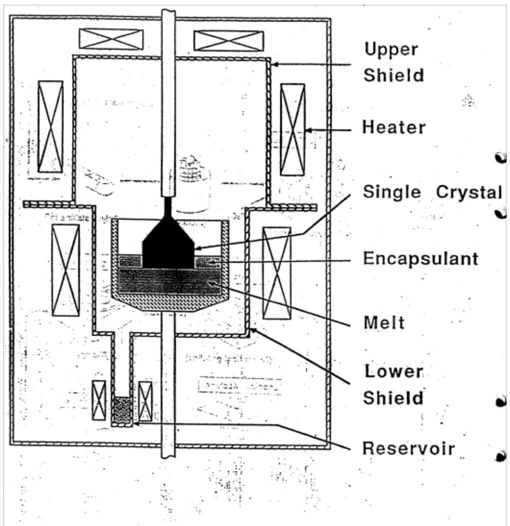
Bulk CdTe is transparent for infrared radiation from 860 nm (which represents the forbidden band energy of 1.44 eV at 300 K) up to the wavelengths greater than 20 µm. CdTe also exhibits fluorescence at the wavelength of 790 nm. When we prepare CdTe nanocrystals, they create quantum dots and their fluorescent peak shifts from the visible to the ultraviolet part of the spectrum depending on the dominant diameter of prepared nanocrystals. CdTe is also used as a material for optical windows and lenses for the area of infrared radiation.



Finally, CdTe is used as an electro optical modulator thanks to its high electro optical linear coefficient ($r41=r52=r63=6.8\times10^{-12}$ m/V).

3.5. Indium phosphide – InP

InP is used in high-performance and high-frequency electronics due to its high electron mobility (compared to Si and GaAs). It is a direct semiconductor with the forbidden band gap of 1.34 eV at 300K. It is used for the production of LED diodes, laser diodes, microwave transistors and Gunn diodes and as a substrate for the epitaxial growth of InGaAs and GaInAsP. These ternary and quaternary compounds are used for the preparation of sources as well as detectors of infrared radiation at wavelengths suitable for the transmission through optical Si fibers (1.3µm a 1.55µm). At these wavelengths, the fibers have the lowest attenuation and dispersion.



Phosphorus Vapor Controlled LEC Furnace

Indium phosphide single crystals are produced by several methods:

1. - Liquid Encapsulated Czochralski Method (LEC method)

2. - Vertical Gradient Freeze Method (VGFM)

3. - Bridgman method in both horizontal and vertical furnaces (Bridgman-Stockbarger method) with and without the seed

4. - Traveling Heater Method (THM)

5. - MOCVD

6. MBE

InP Doping

N-type InP may be prepared by doping with donors e.g., S, Si, or Sn. For p-type production, Zn and Mn acceptors are used.

High-resistance InP

We use doping of Fe atoms, which create in InP a deep acceptor level which compensates most shallow donors and creates a pseudo-intrinsic high-resistance state. Electrical resistance is strongly dependent on the concentration of Fe. When the concentration of Fe exceeds $1x10^{16}$ cm^{-3}, resistance increases sharply by several orders of magnitude. This high-resistive material is used for the preparation of x-ray and gamma-ray radiation detectors.

3.6. Indium antimonide –InSb

InSb is a direct semiconductor with the forbidden band gap of 0,172 eV at 300K. It is the smallest band gap among III-V semiconductors. It has the greatest electron mobility among all semiconductors: 78 000 cm^2/Vs at 300K.

InSb single crystals are produced using several methods (low melting point of 525 °C):
1. - Czochralski method – dominant. We does not have to be used LEC method, as partial pressure of Sb and In vapors is very low and close to the conditions of Si growing.
2. - Liquid epitaxy LPE, also MBE and MOCVD
3. – Horizontal Bridgman method and the Vertical gradient freeze method (limited cases)

3.6.1. InSb Applications

Hall effect sensors (very powerful Hall effect), laser diodes, wavelength $\lambda = 5.2$ μm
Infrared detectors 1-5 μm photodiodes => thermal image => thermal vision
$InSb_{1-x}As_x$ laser diodes emitting infrared radiation with the desired wavelength.

In 2006, Intel introduced a new transistor prototype using InSb in cooperation with QinetiQ. In combination with the traditional silicon, the new material has unique characteristics according to the manufacturer. It will be possible to reach 150% of the performance of the current transistor type and even achieve electricity consumption ten times lower and significant reduction of the emitted heat.

Ternary compounds based on InSb.

(GaIn)Sb – tunable system of wavelengths between 1.5-7μm. It is used in devices for communication, emissions' measuring and photovoltaics.

(InTl)Sb – tunable system for wavelengths 8-12μm (replacement II-VI semiconductors).

3.7. Gallium nitride - GaN

GaN is a direct semiconductor with the forbidden band gap of 3.4 eV at 300K. The material is very hard and crystallizes in hexagonal wurtzite structure. It is used for the production of high-power and high-frequency components, such as highly luminous blue LED diodes and direct violet (405 nm) laser diodes. Thanks to its radiation resistance, it is also used for the production of solar cells for satellites. Due to a great width of the forbidden band, GaN transistors may work at higher temperatures and voltages than GaAs. This makes them ideal for high-power amplifiers in the microwave area.

Preparation

GaN single crystals are prepared using two methods. From the mixture of atomic nitrogen dissolved in molten gallium at temperatures of 1400-1600°C under the nitrogen pressure of 100 atm or by crystallization with the presence of ammonium-NH_3 at temperatures of 500-800 °C and nitrogen pressure of 100 atm according to one of the following reactions:

$2\ Ga + 2\ NH_3 \rightarrow 2\ GaN + 3\ H_2$

$Ga_2O_3 + 2\ NH_3 \rightarrow 2\ GaN + 3\ H_2O$

A special high-temperature and high-pressure growth chamber is necessary.

Thin layers are produced by MOCVD or MBE on ZnO, SiC or sapphire substrates (lattice fit effect).

GaN Applications

GaN blue laser diodes – reading of Blue-ray discs.

Ternary compounds of InGaN or AlGaN, which may tune their forbidden band width by changing the content of In or Al in GaN and enable the production of LED diodes from red to blue.

White LED diodes.

GaN HEMT – use in wireless high-power applications.

GaN doped by Mn is a promising magnetic semiconductor suitable for spintronics.

GaN may be prepared in the form of nanotubes to be used in nano-electronics, optoelectronics and biochemical sensors.

3.8. Sapphire, Al_2O_3

Sapphire (the name comes from Hebrew – blue stone) is a single crystal of aluminum oxide (Al_2O_3), a mineral known as corundum. It either is found as a natural gemstone or is produced artificially for a various applications. While corundum consists of pure aluminum oxide, sapphires always contain an impurities of other elements (iron, chrome etc.), which give it blue, red, yellow, pink, purple, orange or green color. All varieties of corundum including the red ruby are classified as sapphires.

In 1902, Auguste Verneuil, a French chemist invented the method of preparation of synthetic sapphire, now named after him. In the Verneuil method, pure Al_2O_3 powder is poured into oxygen-hydrogen flame, which is directed against the surface on which sapphire slowly growth. Sapphire is also produced by sintering of Al_2O_3 powder (compressing at high temperature) in an inert atmosphere. Artificial sapphire is identical with the natural sapphire. The disadvantages of these methods include the great internal tension in the ingot and a high number of defects. This is why single crystalline sapphire is made predominantly by the Czochralski method, where a rhodium crucible is used. Ingots with the diameter up to 400 mm and weight of 500 kg can be made. In 2020, the global production of synthetic sapphire was about 1500 tons, mostly from the China, Japan, USA and Russia.

Other methods of preparation

Floating zone method, horizontal Bridgman method, pulling from a form-shaper, hydrothermal method.

Hydrothermal synthesis is a solution process that closely mimics natural formation, subjecting the minerals to intense heat and pressure in a sort of “pressure cooker.” Sapphires form around a seed crystal as the mineral solution rises to the top of the cooker.

Synthetic sapphire applications

Sapphire glass (The “glass” designation does not mean an amorphous structure but relates to transmissivity. Sapphire is extremely transparent for radiation with wavelength from 170 nm (UV) to 5.3 µm (IR); at the same time, it is 5x harder than glass (hardness of 9). For this reason, it is used for resistant optical windows in many applications. Sapphire windows are made from monocrystalline ingots cut into slices and optically polished.

Solid-state lasers are another application of sapphire. The very first laser was made from synthetic ruby. Today, titanium-sapphire lasers are popular, which emit radiation with the wavelengths in the red or close infrared area. In these lasers, a sapphire crystal doped by chrome or titanium is irradiated by another intense “pumping laser”, which creates an inverse population of carriers the recombination of which gives rise to a stimulated emission.

Thin sapphire slices are used as insulating substrates in high-power and high-frequency CMOS integrated circuits used in high-frequency application (RF-radio frequency) such as mobile phones, police two-way radios and satellite communication systems. This type of IC is called "SOS" - silicon on sapphire chip and allows integration of digital and analogue circuits in one chip, parallely.

Monocrystalline sapphire slices are also used as non-conductive substrates for epitaxial growth of GaN. A layer of GaN on sapphire is nowadays widely used in the production of blue LED diodes. Sapphire substrates are used for their lower price but also for their low electric conductivity and high thermal conductivity. They are good electric insulators and dissipate the generated heat very well.

3.9. Diamond, C

Diamond is a transparent crystal, which crystallizes in a diamond structure (cubic phase-centered lattice). It is extremely hard (the highest hardness of 10). Diamond has the highest thermal conductivity of all materials (900–2320 W·m−1·K−1) and high electron mobility as well.

Diamond has a great width of forbidden band, 5.5 eV, which represents a wavelength of 225nm (far ultraviolet area). This is why diamond is clear and transparent for visible light. The colour comes from lattice defects and impurities. Some blue diamonds are semiconductors; however, most of them are excellent insulators. The blue colour and increased conductivity come from boron, which replaces carbon atoms in the lattice and behaves as an acceptor.

Most synthetic diamonds are made at high temperature and pressures (High Pressure High Temperature (HPHT) processes). Its yellow colour relates to nitrogen impurities. Chemical vapor deposition (CVD) is another method of growth. The growth occurs at a low pressure (lower than the atmospheric) at temperatures of 700-1000^{o}C. The growth chamber contains a mixture of methane and hydrogen, which decompose, and carbon in the form of diamond deposits on the substrate (Si, sapphire). This method is mainly used to apply several thin (several mm) surface diamond layers.

Diamond applications

Diamond electronics – unfortunately, it is not yet possible to create homogenous large-area wafers.

Diamond coverage – hard, resistant layer

Biosensors

3.10. Silicon carbide, SiC

SiC is a compound of carbon and silicon. Rarely, SiC is found in nature as mineral *moissanite*. It has very high hardness (9.5) and has an indirect forbidden width (2.3-3.2eV). SiC is polymorphous and is found in about 250 crystalline modifications – polytypes (see table below).

Polytype	**3C (β)**	**4H**	**6H (α)**
Atomic number	14+6	14+6	14+6
Atomic weight	40.1 $g \cdot mol^{-1}$	40.1 $g \cdot mol^{-1}$	40.1 $g \cdot mol^{-1}$
Crystal structure	ZnS	Hexagon	Hexagon
Lattice constant	4.3596 Å	4.3596 Å	4.3596 Å
Density	3.21 $g.cm^{-3}$	3.21 $g.cm^{-3}$	3.21 $g.cm^{-3}$
Hardness	9.5	9.5	9.5
Thermal conductivity	360 $W \cdot m^{-1} \cdot K^{-1}$	370 $W \cdot m^{-1} \cdot K^{-1}$	490 $W \cdot m^{-1} \cdot K^{-1}$
Melting point	2730°C	2730°C	2730°C
Forbidden band width	2.36 eV (300K)	3.23 eV (300K)	3.05 eV (300K)
Electron mobility	900 cm^2/Vs (300K)	900 cm^2/Vs (300K)	900 cm^2/Vs (300K)

Silicon carbide is prepared by a reaction of carbon with silicon or silicon dioxide or by a reaction of carbon monoxide with silicone or its oxide:

- $Si + C \rightarrow SiC$
- $SiO2 + 3C \rightarrow SiC + 2CO$ (at 1600-2500 °C, electric arc furnace)
- $Si + 2CO \rightarrow SiC + CO2$
- $SiO2 + 4CO \rightarrow SiC + 3CO2$

Very pure monocrystalline SiC is made by the so-called Lely process, during which SiC powder sublimes in an Ar atmosphere at the temperature of 2 500°C and is then again deposited to a thin SiC monocrystalline substrate.

SiC applications

SiC is used as an abrasive material and is known under the market designation of **carborundum**. By sintering of the SiC grain, very hard ceramic material is obtained, which is used for the production of car brake pads and car clutch and other components under great strain. SiC is widely used in high-temperature/high-voltage electronics, where high electric fields, high temperatures and high currents have to be applied. The very first LED diodes were made from SiC (yellow, blue). However, after mastering the preparation of GaN (10-100 times brighter emission due to a direct forbidden band), the production stopped. Today, SiC is used as an substrate for epitaxial growth of GaN. It is also used for the production of fireproof bricks, as a semiconductor and sometimes as a cheaper alternative of diamond. The light refraction index is 2.56 to 2.68. SiC is used for fabrication of different parts of furnaces (tubes) as it does not melt at any temperature or pressure, has a high SiC sublimation temperature (~ 2700 °C), is chemically stable and has a very low thermal expansion coefficient (4.0×10^{-6}/K).

SiC also may be used for artificial silicon production:

$SiO_2 + 2SiC \rightarrow 3Si + 2CO$

At 2 500°C it decomposes into gaseous silicon and carbon in the form of graphite:

$SiC(s) \rightarrow Si(g) + C(s)$

SiC is used using this reaction also in the preparation of graphene layers at high temperatures.

3.11. Zinc oxide, ZnO

Zinc oxide is a white, powdery substance not soluble in water. It, however, dissolves in diluted acids and hydroxides. In nature, it is present as the mineral *zincite*. ZnO crystallizes in three structures. Most often in the hexagonal wurtzite structure, which is most stable. The cubic ZnS structure emerges in the preparation of ZnO on cubic substrates, and the cubic NaCl structure is the least common, and emerges at the temperatures above 100atm. The lattice constants of wurtzite are 3.25 Å and 5.2 Å. The bond is predominantly ionic. ZnO is soft (hardness of 4.5), has high thermal conductivity and capacity, low thermal expansion coefficient and a high melting point. It is a direct II-VI semiconductor with the forbidden band gap (300K) of 3.4 eV. By adding MnO or CdO, the forbidden gap can be tuned in the range of 3-4 eV.

Zinc oxide can be prepared in laboratory by a thermal decomposition of zinc hydroxide, zinc carbonate or zinc nitrate.

$Zn(OH)_2 \rightarrow ZnO + H_2O$ (most common)

$ZnCO_3 \rightarrow ZnO + CO_2$

$2\ Zn(NO_3)_2 \rightarrow 2\ ZnO + 4\ NO_2 + O_2$

Industrially, it is produced by the zinc burning.

$2\ Zn + O_2 \rightarrow 2\ ZnO$

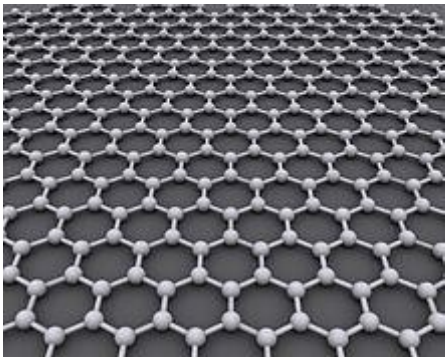
Zinc oxide decomposes into oxygen and gaseous zinc at the temperature of 1975 °C.

ZnO applications

About 50% of zinc oxide is used in the rubber industry. ZnO is an essential part of rubber compounds necessary for the vulcanization of natural rubber to rubber. It is also used as an additive substance into tires due to its high thermal conductivity (it improves heat dissipation in tires). ZnO also has an important use in painting as a white pigment known as the zinc white. It is used in the production of special chemically resistant glass and glazes in the ceramic and glass industry. Zinc oxide is also one of the intermediate products during industrial production of zinc from sphalerite ore. Small particles of ZnO (similar as Ag) have various antibacterial effects. This is why they are used in the clothing industry, in the food industry (food storage), medicine (dental medicine) and dermatology. They are often contained in sunscreens (high refraction index, blocking of UVA and UVB, do not penetrate the skin and are antiallergenic). ZnO nanoparticles, on the other hand, penetrate the skin and may cause unpredictable complications. ZnO is added to many food products as a nutritional source of Zn. ZnO is also used in cigarette filters.

Thanks to its direct forbidden band, ZnO may be used in the production of laser and LED diodes with the wavelength of 375 nm, similarly as GaN (it, however, has a brighter emission spectrum). Due to lattice match, it is used as a substrate for epitaxial growth of GaN. High resistance against radiation damage makes ZnO suitable for electronic applications. ZnO doped by aluminium is conductive, transparent, safe for human health and cheap. This is the reason why it is starting to replace widely used conductive layer of ITO (indium tin oxide), e.g., in solar panels or LCD displays. Due to its very high piezoelectric coefficient, ZnO is used in a wide range of piezoelectric applications.

3.12. Graphene - C



Diamond and graphite are the best-known forms of the chemical element carbon. Graphene has a very similar crystallographic organization to graphite. It consists of one or two layers of carbon atoms, which make up a planar network organized into the shape of hexagons connected through sp2 bonds (a shape looks like honeycomb). Single-atom graphene layer without impurities is highly electrically conductive; two-atom layer behaves as a semiconductor. Graphene was discovered by Andre Geim and Konstantin Novoselov from the University of Manchester in 2004 and they were awarded the Nobel Prize in physics in 2010 for its discovery. These two-dimensional structures were theoretically studied many decades earlier; however, physicists at the time thought that single-atom layers organized into a regular crystalline lattice cannot exist, as they would be very unstable. The reason were thermal vibrations of the crystalline lattice, which were supposed to be so strong that movement of atoms and transformation of parts of the single-atom layer into nanotubes and fullerene structures would occur. Geim and Novoselov prepared graphene from ordinary graphite using adhesive tape. They exfoliated thin surface carbon layers from a graphite crystal. They then obtained little graphene scales by dissolving the tape in a diluent.

Graphene has unique properties. Regular single-atom crystalline lattice of graphene presents a unique physical anomaly, where quantum phenomena take place. Electrons move in graphene with virtually no collisions and at a very high speed, unimaginable in other solid materials, which is crucial for the speed of transistors and thereby for the development of computers built upon them. Scientists from an IBM research centre presented a graphene transistor with an extremely high frequency of 100 GHz (it is difficult to reach frequencies over 5 GHz with silicon). Electrons in graphene achieve the highest mobility of all known materials. Thanks to the unique band structure, electrons virtually behave as if they had no mass and can theoretically move at the speed of light! There are speculations that graphene could be used to proof the so-called Klein paradox. According to the current knowledge, this phenomenon should have occur in very rare conditions only (e.g., near black holes). If this were to be proven, it would pave the way for development of a completely new type of transistors. Graphene, like fullerenes, is expected to become the basis of the next-generation carbon electronics.

Graphene has the highest thermal conductivity from all known materials, it is extremely firm in tension and, similar as silicon, has the properties of a semiconductor. In the future, graphene could be used in “pocket” supercomputers, in touchscreens, solar cells or as part of conductive and extremely mechanically resistant plastics. It could also be used as a construction material in, e.g., the construction of cars, aircrafts and satellites. Graphene is at present one of the most investigated materials in the world.

Graphene is also both the thinnest and strongest material in the world. It is about two hundred times stronger in tension than steel. Graphene is so strong that a force of 20 000 N would have to be applied to cut a 100 μm thick membrane (only slightly thicker than a hair), if the atoms of this layer were bound as tightly as the graphene atoms.

Graphene is also extremely thermally conductive. It was found that although applied to the SiO_2 substrate, the heat dissipation capacity of graphene decreased five to eight times, but it was still twice as good as copper and fifty times better than thin silicon wafers.

There are plans to use graphene in the production of displays, solar panels and various detectors as a thin, electrically and thermally conductive, elastic transparent layer to electrodes. Currently, most applications use the ITO (indium titanium oxide) conductive layer. Apart from the fact that a graphene layer is cheaper, display prototypes are very flexible and elastic. Graphene could therefore be used in the production of flexible or even scrolling displays usable not only in telephones and computers, but also, for example, "e-paper". Presently graphene is also used for a production of electrodes into quickly rechargeable batteries.

Preparation

Only a decade ago, graphene was more expensive than gold and the biggest produced parts of graphene could only be seen under a microscope. Due to the enormous interest in graphene, the preparation technologies are quickly evolving. The price of an inch-sized graphene slice, once estimated at hundred million dollars, is now only few dollars. The global production is estimated at hundreds of tons per year.

1. Scotch tape technology - exfoliation of surface layers using an adhesive tape.

2. Thermal decomposition of silicone carbide (SiC) which is reduced into graphene at the temperature of 1100 °C. However, pieces of material obtained by this method had different size and properties, which is why they were not very suitable for industrial use. They have to be sorted in a complicated way according to size, shape, and conductivity. It is also not possible to prepare a single-atom or a two-atom layer by this method.

3. Deposition of graphene on copper. A copper plate heated to 1000°C is reacted with pure methane. Methane decomposes into hydrogen and carbon, and carbon vapours adhere to the copper surface as thin layers of graphene with the thickness of only one or few atoms. This layer is separated and transferred to a plastic foil. The only disadvantage is, so far, the fact that graphene is produced at very high temperatures in this method.

4. Decomposition using hydrazine (N_2H_4). The base of this method is the ordinary paper immersed in a hydrazine solution. Carbon bound in the paper is reduced to atomic carbon by hydrazine. Surprisingly, this technology keeps the original “paper” structure, and a very thin (maybe mono-atomic) large-area layer is created. However, it has been full of defects and foreign components.

5. The number of pure graphene production methods is quickly growing. Recently, dispersion of graphene oxide in water is becoming dominant. Graphene flakes are connected easily through chemical bonds and create graphene wafers or graphene “paper”. In laboratory, “soluble graphene” may even be created by chemical modification of graphite microcrystals by a mixture of sulphuric and nitric acids with subsequent steps of oxidation and peeling of layers into small flakes.

Biological properties of graphene

Another interesting property of graphene is its ability to kill bacteria together with non-toxicity for humans. Microorganisms do not grow on the graphene-coated surface. Laboratory tests with *Escherichia coli* (indicator of intestinal infections) have shown that bacteria are unable to hold onto or replicate on such surface. The contact with a graphene layer kills practically all of them (viability drops by 98.5%).

It was found that a mere contact with bacteria disrupts their surface membrane and the content of their cytoplasm spills into the environment. At the same time, graphene is not cytotoxic. In the other word, it is able to kill bacteria but is inert towards mammals’ live cells. This duality is unique. On the one hand, biocompatibility exists meaning that body cells can grow on such surface, on the other hand, the contact with such surface becomes fatal for bacteria. For other known materials it valid, what is antibacterial is also cytotoxic. This peculiarity of graphene opens up another wide field of use unrelated to electronics.

A foil from TWEEN composite material (polyoxyethylene sorbitol ester) coated by graphene oxide, which has strong anti-bacterial effects was produced. The method was further developed by Chinese scientists, who prepared a foil from modified graphene (graphene and graphene oxide form the basis) using a common technology of vacuum filtration. Both these methods make it possible to produce a thin foil, which is, in slight exaggeration, called graphene paper. A nano-sheet from oxide graphene has already been prepared, whose thickness is only 1 nm, as well as a sheet of reduced graphene oxide. Though foil can be produced from both components, graphene oxide appears as more suitable for the production of stronger and more flexible material. One vacuum filtration process yields a layer 1.5–4.6 micrometers thick.

In China, they plan to use a flexible and strong graphene foil in the shoe industry. It will also be possible to make self-disinfecting food-wrapping materials from graphene.

Boron nitride- BN

Boron nitride is mostly known as a representative of the so-called non-oxide ceramics used for grinding and cutting tools. In this case, however, it is a cubic structure (c-BN), which is suitable for steel machining. Hexagonal structure (h-BN) is also well known. It has found application in cosmetics or as a solid lubricant ensuring the lubricity of surfaces at high temperatures. Unlike graphene, h-BN is an insulator. It has been known for a relatively long time; however, until during last decade a way, how to implant “islands” of h-BN to a graphene layer, was found. This means, among other things, that we can manage electrical behavior of graphene using BN.

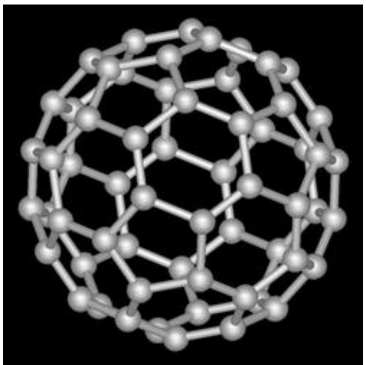
The color of boron nitride is naturally white, so the material was nicknamed white graphene (in older studies also white graphite). BN is complementary material to classic graphene – it is therefore an ideal non-conductive substrate for electronic parts based on “black” graphene. Desirable properties of this material also include the fact that it has a high optical transmittance in the atomic layer and that even in combination with classical graphene it is a transparent material in a wide range of the light spectrum. The resulting mechanical properties of “white graphene” are similar to those of “black” graphene, including its flexibility. Thus, these properties predispose this material to the nano-electronics applications.

One to five-atom thick slice of pure boron nitride (h-BN) was formed using rough boron nitride substrate sublimation at the temperature of around 1000 ^{o}C and the gas phase deposition on the copper substrate. The advantage is that the prepared foil may be removed from such surface and transferred to another substrate. Slices of white graphene with the size of 5 x 5 cm were successfully prepared; however, in theory, their size is not limited in any way. All depends on the dimension of the copper foil to which the application is made and on the dimensions of the furnace in which the steaming of the rough boron nitride substrate is performed.

In the future, h-BN layers in combination with conductive graphene layers could break the so-called Moore law (the number of transistors placed on a surface unit area of the integrated circuit is doubled each two years). Now, it should be the increase in orders of magnitude. But it will not be just transistors, h-BN should enable the production of completely new types of high-capacity small capacitors. Quantum biosensors should follow…

3.13. Fullerenes - C

Fullerenes are spherical molecules composed of five- or six-membered rings of carbon atoms. Spatially, these molecules are organized into a spherical shape and are extraordinarily resistant to external physical influences. They were named after Buckminster



Fuller, an American architect, who designed geodetic domes of a similar shape. Robert F. Curl and Harold W. Kroot were awarded the Nobel Prize in chemistry in 1996 for the discovery and studying the properties of fullerenes. Currently, the investigation of properties and methods of fullerenes' preparations is being studied intensely at many scientific institutions all over the world.

Fullerenes up to the molecule containing 96 carbon atoms have been systematically explored. Organic chemistry developed the most around molecules with 60 and 70 atoms. The most stable fullerene known so far contains 60 carbon atoms. Its pure crystalline form, which is harder than a diamond, was given the name fullerite. Fullerene molecules are often hollow and close to a sphere in their shape. Some small molecules, such as hydrogen, helium, or lithium, may be inserted into the molecule lattice of fullerenes. Molecules thus inserted by this way then affecting the resulting properties of fullerenes. For metallofullerenes, i.e., fullerenes that have one or several metal atoms placed in their hollow space, higher molecules such as C80 or C82 are often used as the base. Carbon nanotubes are also derived from fullerenes. On the one hand, carbon nanotubes respect the main topological feature of fullerenes – composition from a variable number of hexagons and fifteen pentagons; on the other hand, their typical shape is an elongated cylinder. Recently, it has become possible to prepare fullerenes smaller than C_{60}, e.g., C_{36} and molecules of the fullerene shape made up by other elements than carbon. The so-called BN-fullerenes, for example, contain only boron and nitrogen.

There are four known methods of fullerene preparation by pyrolysis:

- The dominant method of preparation uses evaporation of graphite in the electric arc in the inert gas atmosphere.
- The method of fullerene preparation in flames of various organic substances is possible but not very common.
- The third method works with sunlight concentrated into a focal point using a mirror, in which graphite is placed.
- The last method uses laser pyrolysis of organic compounds.

The gradual development of industrial production significantly reduced fullerene prices. The most time-consuming phase of the whole preparation process concerns the separation of individual fullerenes using liquid chromatography.

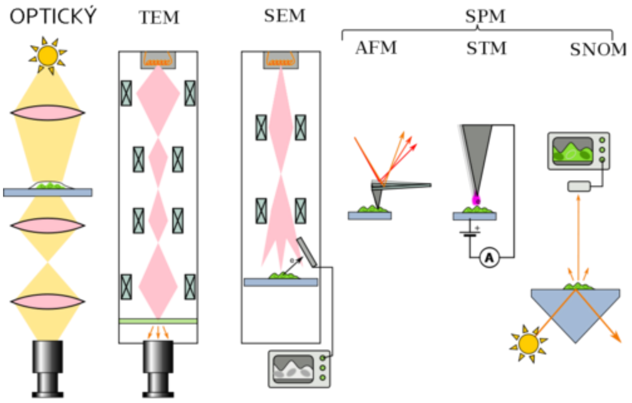
Application

Currently, the largest funds are invested into the research of fullerenes as new promising materials for technology. One of the most important properties is their superconductivity. It has been shown that it is possible to form C60 compounds with alkali metals, which are superconducting at temperatures of 18 K and higher.

Healthcare is another promising field of fullerene application. It could be used them as a carrier of medical drugs

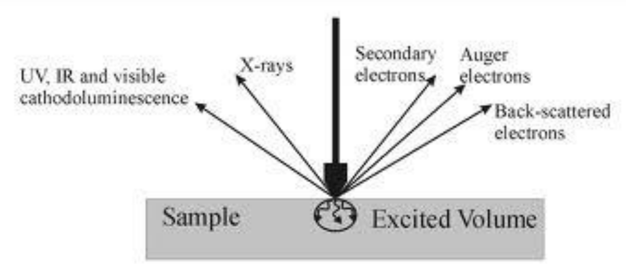
4. Characterization methods

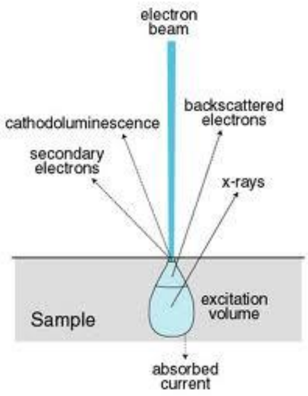
4.1. Microscopic surface analysis

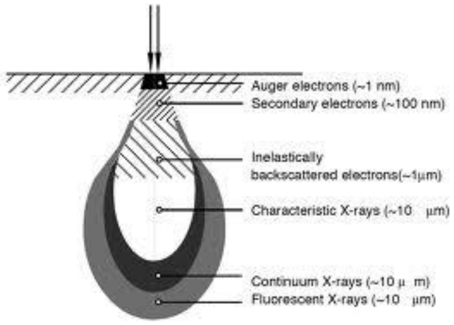


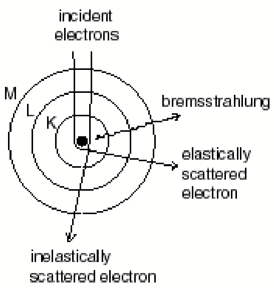
Electron microscope is an electronic equivalent of an optical (photon) microscope. Optical lenses are replaced by electromagnetic lenses and electrons are used to examine objects instead of photons. The problem of optical microscope is that as the wavelength of the used light decreases, the size of usable magnification decreases as well. In the optical microscope, the threshold of smallest observable details is clearly determined exactly as the half of the wavelength of the used light radiation. The wavelength of accelerated electrons (Broglie wavelength), on the other hand, is several orders of magnitude lower than the wavelengths of visible light photons. For this reason, electron microscope has much higher resolution and can thus achieve much higher magnification (up to 1 000 000 times).

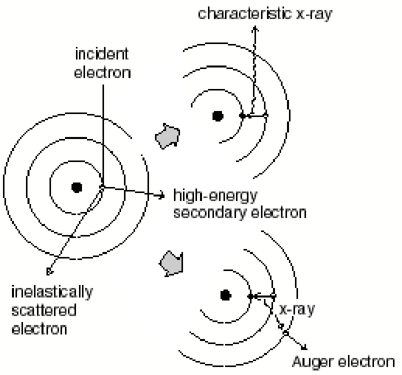
In an electron microscope, optical lenses are replaced by suitably modulated electromagnetic fields. The observed object is placed in a vacuum, which is radiated by a beam of electrons emitted from a hot cathode. Electrons are accelerated by an electric field and hit the object, which they either pass through or are scattered. The interactions between the primary incident electrons and the atoms of the specimen may be divided into two groups: elastic collisions, which are responsible for the backscattered electrons emission, and inelastic ones, during which the primary electrons transfer their energy to the atoms of the sample, followed by a release of the secondary and Auger’s electrons, x-ray radiation and cathodoluminescence.











A variety of techniques are based on the detection of secondary signal, which differ from one another especially by the way of atom ionization and by the type of the secondary signal. These include mainly the following techniques:

X-ray fluorescence (XRF) spectrometry – x-ray radiation is used for the ionization and the generated characteristic x-ray radiation is detected.

Photoelectron spectroscopy (XPS) – x-ray or UV radiation is used for the ionization, and the kinetic energy of electrons generated by ionization is detected.

X-ray microanalysis – a narrow beam of accelerated electrons is used for the ionization, and the generated characteristic x-ray radiation is detected.

Auger electron spectroscopy (AES) – a narrow beam of high-energy electrons is used for the ionization, and the kinetic energy of Auger electrons is detected.

PIXE – a beam of accelerated protons is used for the ionization, and the resulting characteristic x-ray radiation is detected.

X-ray absorption spectroscopy - the spectrum of absorbed x-ray radiation is measured.

Microscope types

- **TEM** - Transmission electron microscope – stationary electron beam, detection of electrons that passed through the sample on a fluorescent screen or through a detector.
- **SEM** - Scanning electron microscope – movable beam, image of the surface is visualized thanks to scattered secondary electrons. In literature sometimes the designation “raster” is used apart the name “scanning” (or in Czech “line electron microscope”), which indicates, that in the microscope’s operation, the primary beam moves over a certain area of the specimen.

Methods

- **SPM** - Scanning Probe Microscopy, is a set of methods designed to determine the surface structure with atomic size resolution.
- **AFM** (Atomic force microscopy) – Atomic force microscopy is based on mapping the distribution of atomic forces on the sample surface. These forces are mapped by getting the tip closer to the surface, creating either an attractive or a repulsive force, which causes bending of the cantilever in which the tip is attached. This bending is detected by a laser sensor. The possibility to study both conductive and non-conductive samples is the advantage of the AFM method.
- **STM** (Scanning tunneling microscopy) – The principle of STM is based on quantum physics. A current flows between the electrode tip and the investigated sample thanks to the tunneling effect, even if the tip does not directly touch the sample. During surface scanning, the tip distance changes so that the tunneling current remains the same. As one of only a few methods, this method is capable of providing even atomic resolution while being relatively simple. Unlike other methods (transmission electron microscopy, field ion microscopy), it does not require complicated sample preparation. On the other hand, it only provides information about the surface.
- **SNOM** (Scanning near-field optical microscopy) – (SNOM or NSOM) is a concept combining the techniques of SPM and optical microscopy. In addition to topography, it also provides optical information. It uses the advantageous properties of the near field, which make it possible to overcome the diffraction resolution limit thanks to the properties of evanescent waves. The location of the source and the radiation detector very close to the sample surface at the distance lower than the wavelength of the incident light, so called the near-field area, is the basis of this method. In this field, there is no significant deviation of the photons from the direction perpendicular to the sample, and therefore the interpretation of the results is easier. Light interacts with the sample and the light, which is reflected or scattered, is detected by photomultiplier tubes and converted into an electric signal.

High-resolution techniques such as AFM, STM or SEM provide good information about the structure and topography but almost no information about the chemical nature of the investigated sample. SNOM, on the other hand, thanks to the electromagnetic radiation can be combined with various analytical methods, which significantly increases their sensitivity.

4.1.1. Transmission electron microscopy (TEM)

Transmission electron microscope uses an electron beam emitted by a hot cathode, accelerated at high voltage and shaped by magnetic lenses. An electron beam that passes through a very thin (semi-transparent for electrons) sample, contains information about the internal structure of the sample. The spatial change in this information ("image") is then magnified by a series of magnetic lenses before hitting the fluorescent screen, a photographic plate, or a sensitive sensor such a CCD camera. The CCD image can be displayed in real time on a monitor or computer.

High-resolution TEM (HRTEM) is limited by a spherical and a chromatic defect. Repair software of the spherical defect recently allowed images to be taken with sufficient resolution (magnification of 50 million times) to visualize carbon atoms in diamond (0.89 A=89 pm) or individual silicon atoms (0.78 A=78 pm). The ability to determine atom positions within a material has made HRTEM an indispensable tool for nano-technology research and development in many fields, involving different types of analyses and development of semiconductor components for electronics and photonics.

4.1.2 Scanning electron microscopy (SEM)

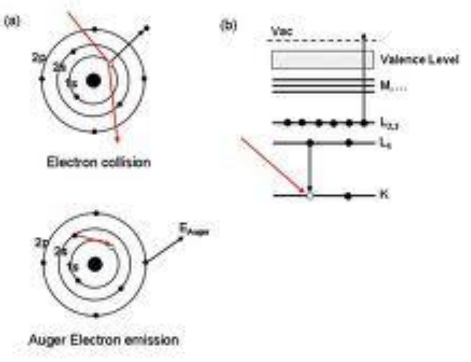
Scanning Electron Microscopy (SEM) analyses the signal of secondary electrons emitted from the sample after the impact of a primary high-energy (~ 10-100 keV) electron beam. In SEM, the primary electron beam is scanned across the surface of the sample. The secondary electron signal (corresponding to the position of the primary beam) is mapped using a detector. It is similar to an optical microscope, but unlike it, the resulting image is made up of a secondary signal - secondary or scattered electrons. Thanks to this, SEM is considered an indirect method. The energy of the primary electrons, given the accelerating voltage, affects the shape of the excitation volume, which becomes shallower as the applied voltage value decreases and the area from which the secondary electrons are released does not match the diameter of the primary electron beam. This reduction in the area from which secondary electrons are released is another cause of the reduction in the resolution of the microscope when the accelerating voltage decreases. The depth of penetration of primary electrons is also influenced by the composition of the sample. Obviously, a sample consisting of heavier elements (e.g. metals) will produce more scattered electrons than a sample prepared from light elements, and the depth of penetration of primary electrons will be smaller.

Secondary electrons are used in the SEM to show the surface morphology of the sample. Secondary electrons are electrons excited from the atomic shell's orbital. They are distinguished from back-scattered electrons by their low energy and speed. To be able to access the secondary electron detector, electrons need to be accelerated using a grid with a bias of about 10 kV. The detector is usually a scintillator (e.g. YAG). The brightness of the beam synchronously scanning across the screen is directly proportional to the signal from the secondary electron detector. Because of the low energy of the secondary electrons, more secondary electrons reach the detector from the surface hills, which results into a higher signal intensity from the detector and thus a bright spot on the screen, the opposite occurs in the case of the valleys. This gives a topographical contrast that allows the surface of the sample to be displayed in multiple magnifications. The production of scattered electrons, as mentioned above, depends on the mean atomic number of the sample. It follows that the bright areas will appear on the screen to be those with a higher middle atomic number (i.e., composed by heavier elements). On the contrary, areas composed by light elements will appear as dark places. The image in the scattered electrons is able to distinguish regions with different element compositions.

Inspecting a sample by SEM is often accompanied by interfering phenomena. These include, in particular, charging the sample surface onto which the negatively charged primary electrons fall when it is not electrically conductive enough. As a result, the primary electron beam deflects, oversaturating the secondary electron detector.

SEM is a widely used technique in the industry. The number of other signals generated in the sample chamber during interactions of accelerated electrons with the sample mass such as x-ray radiation, Auger electrons, cathodoluminescence and others, in addition to the above-mentioned signals, is used for detail characterization of the sample. When detecting them, it is possible to determine, for example, the composition of the sample in a given area and, when compared to an appropriate standard, the quantitative representation of the individual elements. In combination with an Energy Dispersive X-ray Spectroscopy (EDS) detector, it allows local identification of the chemical composition.

4.1.3 Auger Electron Spectroscopy - AES



Auger Electron Spectroscopy is a surface analytical method, in which the primary excitation of a sample is performed using high energy *electron beam* (E-AES) or *x-ray photons* (X-AES). The incident electrons excite the electrons in the electron shells of the surface atoms, and when they recombine, the "Auger" electrons are emitted. When the incident energy is high enough, the electron is kicked out from some inner electron level and the free place is immediately filled with an electron from the higher shell. The excess energy is either released by the emission of X-rays or is passed to another electron that can also be emitted (double atom ionization). In AES, the kinetic energy of emitted x-ray photons is measured, which is directly proportional to the atomic number of individual elements present on the surface and just below the surface. The peak intensity is proportional to the number of atoms present (quantitative analysis).

The electron beam can be scanned across a certain surface area or can be aimed at exactly one spot on the surface. This focus (10nm or less) makes AES an extremely sensitive method for element analysis of small surface objects. The Auger scanning microprobe is an example of Auger spectroscopy, which provides an image of area distribution of the observed element. In combination with ion-beam etching, it also allows for in-depth chemical analysis.

- Defect analysis
- Particle analysis
- Surface analysis
- Small-area depth profiling
- Process control

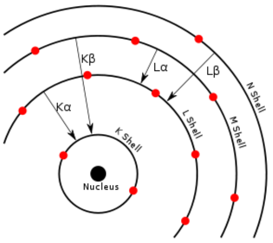
- Thin film analysis composition

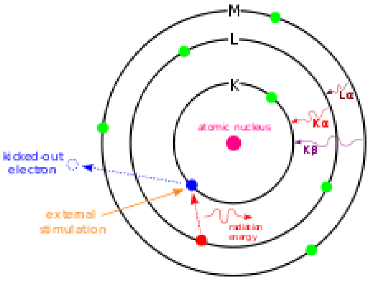
4.1.4. Rutherford Backscattering Spectrometry (RBS)

RBS is a method used for the element analysis of thin layers that does not require the use of reference standards. During RBS, the examined material is shot at a certain angle by high-energy (MeV) charged particles, protons, or helium ions (He^{2+}). The particles are elastically scattered in the material, and the distribution of energy and the yield of the backscattered particles at a certain angle are measured. Since the capture cross-section for backscattering is known for each element, it is possible to obtain the depth profile of each element (only for layers thinner than 1mm). The magnitude of the energy of the backscattered ions that we detect at a certain angle is affected by the loss of energy due to the momentum change upon the collision with the nucleus of the sample atom and the loss of energy caused by the particle passing through the crystalline lattice of the sample. The detection limits of the RBS method are in the range of 10^{13}-10^{15} atoms/cm^2, the depth resolution averages 10 nm, and can be improved in the appropriate measurement geometry. Mass resolution can be significantly improved by using heavier ions. It is possible to analyze dielectrics, thin metal films, inducing dopants in silicon, III-V and II-IV compounds.

- Thin film composition/thickness
- Determine real concentrations (atoms/cm^2)
- Determine film density (when thickness is known)

4.1.5. Energy Dispersive X-ray Spectroscopy (EDS)



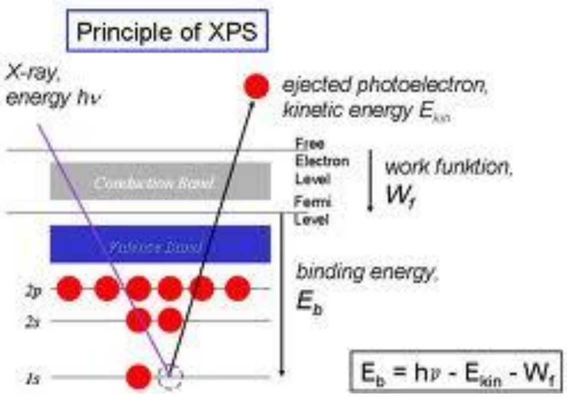


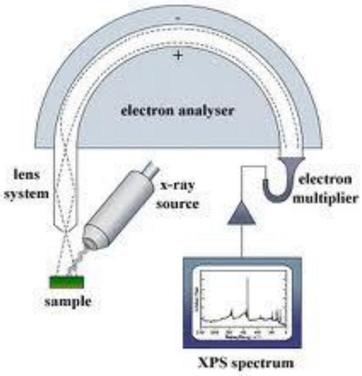
Energy-dispersive X-ray spectroscopy (EDS or EDX) is an analytical method used for local chemical analysis of samples. In this method, the surface of the sample is irradiated by a primary beam of high-energy electrons, which stimulates the emission of characteristic x-ray radiation. **X-ray fluorescence** - X-ray fluorescence spectroscopy- **XRF**, which uses a beam of protons, rtg or gamma rays is an alternative. The incident beam kick-out electrons from the inner shells of the atomic orbital. Subsequently, electrons from the outer shells (higher energy) fill the places after the electron has been kicked out, and the energy difference is emitted in the form of photon X-ray radiation. The energy will be lower (greater wavelength) than the energy of the original X-ray photon (smaller wavelength). The amount and magnitude of energy is measured by an energy-dispersive spectrometer. From this energy, which is characteristic of individual atoms, it is possible to determine the chemical composition of the sample. The energy-dispersive spectrometer is mostly embedded in a SEM (SEM-EDS) or electron microprobe. It consists of an x-ray detector, which converts the energy of the radiation into a voltage signal, which is measured by a pulse processor and analyzed in a multichannel analyzer.

- Identification of chemical composition in small areas.
- Mapping of presence of different elements in structural defects.

4.1.6. X-ray Photoelectron Spectroscopy (XPS)

The XPS method can be used for qualitative, quantitative, and structural analysis. The surface of the sample is irradiated by high-energy x-ray, which excites the electrons on inner shells with binding energies between 0 eV and 1500 eV, which release the sample due to the external photoelectrical effect. Electrons are emitted only from a thin layer of the sample surface. The signal is magnified and transferred into an electromagnetic Time-of-flight analyzer. The binding energies of inner electrons are characteristic for each element, and it is therefore possible to clearly establish the presence of elements in any mixture.





- Overlap of individual element lines is virtually absent here
- Under constant conditions, qualitative analysis may be performed by determining the height or the area of the peak
- The calibration curve method may also be used for quantitative analysis
- The fact that inner electron orbitals close to valence orbitals are partially influenced by the emergence of a chemical bond may be used for structural analysis – ***chemical shifts*** occur.
- Correlation of chemical shifts with the change of charge may be used to establish the oxidation number of the atom and to determine the change in bond polarity.
- When studying organic compounds by the XPS method, also shifts caused by the connection of different functional groups with the given carbon atom may also be observed.

4.2. Bulk and surface chemical analysis - Destructive methods

4.2.1. Glow Discharge Mass Spectrometry (GDMS)

In mass analysis, ions are separated according to the mass-to-charge ratio (m/z). The movement of charged particles in an electric and magnetic field in a vacuum can be described by two equations:

Lorenz force $\mathbf{F} = Q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$
Second Newton law $\mathbf{F} = m\mathbf{a}$ (non-relativistic case, applies only for speeds below the speed of light)

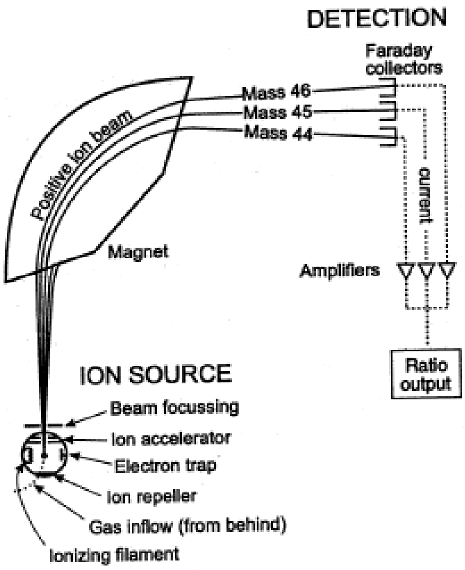
From these equations, we get a classic differential equation describing the movement of charged particles in time and space:

$$(m/Q)\mathbf{a} = \mathbf{E} + \mathbf{v} \times \mathbf{B}.$$

Usually, a dimensionless ratio of m/z is introduced, where z is the electrical charge of the ion (z=Q/e).

A variety of mass analyzers using a combination of static or dynamic electric and magnetic fields are used to analyze individual ions. Most often, a combination of electrostatic and magnetic field (spatially separated - ***sector instrument***) is used for mass analysis, where the fields bend the trajectory of individual ions, which then fall onto the mass detector. The combination BEB (magnetic-electric-magnetic) is widely used. One of the most modern principles is the double focusing sector instrument, in which ions are separated at the same time according to their trajectory and velocity.

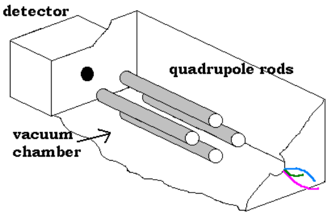
Time-of-flight (TOF) mass analyzers use static electric field to accelerate ions and measures the time it takes for the ions get to the detector. If ions are equally charged, they will have the same kinetic energy and their speed will depend on their mass only. Lighter ions will reach the detector sooner.

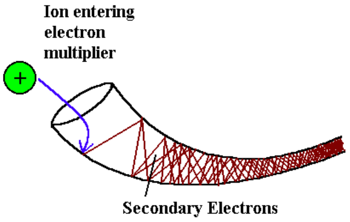


A quadrupole mass analyzer consists of four metal rods put into a rhombus shape. A combination of DC and AC oscillating voltage pulsing with a radio frequency (RF) of 13.5 MHz, is put to individual rods, so that the opposite rods always have the opposite potential. The analyzed ions are directed into the middle between the rods. When a specific DC and AC bias is applied, only ions with a selected m/z - mass/charge ratio (atomic mass/atomic number) passes along the rods to the detector. Other ions are pushed outside the center. There are many combinations of voltages, which allow for individual detection of most ions. Quadrupole analyzers are compact and simple; however, their resolution capability is low for ions with the same or very similar m/z ratio. A mass spectrometer often uses a triple quadrupole configuration with three subsequent quadrupoles.

The final element of a spectrometer is a detector that measures either an induced charge or a current produced by passing of an ion or by its collision with the detector surface. Ion-electron multipliers, Faraday cells or ion-photomultipliers are usually used as detectors. Microchannel planar detectors are used in modern instruments.

The Channeltron - ion electron multiplier - is a special type of commonly used ion detector. It is a tube in the shape of a horn, to which high voltage of opposite polarity to that of detected ions is applied. Ions which leave the quadrupole analyzer are sucked into the tube. When hitting the wall, secondary electrons are emitted in an avalanche-like effect, which reach the quantity of 10^8 electrons at the end of the tube (a photomultiplier works similarly).



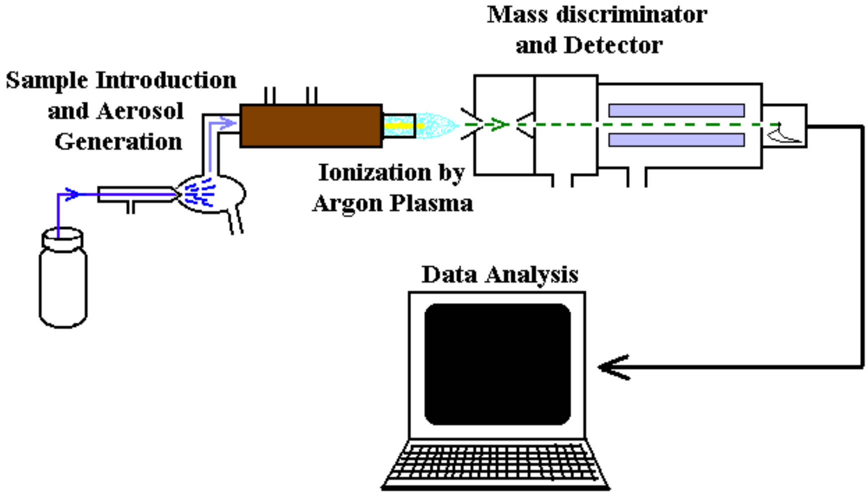


For both the bulk and surface analysis, glow discharge mass spectrometry (GDMS) or glow discharge optical emission spectrometry (GD-OES) is widely used. The sample represents a cathode which is located in a low-pressure (~ 100 Pa) glow or plasmatic discharge. In the discharge, ionization of a gas occurs (argon is the most commonly used gas) and positive Ar ions are then accelerated in the electric field towards the sample surface (cathode) reaching the energy of 100 eV-1000 V. This energy is sufficient for de-sputtering of surface layers of the sample, which become atomized. Neutral atoms de-spattered from the surface are ionize again and are directed to a mass spectrometer-quadrupole analyzer. As a matter of principle, depending on the sputtered time, individual surface layers are gradually de-spattered, so that a depth profile analysis ranging from 2 nm to 500 µm can be performed by monitoring the intensity depending on the sputtered time. A significant advantage of this method is that we can also get information on elements such as hydrogen, nitrogen, oxygen or chlorine.

The use of GDMS is limited to conductive samples. Non-conductive materials may only be analyzed when a sample surface will be covered by a conductive layer as a secondary cathode. In this case, GDMS sensitivity depends on the purity and properties of the material used for covering. The main drawback is that the signal stability and intensity may be low and the background signal from the secondary cathode may not be separated from the analyzed signal. In the case of powder materials, there is a possibility of mixing with a pure conductive substance such as graphite or metals such as Ag, Cu, In etc. These materials may be pressed into tablets or onto a surface of a conductive metal. All these methods, however, increase the risk of contamination during the sample preparation and reduce the sensitivity of the determination for elements present in trace amounts.

- Trace and ultra-trace element determination in high purity metals, alloys, manufactured carbons and graphites, semiconductor materials, various oxides, carbides and ceramics
- Depth profiling of flat surfaces for major, minor and trace elements.
- Identification of unknowns from minute amount of specimen
- Full element characterization of powders or particulate materials.

4.2.2. Inductively-Coupled Plasma Spectroscopy (ICP-OES/MS)



In the ICP-MS method, all samples are transferred to a liquid state (solution, emulsion), in which they are decomposed into neutral atoms in a high-temperature Ar plasma and analyzed based on the m/z ratio (mass to charge-atomic mass/atomic number). We can divide this method into four steps: Transferring sample into aerosol, Ionization of the aerosol in argon plasma, Mass/ charge ion separation, and Detection.

Liquid samples are sprayed using a nozzle, through which argon flows at a high speed creating fine fog. This fog is sucked into a spraying chamber, through which pass only very small droplets (about 2 % of droplets from the original fog), which are then gassed gas in an inductively-coupled plasmatic discharge, which produces a temperature of about 8,000°C. In the hot plasma, all solvents are removed, droplets are atomized and then ionized. ICP-MS can also analyze solid samples, but these either have to be solve in very pure solvents or undergo laser ablation first.

To detect the ions, a quadrupole analyzer is used. When a specific DC and AC bias is applied, only ions with a selected m/z - mass/charge ratio (atomic mass/atomic number) passes along the rods to the detector. The most commonly used ion detector in ICP-MS is the channeltron-ion electron multiplier, the principle of which is similar to a photomultiplier.

ICP-MS is an extremely sensitive method with a low detection limit for a wide range of elements. The normal detection limit is ppt (particle per trillion), for some elements up to ppq (particle per quadrillion).

Optical emission spectroscopy (**OES**) is an alternative to ICP-MS where ions are thermally excited in hot Ar plasma, and during recombination emit light photons of a specific wavelength. The light is analyzed by a spectrometer with monochromator, where intensity of individual wavelengths is measured. This can be recalculated using standards for individual elements' concentration.

- Bulk quantitative survey analysis of major, minor or trace constituents
- High accuracy determinations of major and minor constituents in a wide range of materials
- Quality and process control

4.2.3. Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS)

Laser ablation causes a surface destruction of material, which includes melting, evaporation, sublimation, erosion, explosion and injection of macroscopic particles, molecule fragments, atoms, ions or electrons. These processes depend on the wavelength and the radiant power density of the laser radiation and also on the optical and thermal properties of the sample. Two types of mechanisms can be distinguished according to the radiant power density of the laser beam. In the case of a lower power density ($< 10^6$ W/cm^2) and longer pulse duration, this is a thermic mechanism (evaporation). In the case of higher power density and shorter pulse duration, non-thermic mechanism (ablation) occurs.

In LA-ICP-MS, a pulsing laser beam is used to evaporate the surface layers of the examined material. The produced aerosol is transposed into a cell with inductively-coupled Ar plasma, similarly as in the case of ICP-MS.

LA may be used to analyze any solid sample, for which a very small amount (micrograms) are required unlike for ICP, where at least milligrams are needed. In addition, the focused laser beam enables both area and depth analysis of very small spots (typically micrometers).

- Major, minor and trace level compositional analysis of conductive, semi-conductive and nonconductive materials
- Contamination of plastics, pharmaceuticals organics or biological materials
- Failure, contamination and inclusion analysis
- Forensics analysis
- Environmental and mineral sample analysis
- Elemental distribution analysis

4.2.4. Atomic Absorption Spectroscopy (AAS)

AAS is an analytical method used to determine the content of trace as well as significant concentrations of individual elements in the analyzed solution. Using this method, over 60 elements of the periodic table may be analyzed with the sensitivity from hundredths to hundreds ppm. The solution of the analytical sample is turned into an aerosol using ultrasound spraying and produced aerosol is inserted into a flame or graphite atomizer, where the solution evaporates immediately and breaks down chemical bonds in the molecules of the present compounds. The conditions of the atomization are chosen so that as large a population as possible of measured atoms remains in a neutral state and there is no ionization accompanied by the formation of charged particles like Metal+.

The light beam produced from a special discharge lamp and passes through the flame ionizes analyzed atoms localized in the flame. During ionization, photons are absorbed, the intensity of the passing light thus decreases following the Lambert-Beer law:

$$I = I_0\, e^{-(k * n * l)},$$

where I_0 is the intensity of the incident radiation, I is the radiation intensity after passing through the absorbing environment (flame), k is an atomic absorption coefficient for the given absorption line, n is the number of atoms of the analyzed element in a volume unit and l is the length of the absorption layer (diameter of the flame).

In practice, the logarithm of the light energy loss, so-called absorbance (A), is used as the measured value, for which the following formula applies:

$$A = \log (I_0/I) = 2{,}303 * k * \mathbf{n} * l,$$

which means that for absorbance, there is a very simple linear relationship to the concentration **n** of the atoms of the measured element. All AAS spectrometers therefore give the measured signal in units of absorbance after mathematical processing of actually measured intensities of the passing light.

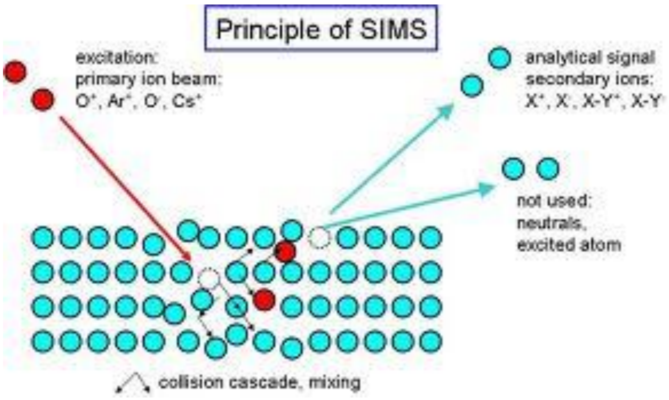
The measurement setup of the AAS method has 3 basic components:

1. Sources of electromagnetic radiation (discharge lamp)
2. Measurement cells in which atomization occurs (flame, graphite cuvette)
3. Optical detection system which measures the intensity of the passed-through radiation with a specific wavelength

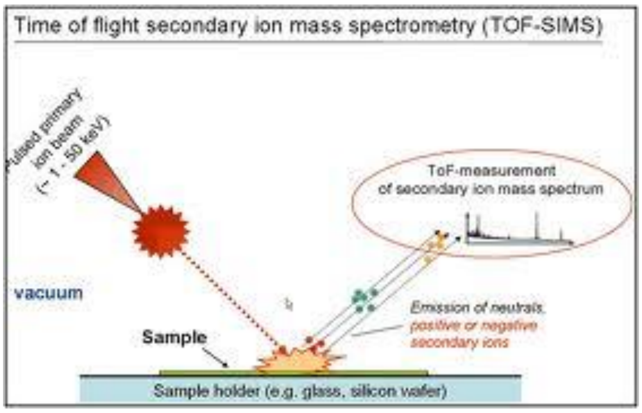
Besides this, each AAS setup is equipped with a setup transporting the liquid sample into a measurement cell and with a computing system which performs whole analysis and evaluates the measured signals, which it converts directly into the target units (e.g., concentration of the measured element in the sample).

4.2.5. Secondary Ion Mass Spectrometry (SIMS)

Secondary Ion Mass Spectrometry is an analytical method that allows to analyze the composition of solids, thin films or detect very small concentrations of dopants and impurities. In this method, the surface of the investigated material is bombarded with positively charged



primary ions (mainly Ar^+, Xe^+, O^+, Cs^+). Secondary ions are ejected from the surface and directed by a magnetic field to a mass spectrometer where they are analyzed. Both area and depth analysis can be performed. SIMS requires a high vacuum with a pressure of less than 10^{-4} Pascal (10^{-6} mbar or torr). This is needed to avoid both collisions of secondary ions with gas ions and also surface contamination of the samples. Secondary ions are analyzed by a sector, quadrupole or time-of-flight analyzer. The TOF analyzer separates ions according to their kinetic energy. This requires pulsed generation of secondary ions using a pulsed source of primary ions or pulsed extraction of secondary ions. This is the only analyzer that allows all generated secondary ions to be detected at once. Detection of individual ions is performed similarly to GD-MS.



This method can be used to measure concentrations up to ppb scale, roughly also the concentration of elements in the matrix. Very often this method is used to determine the diffusion or implantation profiles. The method is used in two modes. Static SIMS is used to analyze surface monoatomic layers, usually with a pulse source and time-of-flight mass spectrometer, while dynamic SIMS is used to analyze bulk layers using significant de-sputtering caused by primary ion beam. A magnetic sector or quadrupole mass analyzer is used as the analyzer.

- Dopant and impurity depth profiling
- Composition and impurity measurements of thin films (metal, dielectric, SiGe, III-V, and II-VI)
- Ultra-high depth resolution profiling of shallow implants and ultra-thin films (ULE implants and gate oxides)
- Bulk analysis, including B, C, O, and N in Si
- High-precision matching of process tools (ion implanters)

4.3. Nondestructive methods of the surface chemical analysis

Scanning Probe Microscopy, SPM

SPM – Set of methods designed for determining the surface structure with the resolution at the atomic level. The best-known methods include S**canning tunneling microscopy** and **Atomic force microscopy.**

4.3.1. Scanning tunneling microscope (STM)

The principle of the **Scanning tunneling microscope** is using tunneling effect to measure the surface topography of solid substances. In the STM, we measure the electric current between a sharp tip and a conductive sample, even though both are separated by a

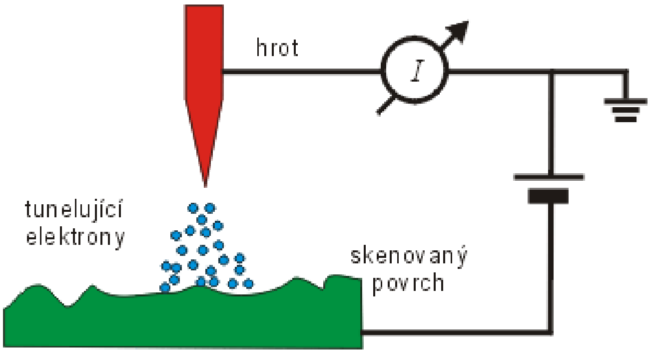


Fig. Schema of the tunneling microscope

vacuum barrier and are not in a direct contact. The number of electrons that are tunneled from the material to the tip is exponentially dependent on the distance, and a map of the surface can be drawn using the measured current. From these maps, we can then construct an image of arrangement of the individual atoms in investigated sample. **STM** was invented in 1981 by G. Binnig and R. Rohrer and both were awarded the Nobel Prize in 1986 for the discovery of the scanning tunneling microscope. STM can be compared to a gramophone, when the needle drawn above the surface changes its vertical position depending on the vinyl plate morphology.

The measurement of surface topography itself can be performed in two basic modes. In the constant height mode, when the tip is firmly fixed at a constant height above the surface and a change of tunneling current is recorded, or in the topographic mode, when the height of the tip above the sample is regulated so that the tunneling current remains constant for the whole time (feedback). Recording the xyz-position of the tip provides a direct picture about the atomic topography of the surface of the investigated material. The STM not only makes it possible to display individual atoms on the solid substance’s surface, but also allows tracking of their thermal movement (diffusion) over time.

Using STM, it is possible to obtain spectroscopic information about the local electron structure (i.e., to map electron states for different energies) at a place on the surface, where such information cannot be obtained by other techniques (e.g., near defects).

STM becomes a key tool, among other areas, in studying the transport properties of atomic contacts, the transport of electrons through organic molecules or in sequencing of DNA in the field of molecular electronics. Despite its relatively simple design and operation, the wider use of STM is significantly limited. Because tunneling current has to be measured, only electrically conductive materials can be examined, which greatly limits the scope of use, especially for biological materials.

4.3.2. Atomic force microscope (AFM)

The principle of the **Atomic force microscope** is the same as STM except that, instead of current, we measure the force between the tip and the surface of solid sample. AFM scans the surface of the material using a tip attached to an elastic cantilever. The tip is attracted to electrostatic forces (attractive and repulsive forces between atoms) and van der Waals forces. The AFM microscope is not limited to conductive materials like the STM microscope. The ATM microscope was discovered by G. Binning, C. Quat and C. Gerber in 1986.

We have again two working configurations. In direct contact mode, the tip is slightly pressed into the sample and the cantilever bends due to the electrostatic and van der Walls forces in accordance with Hook’s law. During the measurement, the tip moves across the surface of the sample in a regular pattern (scanning) such that the height of the second end of the cantilever is fixed. If the surface of the sample is not smooth, the cantilever bending is different at different places of the sample, and by monitoring the relationship between the bend and the location on the sample, we are able to construct a magnified image of the sample. The bending of the cantilever is monitored using a laser beam.

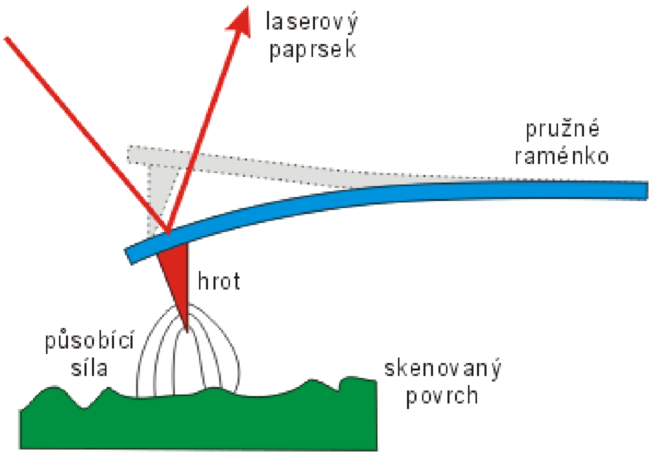


Fig. Schema of the Atomic force microscope.

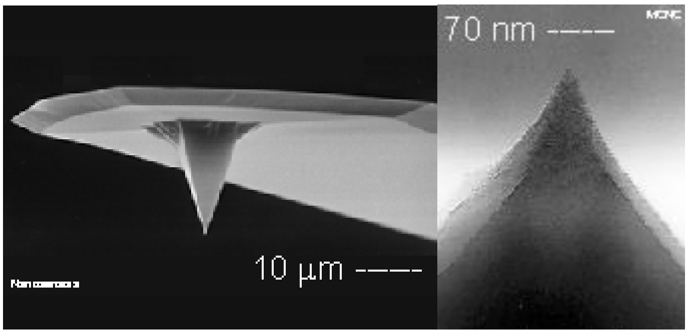


Fig. Cantilever.

A direct contact measurement, however, often results in the tip damage, if the sample surface roughness is too high. Therefore, an indirect contact mode, so-called constant bending mode, using a feedback is more often used. In this mode, the current value of the bending is compared to the pre-set value at each point of the raster (scan), and if these are different, the cantilever with the tip gets closer or further from the sample by such distance (z) so that the bending value again matches the pre-set value. Instead of the size of the bending, the z values are used to collect the image. At the same time, a constant bending value means that a constant force is applied to the sample. This mode may also be used for more rough samples, however, the image collection takes more time.

The detection of the cantilever bending is most often performed by a laser. The laser beam from a laser diode impact the cantilever, from which it is reflected onto the photodetector. If the cantilever bending changes, the angle of impacted and refracted beam also changes and therefore the beam hits another point of the photodetector. If the photodetector is position sensitive to the impact point of the beam, the cantilever bending can be determined from its output signal.

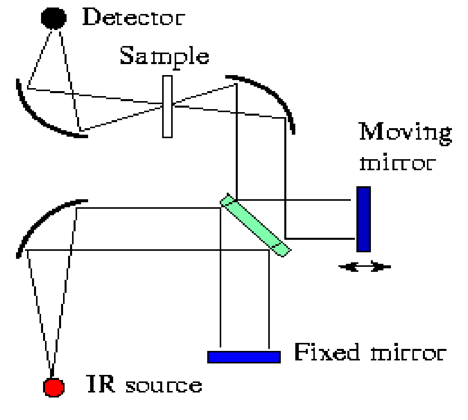
Both mentioned so-called *contact* regimes, however, may lead to damage to the sample, because there is a large friction force between the tip and the sample during the move from one point to another. For this reason, the so-called *contactless (dynamic)* mode is used, where no direct mechanical contact occurs between the tip and the sample, and the tip and the sample interact each other through van der Waals forces. The arm with the tip is introduced into an oscillating motion at a certain frequency and amplitude, and the amplitude size is measured instead of the arm’s bending. Because the amplitude depends on the distance between the tip and the sample, it is possible to build an image of the sample surface by monitoring the amplitude changes.

The accuracy of the AFM is dependent on the accuracy of the tip position maintaining, the accuracy of the tip’s movement and the ability to bending detection. For the movement of the tip, piezoelectric scanners are exclusively used, which are able to execute movements smaller than one tenth of a nanometers. In order to maintain the exact position of the tip, AFM microscopes are built very strong mechanically and are usually placed on anti-vibration tables.

New scanning microscopes already use both STM modes and dynamic AFM, where both the tunneling current and atomic forces can be measured at the same time. This approach offers a unique opportunity for a complex analysis of surfaces and nanostructures in the atomic scale. It also makes it possible to study the fundamental questions of quantum mechanics, i.e., the mutual dependence of the tunneling current and atomic forces, where both quantities are a function of the overlap of wave functions between the tip and the surface.

4.3.3. Fourier Transform Infrared Spectroscopy (FTIR)

The method is based on the absorption of infrared radiation passing through the sample, during which generates transitions involving changes in both vibrational and rotational energetic states of the molecule depending on the changes of bipolar moment of the molecule. When such transitions emit or absorb photons (electromagnetic radiation), the frequency is proportional to the difference in energy levels and can be detected by certain kinds of spectroscopy. The resulting infrared spectrum is a functional dependence of energy, usually expressed in percentage of transmittance or in units of absorbance on the wavelength of the incident radiation.

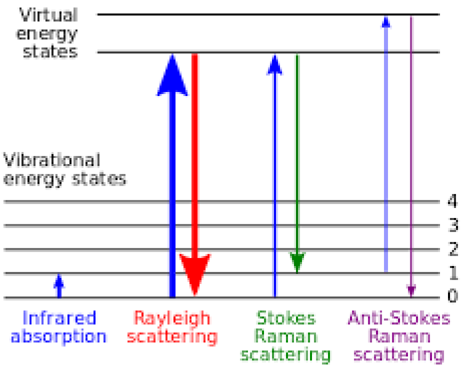


Schema of the Michelson interferometer

FTIR uses the Michelson interferometer, which contains a “white” or infrared source, a beam splitter, a fixed and moving mirror, and a detector. The spectrum is measured depending on the position of the moving mirror. Typically, both spectra with and without the sample are divided and the resulting intensity spectrum is converted to frequency dependence using Fourier transformation. The absorption and thus the reflection of the sample can be measured by the same process. FTIR provides information on chemical bonds and molecular structures of both inorganic and organic compounds. The bonded atoms oscillate with a vibration frequency characteristic for the given substance. When the material is irradiated by infrared radiation, the atoms absorb that radiation at frequencies which resonate with the vibrations of atoms. Measuring the frequency of absorption can therefore provide information on the chemical composition of the material, or its admixture.

4.3.4. Raman Spectroscopy (Raman)

Raman spectroscopy uses the **Raman effect** also known as **Raman scattering**, which occurs when photons of incident light interact with vibrational and rotational energetic states of atoms or molecules (scattering on optical photons), when the scattered radiation has a different wavelength (photon energy) than the incident light. A laser beam is used for excitation and a CCD camera is used to detect emitted photons. When the sample is radiated with a laser beam, most of the radiation passes through, some of the photons are absorbed and a small portion (10^{-4}) is elastically scattered. The laser beam excites the electron from a basic state into a virtual state, and when recombine back to the basic state, a photon with the same wavelength as the original photon is radiated. This effect is so-called *Rayleigh scattering*, which carries no analytical information at all. Only 10^{-8} of photons are non-elastically scattered and interact with the sample (Stokes and-Stokes lines) – Raman spectroscopy uses this non-elastic scattering.



1. If an electron recombines back, after the excitation to a virtual state, to a higher quantum level than to the one from which it was kicked-out, a photon with a greater wavelength is emitted – the so-called Stokes photons.

2. On the other hand, if the electron was not originally in the basic state, but at a higher level and recombines to the basic level, a photon with a smaller wavelength is emitted – the so-called anti-Stokes photons.

The shifts of frequencies in Stokes and anti-Stokes photons from the frequency of the used laser source then carry analytical information about the differences of individual quantum levels (most often vibration levels).

Raman spectroscopy is a largely a supplementary method to infrared spectroscopy. It is almost a rule that transitions that are intense in the Raman spectra are weak in the infrared spectra and vice versa.

Raman spectroscopy is widely used for qualitative analysis of both organic and inorganic compounds. It makes it possible to determine the chemical composition of samples and identify additives present in the sample.

- Identification of organic molecules, polymers, biomolecules, and inorganic compounds in elementary or volume structure (identification of minerals in mineralogy).
- Determination of the relative concentration of individual chemical compounds in mixtures such as medicaments, drugs but also colours, inks or organic fibers – forensic chemistry.
- Determination of presence of different types of carbon (diamond, graphite, amorphous graphite, graphite nanotubes) and their relative quantity in compounds.
- Determination of inorganic oxides and their charge states.
- Measurement of the stress and crystalline structure in semi-conductors and other compounds.

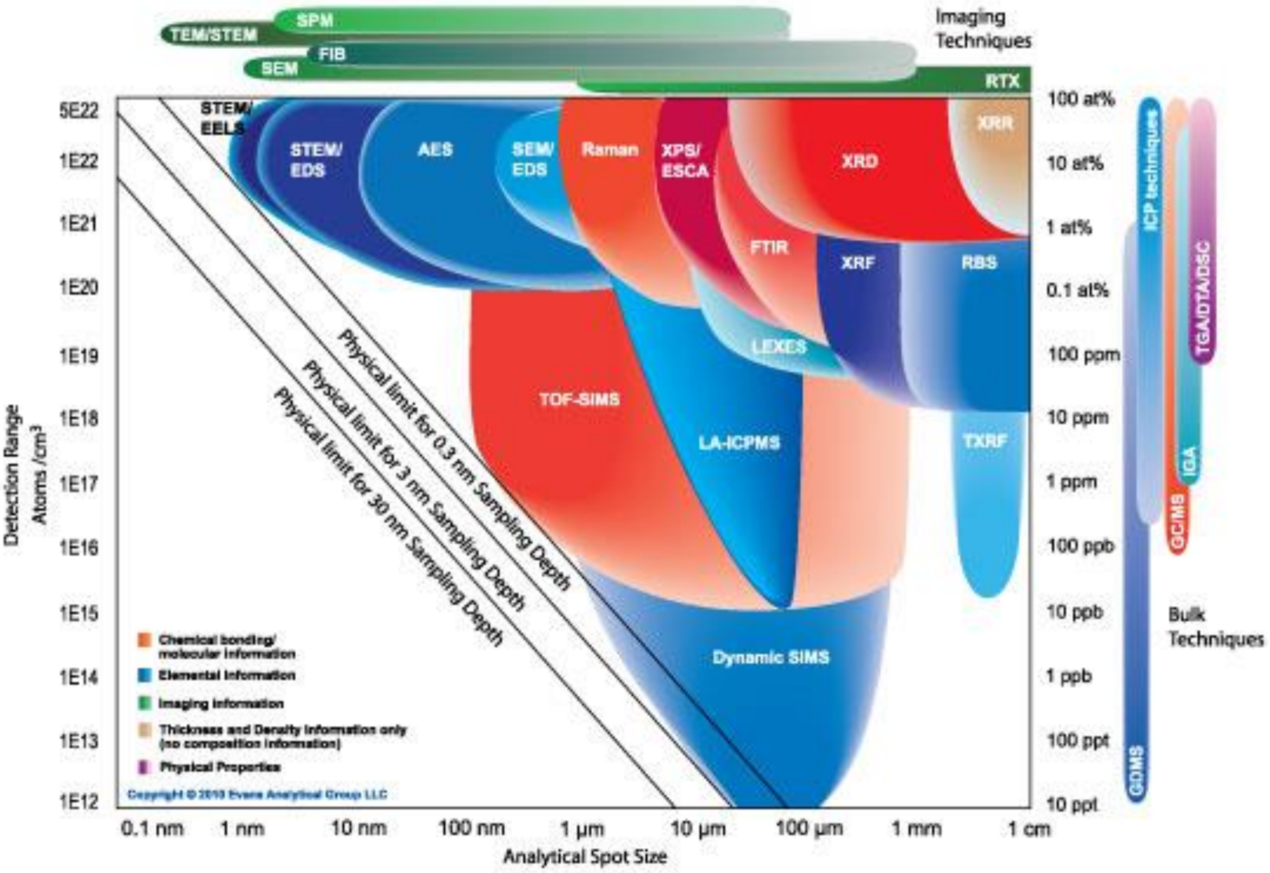


Fig. Analytical Resolution versus Detection Limit

5. Vacuum technology

For pressure measurement, the unit **1 Pascal=1 Nm^{-2}**, [Pa], is used in physics according to the SI system (Système international, International system of Units). In vacuum practice, multiples of Pascal unit are historically also used (1 hPa = 100 Pa = 1 mbar) and other units like **bar**, t**orr**, **atm**, **psi**, **at, mm Hg, mm H_20** and their multiples.

In technical practice, the positive convention is mainly used, with a pressure of 0 (no matter the unit) as the lowest pressure level. The zero level refers to the so-called *perfect vacuum*, a technically and physically unattainable pressure value of 0. However, sometimes a negative convention is used, where the standard value of atmospheric pressure is set at zero and the vacuum (vacuum) is then calculated as the difference (negative value, e.g. -100 hPa). There is also a way of labelling as % of vacuum (e.g. 90% of vacuum means a pressure of 10% of the atmospheric pressure value = 101,325 hPa).

Measuring pressure units – transposing

Unit	1 Pa	1 hPa=mbar	1 torrr	1 atm	1 bar	1 psi	1 at
1 Pa	= 1 Nm^{-2}	100	$7{,}5006\cdot10^{-3}$	$9{,}8692\cdot10^{-6}$	$1\cdot10^{-6}$	$1{,}4504\cdot10^{-4}$	$1{,}0197\cdot10^{-5}$
1 hPa = mbar	100	1	$7{,}5006\cdot10^{-5}$	$9{,}8692\cdot10^{-8}$	$1\cdot10^{-8}$	$1{,}4504\cdot10^{-6}$	$1{,}0197\cdot10^{-7}$
1 torr = 1 mm Hg	133,322	1,33322	1	$1{,}3158\cdot10^{-3}$	$1{,}3332\cdot10^{-3}$	$1{,}9337\cdot10^{-2}$	$1{,}3597\cdot10^{-3}$
1 atm	101 325	1013,25	760	1	1,01325	14,69595	1,033227
1 bar	100 000	1000	750,06	0,98692	1	14,5038	1,0197
1 psi	6 894,76	68,9476	51,715	0,068046	0,068948	1	0,070306
1 at = 1 $kpcm^{-2}$	98 066,5	980,665	735,58	0,9678	0,98067	14,224	1

Notes: 1 Torr is a hydrostatic pressure of 1 mm Hg columns on the sea level at 0°C; ATM is a physical atmosphere and is based on standard atmospheric air pressure; PSI (from English "Pound per Square Inch") corresponds to pressure g.lb.in^{-2} on sea level; **kp.cm^{-2}** is kilopond per cm square, and so-called technical atmosphere (AT) corresponds to pressure **g.kg.cm^{-2}** on sea level.

Pressures can be indicatively classified according to their size, but these ranges have only an informative character. This classification allows faster orientation in the availability and applicability of technological processes for generating and measuring pressure.

1) Perfect vacuum 0 Pa (absolutely empty space, it's more of a hypothetical situation)
2) Extremely high vacuum < 10^{-10} Pa
3) Ultra-high vacuum 10^{-10} – 10^{-7} Pa
4) High vacuum 10^{-7} – 10^{-2} Pa
5) Low vacuum 10^{-1} – 10^{2} Pa
6) Normal atmospheric pressure $^{\#}$ = $1{,}01325\cdot10^{5}$ Pa = 1013,25 hPa = 760 torr
7) Increased pressure, up to ~ 3.10^{6} Pa
8) High pressure, up to ~ 4.10^{7} Pa
9) Very high pressure, up to ~ 5.10^{8} Pa
10) Ultra-high pressure, above ~ 5.10^{8} Pa

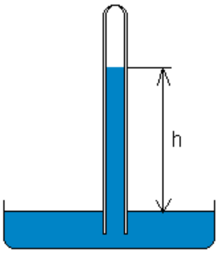
Normal atmospheric pressure is defined as the average value of air pressure at sea level at 45° norths at 15°C and weight acceleration g_n = 9,80665 ms^{-2}.

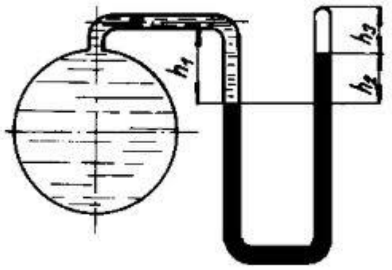
The highest vacuum achieved in the Earth laboratory was close to 10^{-11} Pa, the vacuum in space reaches 10^{-15} Pa and less - here it is necessary to take into account the number of particles found in a certain volume. The pressure on the surface of the moon, which has no own atmosphere, but is constantly bombarded by particles contained in the so-called "solar wind", is about 10^{-9} Pa, which represents the number of 4.10^5 particles in 1 cm^3. There are approximately 10 particles in interplanetary space, in interstellar space 1 of the particle and even 10^{-6} particles in 1 cm^3 in the intergalactic space. So, the pressure here is close to the perfect vacuum. In contrast, the highest artificially achieved pressure by the static method, i.e. long-term force action on the sample, was close to the value of 5.10^9 Pa. Dynamic methods based on shock waves in explosion can achieved pressures of an order of magnitude higher, but only for a very short time.

5.1. Principles of vacuum measurement (under pressure)

5.1.1. Absolute vacuum meters

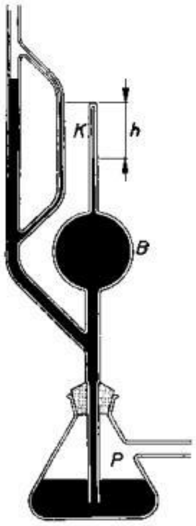
5.1.1.1. Torricelli tube, U-Tube





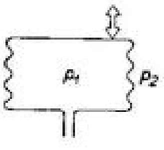
Tube filled by mercury is placed inverted on the dish full of mercury. This causes the mercury in the tube to fall down until the difference between mercury on the surface and in the tube is about 760 mm (Torricelli experiment for proving of the atmospheric pressure).

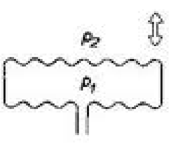
5.1.1.2. McLeod compress vacuum gauge

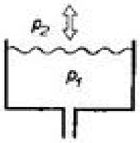


It is a scientific instrument used to measure very low pressures in the range of 10^{-3} Pa do 10^5 Pa. It serves as a pressure (vacuum) standard with calibration uncertainty 0.5% for pressure greater than 10^{-1} Pa and about 3% at pressures lower than 10^{-2} Pa. The principal is based on the Boyl Law, where the known amount of gas is enclosed above the mercury cap, which is compressed (expanded) by the hydrostatic pressure of the mercury column and the pressure in the measured space.

5.1.2. Membrane vacuum gauges







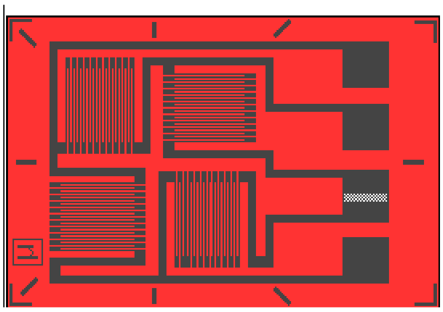
These manometers use the elastic properties of the membranes, which stretch and relax depending on the pressure difference outside and inside the space they separate. The membranes may also have the shape of a wavy surface tube, the so-called bellows, the length of which varies depending on the pressure difference inside and outside the tube. They can measure the pressure up to about 10^{-2} Pa. A mechanical position reading device is attached to the membrane, or the position is read using a laser beam.

5.1.3. Piezoelectric tenzometers

The piezoelectric crystal is a major component of active electrical pressure transducers. Crystals lacking the center of symmetry can produce surface potential if they are exposed to the deformation in a specific direction. Quartz (SiO_2), Seignette salt (Potassium sodium tartrate - $KNaC_4H_4O_6.4H_2O$), barium titanate ($BaTiO_3$) and others ($PbZrO_3$, $LiNbO_3$, $LiTaO_3$ …) are the most common crystals that exhibit a piezoelectric phenomenon. The sensor itself is made from a crystal cutlet that is provided with conductive electrodes. When pressure is applied, the piezoelectric sensor acts like a charge generator. The sensor output voltage is relatively high, yet it is not possible to measure it directly. A large input resistance amplifier must be included between the measuring instrument and the sensor for impedance adjustment, less of the signal amplification. Piezoelectric converters can be used to measure very rapidly changing pressures. Another application of the piezoelectric effect is a quartz crystal resonator having a unique frequency that can be changed applying pressure on the resonating crystal. Combining with the reference resonator (to compensate the ambient temperature fluctuations), piezoelectric tenzometers form a very precise pressure gauge. They are used to measure high pressures (up to 70 MPa), but can also be adjusted for vacuum measurement. They are used as a secondary standard for its accuracy. The disadvantage is the high price of complete device.

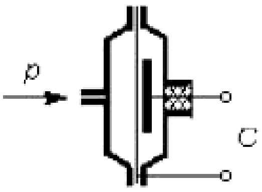
5.1.4. Resistance strain gauge

Strain gauges (sometimes called variable resistance sensors) are the most commonly used electrical pressure transducers. A resistance strain gauge is a sensor that uses the so-called piezoresistive effect. A strain gauge indicates the strain of material or structure at the point where it is installed. When a material is deformed the strain gauge varies its resistance. This change can be used to determine the force, weight, pressure, tension, compression, torsion or torque. During mechanical stress in the area of elastic deformations, metals or semiconductors undergo changes in the crystal lattice that lead to a change in resistance. Like piezoelectric sensors, these devices are mainly used to measure rapidly changing pressures. Metallic resistance strain gauges are made from resistance wires with a diameter of 0.02 to 0.05 mm, which are glued to a thin paper or plastic substrate. They are shaped like a multiple meanders. The whole strain gauge sensor is glued to the measured place with a special glue. Tensometric sensors of complex shapes (radial, spiral, etc.) are now produced by patterning (etching out) a thin resistance foil (same procedure as the production of printed circuit boards). Semiconductor strain gauges are made from silicon, either just using shaped single crystal or by planar technology on a silicon or on other semiconductor substrate. The mechanical stress in a particular crystallographic axis of the single crystal or in the diffused layer of the semiconductor cause a change of the electrical conductivity. The change in resistance depends on the type of semiconductor and the concentration of impurities. The active part, the semiconductor strip, is glued with a suitable adhesive to the surface subjected to deformation.



5.1.5. Capacitive strain sensors

The basis of a capacitive sensor is a two or several electrode system. One of the sensor electrode is fixed and the other is made up a membrane that changes its position due to the action of measured pressure. The distance



of these electrodes then determines the capacitance of the capacitor. The capacity of the measuring element is further electronically processed and converted to a pressure value. They are commonly used for measuring pressures in the range of 10^{-1} Pa – 10^{6} Pa with an accuracy of 0,08 % – 0,25 %, but there are also capacitors for measuring pressures up to 10^{-3} Pa.
The linearity and repeatability of capacitive sensors is better than that of resistive sensor during a long period, on the other hand the resistive sensors have higher sensitivity.

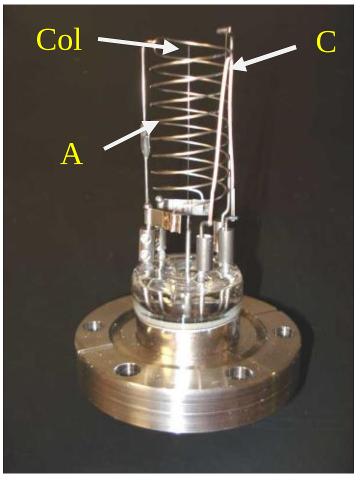
5.1.6. Thermal conductivity vacuum gauges

These gauges use the dependence of the thermal conductivity of the gas on the pressure, resulting in the cooling of a suitable metal filament with gas in the measured chamber.
The change of the metal filament temperature is converted into:
- Change in electrical resistance of filament - resistance vacuum gauge (Pirani).
- Thermoelectric voltage change – thermocouple vacuum gauge.
- Change in the shape of a suitable element (bimetallic, expansion vacuum gauge).

Pirani vacuum gauges are based on the principle of measuring the electrical resistance or temperature of a metal filament heated by a constant electric current and cooled by the surrounding gas. A heated filament suspended in a gas will lose heat to the gas as its molecules collide with the filament and remove heat. If the gas pressure is reduced the number of molecules present will fall proportionality and the filament temperature will decrease more slowly. Heat transfer depends on the density of the gas and therefore on its pressure. The metal filament has to have a strong temperature dependence of resistance - platinum is most often used. These vacuum gauges allow the measurement of pressures from 10^{-2} Pa do 10^{5} Pa (accuracy increases with decreasing pressure). These gauges may differ in design and the way in which the temperature change in the sensor is indicated. Some, for example, consist of two surfaces (filaments, strips) with different temperatures, between which heat is transferred. Thermoelectric sensors contain electrically heated metal filament and a thermocouple. The filament temperature as a function of pressure is measured by a thermocouple. The temperature depends on the type of gas being measured. Thermistor gauges often use semiconductor thermistors instead of a thermocouple to measure filament temperature. Pirani gauges are very common and relatively cheap. They are simple, reliable, but not very sensitive. However, they have all the disadvantages of thermal vacuum gauges – relatively short pressure and frequency range.

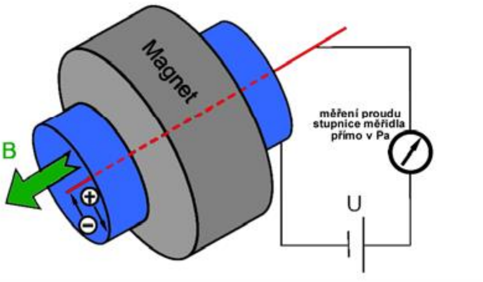
5.1.7. Ionization vacuum gauges

The principle of an ionization gauge is the measurement of the electric current between the electrodes, which are placed in a chamber with the measured pressure. The electric current between the electrodes is realized by the movement of electrically charged ions, which are generated in the vacuum gauge by the ionization (collisions of gas molecules with electrons emitted from the cathode (ionization vacuum gauges with hot or cold cathode) or by the effect of α-radiation from the alpha source (in so-called alphatrons).

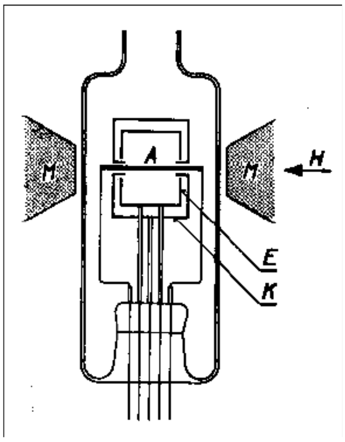


a) **Hot-cathode ionization gauge** (Bayard-Alpert gauge) consists of a cathode filament (C) heated by electric current, an anode helical grid (A) and a central ion wire collector (Col). The construction is similar like a triode. The hot cathode acts as a source of electrons that ionize the residual gas. The amount of generated charged ions is proportional to the density of the gas and thus its pressure. The anode - cathode bias is +150 V to +200 V. The collector to cathode bias is –25 V to -50 V. The gas pressure is determined from the ion current measured at the collector. Ionization is linearly dependent on pressure. The gauges are used for measuring pressures 10^{-1} Pa – 10^{-6} Pa (after special modifications up to $\underline{10^{-9}\ \mathrm{Pa}}$). Air must not enter the apparatus during the measurement, as the hot cathode filament would burn immediately. Moreover the filament material gradually evaporates and the vacuum gauge must be replaced over time. The gauges have relatively low sensitivity because the number of ionized particles is small.

b) **Cold-cathode ionization gauge** (Penning gauge or inverted magnetron). Both they use a glow discharge of ionized gas molecules generated by collisions with electrons accelerated in an electric field applied by a 4 kV DC voltage. They have no heated cathode. To increase the probability of electron collisions with highly dilute gas molecules, the electrodes are placed in a strong magnetic field that forces the electrons to circle in spiral paths due to the Lorenz force. In a standard Penning gauge (a cylinder with a diameter of about 2 cm and a length of 10 cm), the electron trajectory is several tens of meters. On the other hand, due to a relatively large mass of generated ions, their trajectory is almost unaffected by the magnetic field. They directly drift toward the anode and thus create a measurable current that depends on the gas pressure. The gas pressure is determined from the electric current passing through the collector. The Penning vacuum gauge is suitable for measuring pressures (discharge must ignited) $\underline{10^{-1}\ \mathrm{Pa} – 10^{-7}\ \mathrm{Pa}}$.



The inverse magnetron is used to measure a vacuum up to $\underline{10^{-11}\ \mathrm{Pa}}$. The difference in the design of the Penning gauge and the inverse magnetron is mainly in the mutual arrangement of the electrodes. In addition, the inverse magnetron contains a shielding electrode, which prevent photoelectrons emitted from the cathode (soft X-rays that are generated during the discharge) and thus significantly increases the sensitivity of the gauge. The penetration of air into the cold-cathode ionization gauges during the measurement has no destructive effect on them, as in the case of hot-cathode ionization gauges, and its functionality practically does not change over time.



Memory effect of the Penning ionization gauges

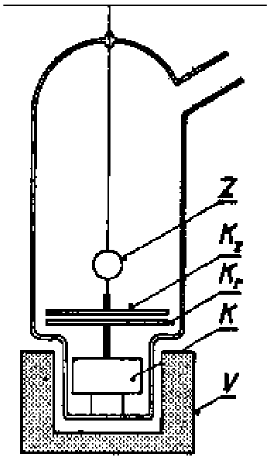
- The glow discharge ignites, the glow discharge ignites, the cathode is desputtering and desputtered metal absorbs the gas, which is released again at a lower pressure - the vacuum gauge shows a higher pressure. The memory effect can be reduced by reducing the anode volume.

Modern construction

- A design, where magnetic and electric fields are perfectly perpendicular to each other
- High voltage is less than 2 kV.

5.1.8. Viscosity manometers *(Langmuir – Dushman molecular gauge)*

These gauges are based on the fact that at low pressures, the viscous drag between two surfaces in relative motion is proportional to P/M, where P is the pressure of the gas and M is its molecular weight. The lower disk Kr rotates and the gas molecules transmit the impulse to the upper disk with the mirror Kz. The pressure is determined from the angle of rotation of the mirror. The angle depends on the molecular weight of the gas and the viscosity. Viscosity gauges are resistant to aggressive gases. Their measuring range $\underline{1\ \mathrm{Pa}\ \text{to}\ 10^{-5}\ \mathrm{Pa}}$ depends on their design.



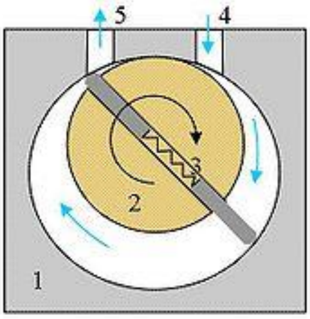
5.1.9. Mass spectrometers

The principle is the same as in mass spectroscopy. In the ionization chamber, residual-gas particles are bombarded with electrons to generate charged ions. These move along a curved path to the ion mass quadrupole analyzer. The radius of curvature depends on the mass and energy of the ions (m/z ratio). The intensity of the peaks that are characteristic of a particular substance is measured. It is possible to measure the partial pressures of residual gases.

5.2. Principles of vacuum generation

Vacuum pumps, of various types and designs, are used to generate a vacuum. A vacuum pump is a device that draws air or other gases from an enclosed space and thus creates a partial vacuum. We can divide vacuum pumps according to both the physical principle of operation and the degree of vacuum they can achieve. The most well-known types include oil rotary, piston, screw, turbo-molecular pumps and, to achieve high vacuum, ionic, cryogenic, sorption, sublimation and so-called getter pumps.

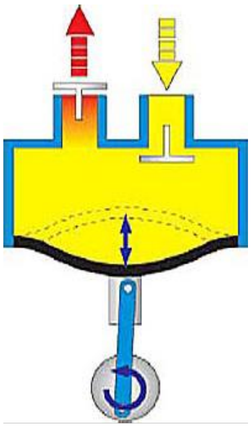
5.2.1. Rotary vane vacuum pump



Rotary pumps are used to achieve low vacuum and as the first stage of high and ultra-high vacuum systems. The design consists of a cylindrical rotor (2), which is mounted eccentrically inside a cylindrical housing or stator (1). The one side of the rotor is in close contact with the wall of the stator. Two blades (also known as ‘vanes’), mounted inside the rotor and connected by the spring, move in and out due to centrifugal force following the internal surface of the stator. The pump draws gas from the right inlet port (4), compresses and put away through the left outlet port (5). The whole system is immersed in oil. With good adjustment, it can reach a vacuum up to a pressure of 10^{-1} Pa. A major drawback of oil rotary pumps is the possibility of oil entering the pumped vessel or area.

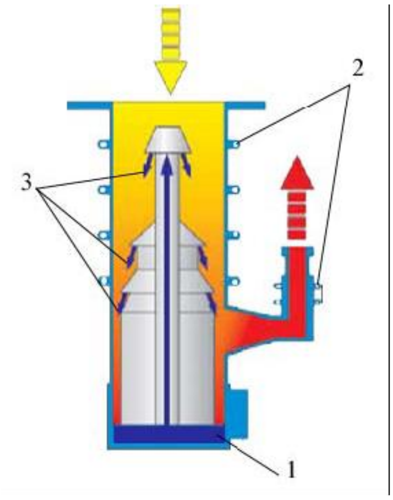
5.2.2. Diaphragm pump

The mechanical piston pump works on the principle of cyclic volume expansion formed by moving the piston in a tight cylinder (Sprengler pump). The piston can also be formed by a mercury column, such as in a Toepler pump. Piston pumps with this design have low efficiency, due to their relatively large dead (inexhaustible) volume and are currently barely used. A modern variant of a piston pump is a diaphragm pump that works on an air displacement principle using inlet and outlet valves. There is no contact of the pumped environment with the mechanical system of the pumps, as they are separated from each other by a diaphragm, which is mechanically pushed and pulled in and out of a pumping vessel. Pumps are therefore suitable for pumping aggressive substances, provided, of course, that the diaphragm and valves are made of chemically resistant materials, such as Teflon. These pumps are then universally applicable. Diaphragm pumps have a number of advantages. They are gas-tight, with a 220V power supply, oil-free and require virtually no maintenance. The construction parts that come into contact with the pumped medium are usually passive, so that relatively little heat is generated during the operation of the pump, which facilitates the possible condensation of the vapors of the exhaust substances at the outlet. However, they have a lower limit of achievable vacuum than rotary pumps of the order of 10^{2} Pa. The limit pressure is limited by the pressure at which the inlet valve can be still opened. Drawback - when the diaphragm collapsed all of the air is forced out of the diaphragm vessel.



5.2.4. Diffusion pump

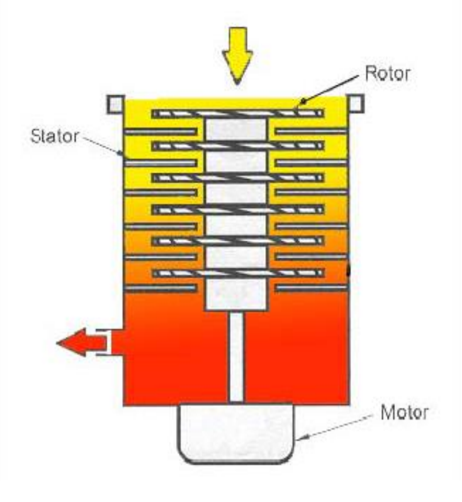
The diffusion pump uses the vapors of a boiling fluid (oil, mercury) to capture air molecules. The vapor is then moved to another location where cooled and condensate. The air molecules are released back during cooling. The combination of gravity and the downward direction of the vapors move the air molecules toward the bottom of the pump. The oil diffusion pump has no moving parts and works by sharing the impulse between the pumped gas molecules and the fast-moving stream of oil vapor (supersonic speed). However, in the higher vacuum, only the drag of gas molecules by a fast-flowing liquid or steam works. The achievable vacuum is limited by the surface tension of the working liquid or vapor, which is why mercury was used in the past, today mostly oil vapor.



The diffusion pump consists of a boiling section (1), where the working fluid (oil or mercury) is heated to boil by an electric current and its vapors are jet by a nozzle into the working place of the pump (3). The flow direction of these vapors is identical to that of the pumped gas. The pumped gas diffuses into the working fluid vapor and this mixture is then guided to an area where only the working fluid condenses (and flows back into the boiling section) and the pumped gas is drained away from the system by a primary vacuum pump. For proper operation, the pump housing, on its inside wall the condensed working fluid flows, must be efficiently cooled, usually with water (2). The limit pressure of oil diffusion pumps ranges from 10^{-2} Pa to 10^{-7} Pa (depending on the quality and type of oil). The diffusion pump requires a pre-vacuum in the order of units up to tens of Pa (i.e. their output is connected to the input of the primary pump, usually a rotary oil pump).

The pumping speed of diffusion pumps is relatively low and in order to reach the limit pressure in the evacuated area, it is necessary to pump the gas for a very long time (even tens of hours). The pumping speed is higher for diffusion pumps with a large pump body diameter. The advantage of diffusion pumps is the relatively low price, design simplicity, durability, and last but not least, the ability to pump any gases. The drawback is the possible contamination of the vacuum by the vapors of the working fluid, and therefore it is necessary to include freezing fingers and oil traps between the evacuated space and the diffusion pump. They also have a relatively long rise time (it is necessary to heat the oil to a working temperature of approx. 200 ° C before starting) and a long standby time. They must always be placed vertically and it is necessary to prevent air from entering the pump during pumping, as there is a risk of degradation of the working fluid at high temperatures by the interaction with oxygen.

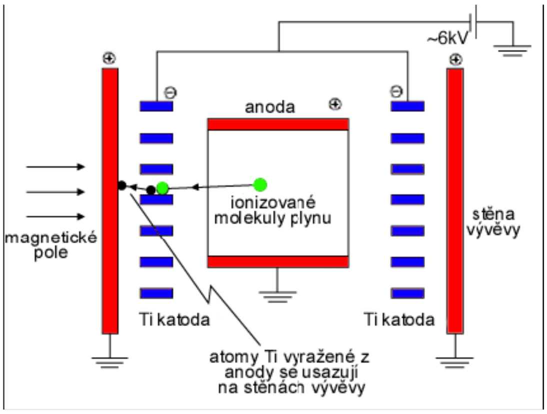
5.2.5. Turbomolecular pump



Turbomolecular pump consists of a quickly rotating rotor blade and stationary stator blade pair. The principal is similar in design to a turbine or fan, where residual gas molecules are reflected in the pumping direction by means of a rotor with blades. The rotation rate of this pump is very high (10-90 thousand revolutions per minute) and requires a perfectly balanced rotor and a shaft bearing in magnetic levitation. The diagram of this pump is shown in the figure. This type of pump is difficult to manufacture and therefore expensive, but it can pump from a very low vacuum up to a pressure of 10^{-8} Pa (vacuum is very clean). The turbomolecular pump cannot operate at atmospheric pressure because the rotor blades would heat up and destroy with strong friction against the pumped gas, so it is used in combination with a primary pump (e.g. oil rotary or diaphragm). The turbomolecular pumps are widespread and are now part of a wide range of spectroscopic devices, electron microscopes and particle accelerators.

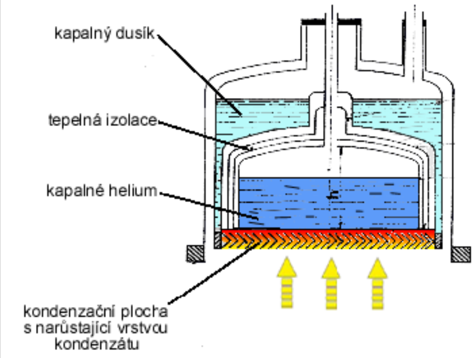
5.2.5. Titanium ion-sorption pump

Ion pumps combine the principle of momentum pumps (gas molecules are given an impulse in the direction of pumping charged particles) with the principle of adsorption pumps (gas molecules are trapped in the vacuum pump-gettering). The pump consists of a reactive surface of a suitable substance that captures residual gas molecules to itself. Chemisorption of molecules on the active metal layer is used. The metal is either evaporated or cathode-sputtered. The scheme shows the function of an ion pump consisting of two cathodes (e.g. made of titanium) and a tubular anode. A bias of several kV is applied to the anode, so that a glow discharge is created, which is the source of the electrons, like in an ionization gauge. The discharge takes place in the magnetic field of the permanent magnet, which is perpendicular to the plane of the electrodes, and therefore the electrons move in long helical paths, which increases the probability of their collision with the molecules of the pumped gas. During the collision, the molecules are ionized and the formed ions are entrained towards the cathodes, where they are adsorbed on the titanium surface. The efficiency of the pump is further increased by the sputtering of titanium and captured on other parts of the pump, where it binds the residual pumped molecules. The ion pump works in cycles, it cannot operate continuously. Its cycle has 3 phases: sorption, desorption and regeneration. In the sorption phase gas molecules are either condense or be adsorbed by the large reactive surface of the titanium. In the desorption phase the pump is allowed warm up to room temperature and the gases escape through the pressure relief valve or other opening to the atmosphere. The pump can be used in a cycle of sorption and desorption until it loses too much efficiency and is regenerated or in a cycle where sorption and desorption are always followed by regeneration. In the regeneration phase the pump body is heated to drive gas molecules that does not desorb at room temperature and accumulates in the titanium. A pressure of 10^{-9} Pa to 10^{-10} Pa. can be achieved by this pump. The main advantages are the absence of oil or other contaminants, low cost and vibration free operation because there are no moving parts. The drawbacks are high electrical input, high weight, lower pumping speed, the need for a strong permanent magnet and the need to renew the electrodes. These pumps are used to obtain ultra-high vacuum in special instruments, such as mass spectrometers, accelerators, microwave generators or electron microscopes. They are usually classified as the third stage after a rotary or diaphragm pump and a turbomolecular pump.



5.2.5. Cryopumps

Cryopump traps gases and vapours by condensing them on a cold surface, but are only effective on some gases. The effectiveness depends on the freezing and boiling points of the gas relative to the cryopump's temperature. A cryopump usually consists of two insulated vessels containing liquid nitrogen and liquid helium. The outer vessel is filled with liquid nitrogen and thermally insulates the inner vessel with expensive liquid helium, which is connected to a condensation smooth surface extending directly into the evacuated space. Residual gas molecules condense on this cold surface and the pressure in the pump drops. As a layer of solid condensate, which has poor thermal conductivity, grows on the condensing surface, the efficiency of the pump gradually decreases during the pumping cycle. At the end, the pump must be stopped, heated and the condensation surface desorbed. To increase the pumping effect, instead of a smooth condensing surface, a material capable of absorbing gas in its cavities (zeolites or activated carbon) is used,

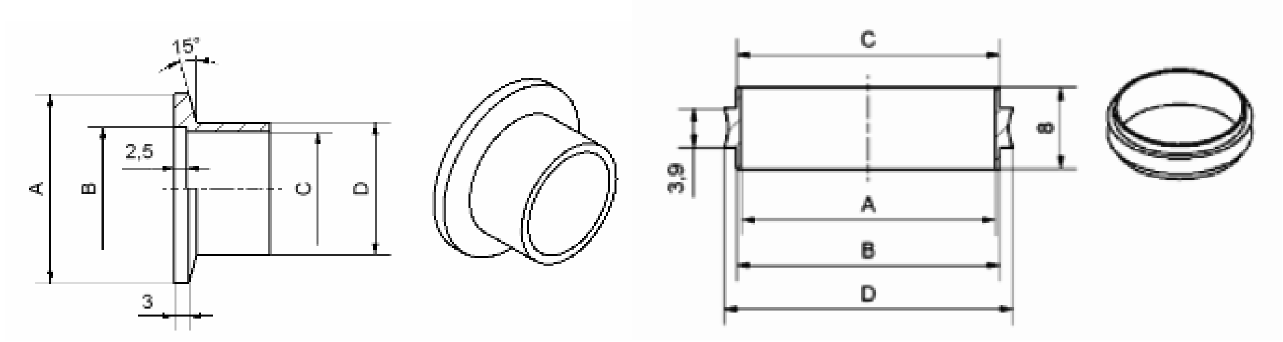


which is the principle of a cryosorption pump. These pumps are very simple in construction, have a high pumping speed, provide a clean vacuum and can reach pressures of 10^{-7} Pa (pumps cooled only by liquid nitrogen) and pressures lower than 10^{-10} Pa (pumps with liquid helium flow).

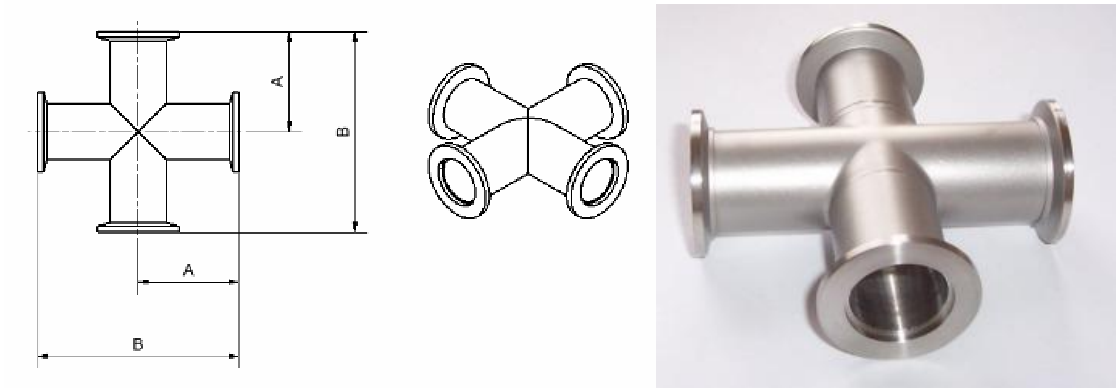
5.3. Vacuum clamps, fittings and flanges

5.3.1 Quick-release fastening elements according to ISO-KF DIN 28403.

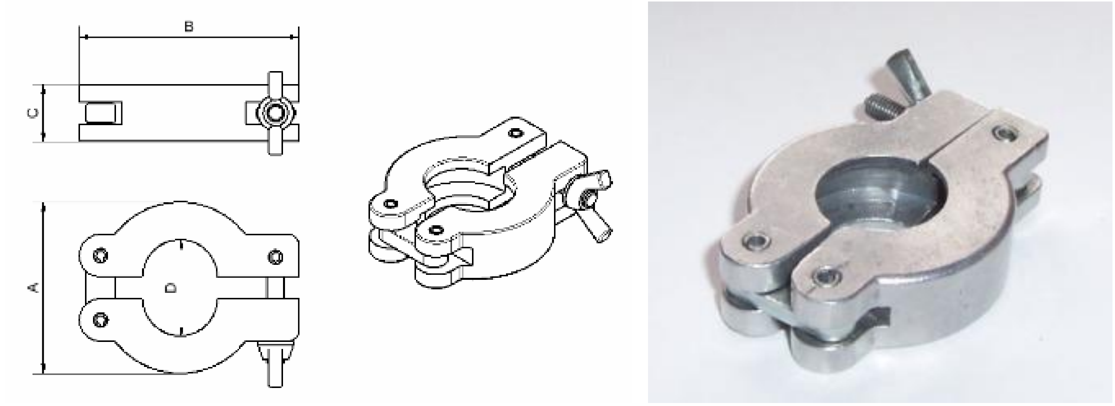
Size DN 10, 16, 25, 40, 50



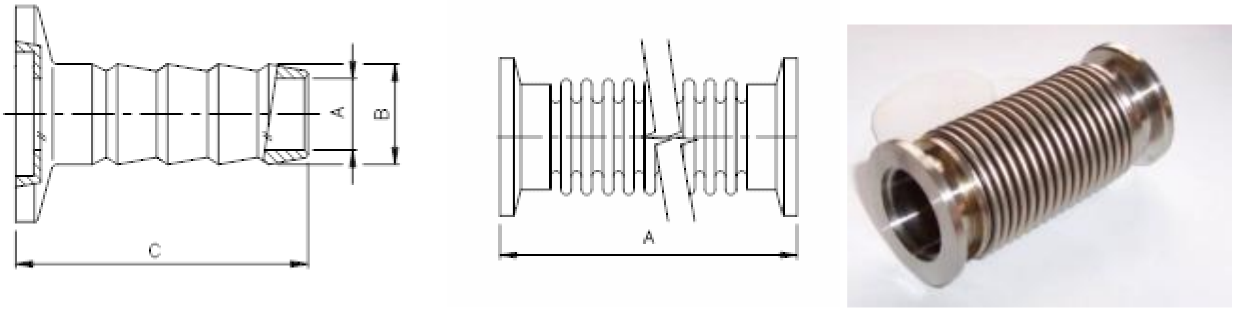
Flanges with tabulation, Centering Rings



4-way cross

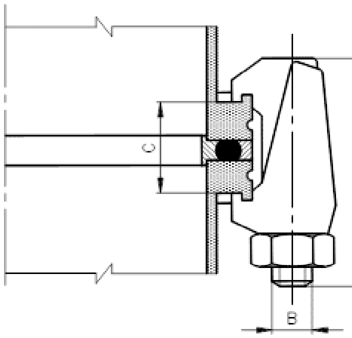


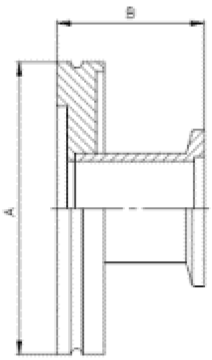
Clamping Ring



Flange with Hose nozzle, Flexible Bellows with Flanges

5.3.2. Vacuum clamps according to ISO-K DIN 28404






Adaptor/Reducer

5.3.3. Gaskets

So-called "O" rings. Special rubber called NBR (-30-90^{o}C) or Viton (-15-150^{o}C).

Indium, Copper, brass gaskets.

Feedthrough

Individual vacuum components

Flanges and Fittings

Bellows with Flanges

6. Applications

6.1. Energy-saving light sources

6.1.1. Photometry - Basic concepts and units

Luminous intensity is defines as a spatial density of luminous flux (power) per unit solid angle emitted by a point light source in a particular direction. The candela [cd] is the unit of luminous intensity and is one of the base units in the International System of Units (SI). Candela is defined as the luminous intensity in a given direction of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and has a radiant intensity in that same direction of 1/683 watt per steradian (called unit solid angle). A common wax candle emits light with a luminous intensity of roughly one candela. In fact, the luminous intensity is the light flux density that radiates at a given angle. If the luminous intensity value of the light sources is indicated, then it is usually also stated for which angle the given luminosity value is valid, or a full graph is available showing the luminous intensity values at all angles.

Luminous power or **luminous flux** [lumen (lm)] indicates to the luminous energy emitted by the source per unit time in all directions. The luminous flux value in Lumens expresses the light energy depending on the **sensitivity of the human eye** to the light of different temperatures (different light wavelengths). Thus one lumen is the luminous flux emitted within unit solid angle (one steradian) by a point source having a uniform luminous intensity of one candela, so that 1 lm = 1 cd sr, and the total flux in all directions is 4 π lm. Luminous flux is one of the important aspects when comparing replacements for conventional light bulbs. When assessing between each other LED lamps, this unit most closely matches the classic assessment of filament bulbs among themselves according to wattage power. In the near future, light bulbs will be judged not by wattage power, but by lumens. The usual values of the luminous flux of ordinary light bulbs are given in the following table.

Input power of ordinary light bulb [W]	15	25	40	60	75	100
Luminous flux [lm]	90	200	400	700	900	1300

Luminous efficiency (lm/W) is a measure of how well a light source produces visible light. It is the ratio of luminous flux to the electrical power over the same area, measured in lumens per watt. Represents the value of the total light output of the source per 1 Watt of energy consumed. It is used to compare the efficiency of different types of LED lamp. You have, for example, two LED lamps, the first of which has a light output of e.g. 350 lm and the second 500 lm. The first of the LED bulbs then has a consumption of 5 W and the second 10 W. The first LED bulb then has a light efficiency of 70 lm/W and the second 50 lm/W. It is clear from the above example that the first of the bulbs has a higher efficiency, i.e. it is more economical (it gives more light per 1 W of energy consumed).

Illuminance [lux (lx)] is a measure of how much luminous flux is spread over a given area. One lux (Latin for “light”) is the amount of illumination provided when one lumen is evenly distributed per unit area. 1 lux is equal to 1 lumen per square meter: $1\ \text{lx} = 1\ \text{lm/m}^2 = 1\ \text{cd·sr/m}^2$.

6.1.2. Light sources

Incandescent light bulbs

They work on the principle of heating a thin metal filament (mainly from tungsten) with an electric current supplied through it. At high temperatures (Joule heating), the filament emits photons (due to the photoemission) mainly in the infrared region, partly in visible light. For highly powered incandescent light bulbs (projection types, halogens, etc.) ultraviolet radiation can also be found in the spectrum, but a standard bulb of a soft sodium-calcium glass bulb is practically non transparent to UV radiation. For standard bulbs up to power 15 W, the bulb is usually evacuated, for stronger bulbs it is filled with a mixture of nitrogen and argon, but also rarely with krypton or even xenon. These filling gases allow higher operating temperatures of the filament and simultaneously suppressing its aging by spraying or evaporation. From an optical point of view, the filament of the bulb does not behave like an absolutely black body emitter, but as if it were several hundred kelvins warmer (tungsten is a selective emitter). The main advantages of the incandescent bulbs include low price, simple construction, highly automated production, continuous brightness control (dimming), excellent color rendering, direct power supply, absence of harmful substances, emits no hazardous radiation, emits a continuous spectrum of light similar to this in sunlight. The main drawback include low efficiency and low luminous efficiency (around 10-15 lm/ W), low service life and strong dependence of parameters on the supply voltage.

Halogen classic light bulbs

Standard halogen bulbs are just improved classic light bulbs. They also contain tungsten filament in a small transparent chamber filled with an inert gas and a small amount of halogen (iodine, bromine). The bulb undergoes a so-called halogen cycle, where the evaporating tungsten combines at a high temperature and decomposes back for example, with bromine. Thanks to the tungsten vapor pressure near the filament, its evaporation is suppressed, therefore the halogen bulbs can operate at higher temperatures than conventional gas-filled bulbs resulting in a longer service life and increased luminous efficiency (up to 30 lm/W). The color temperature of the light is higher than with conventional light bulbs. Standard halogen bulbs (with lower efficiency) are not much more efficient than conventional bulbs (<15%).

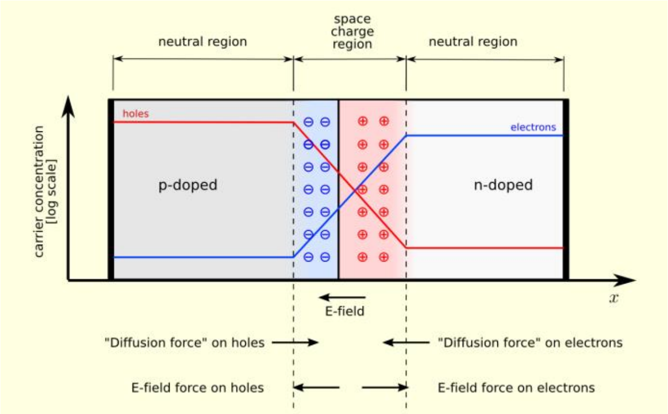
The new type of halogen bulbs filled with xenon or krypton enables 25-30%, sometimes up to 50% higher efficiency at the same light output compared to conventional bulbs. The bulbs are available both with a standard halogen bulb socket and in a classic bulb socket E27/E14. In the latest halogen technology, halogen bulbs have a coating that reflects back infrared radiation. This allows 45% lower energy consumption with the same light output compared to conventional light bulbs. Based on this technology, energy class B efficiency can be achieved. Improved halogen bulbs have a longer service life about 2000-3000 hours compared to conventional bulbs (1000 h). Due to the longer service life and higher efficiency these bulbs bring consumers net savings despite the higher purchase price. However, they are still relatively inefficient compared to compact fluorescent lamps.

Energy-saving compact fluorescent lamps

Energy-saving compact fluorescent lamps typically consist of a tube sealed at both ends with metal electrodes and an electronic ballast. It is necessary to use a ballast to limit and stabilize the current in the lamp. Electrons generated by hot cathode pass through a tube that is filled with the mixture of mercury vapors and inert gas (argon). Electrons are accelerated under applied electric field, ionize mercury molecules and glow discharge is ignites. During crossing to the basic energy level invisible ultraviolet radiation is emitted (wavelength of 253 nm). This emitted UV light is converted into visible light exciting the fluorescent coating (luminophore - phosphorus) deposited to the inside of the tube, and into heat when absorbed by other materials such as glass. Compact fluorescent lamps have 60-80% higher energy efficiency compared to standard incandescent light bulbs. The average service life of compact fluorescent lamps (about 6.000 to 20.000 hours) is 8 to 20 times longer than the service life of conventional incandescent light bulbs (1.000 hours). Some (rather more expensive models) may even be dimmable. They are produced in the form of tubes, but also in the design of a classic bulb with an E27/E14 thread. Today's fluorescent lamps do not pose a health risk in normal use. The effects of the UV radiation emitted by the fluorescent lamp can only be seen when exposed to a distance of less than 20 cm for a long time. The drawback of fluorescent lamps is that they flash at the electricity frequency (50/60Hz). For this reason, their use is limited where there are rapidly rotating objects, especially in machine tool workshops. Because of the stroboscopic effect, a rotating object illuminated by intermittent light may appear to be standing. A significant environmental impact of compact fluorescent lamps is the problem associated with their insufficient recycling. A significant number of compact fluorescent bulbs are still being disposed of as conventional waste. However, compact fluorescent lamps, due to their mercury content, must be disposed of as hazardous waste and subject to recycling. The retail price already includes the amount to be paid for recycling, and manufacturers and importers are obliged to collect and recycle compact fluorescent lamps. Recycling makes it possible to recover up to 90% of all materials and 95% of the mercury contained in the product.

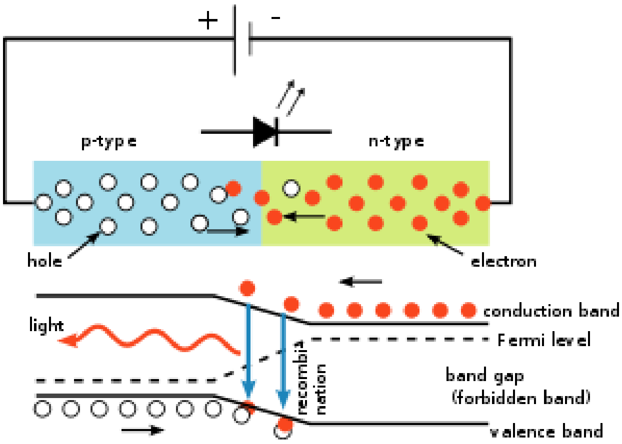
LED diodes (Light-Emitting Diode)

A light-emitting diode (LED) is a semiconductor light source operating as a forward biased standard diode with the p-n junction.



P-n junction is a boundary or interface between two types of semiconductor materials, p-type and n-type, inside a single crystal of semiconductor. Free electrons in the n-type diffuse in the concentration gradient into the p-type and recombine with the holes. In a similar way, holes in the p-type diffuse into the n-type and recombine with the free electrons. Near the junction positively charged ionized donor dopants in the n-type and negatively charged ionized acceptor dopants in the p-type are unmovable. Thus, the n-type region near the junction becomes positively charged and the p-type region near the junction becomes negatively charged. There are no mobile carriers in the regions near the p–n interface, forming the space charge region or depletion layer (see fig.). The electric field created by the space charge region has an opposite direction than diffusion process for both electrons and holes. There are two concurrent phenomena: the diffusion process that tends to generate more space charge, and the electric field generated by the space charge that tends to counteract the diffusion. In the p–n junction, without an external applied voltage, an equilibrium condition is reached in which a potential difference forms across the junction. This potential difference is called *built-in* potential V_{bi}.

When an electric field is applied on the p-n junction *in the forward direction*, electrons from the n-type and holes from the p-type are injected into the p-n junction area, where they recombine with each other and the recombination energy is irradiated in the form of photons (electroluminescence). The LED emits incoherent light with a narrow spectrum. The range of the irradiation wavelengths depends on the width of the semiconductor forbidden band gap. LED lamps are manufactured with emission range from ultraviolet to the infrared light including whole visible spectrum. By design, they represent an LED component in which the contacted chip (or combination of chips) is overlaid with spherical caps made of epoxy resin or acrylic polyester with the required optical properties. The materials from which LEDs are made have a relatively high refractive index, and without the cap much of the emitted light would be reflected by total reflection back at the planar interface with the air. LEDs can be manufactured in the spot or scattering design, with different emitting beam angles. Originally, the LEDs emitted monochromatic light (the first LED was red) and were used mainly for indication. An important milestone was the invention of the technologically demanded blue diode, which opened the way to the white diode. Another milestone was the introduction of a high-power LED, which gave consideration to the use of LED technology for general lighting. Compared to other electric light sources (bulb, discharge lamp, incandescent lamp), LEDs have the advantage that they work with relatively small current and voltage values.



LED material for different spectrum range.

Color	Wavelength (nm)	Bias (V)	Material
Infrared	$\lambda > 760$	$\Delta V < 1.9$	GaAs, AlGaAs
Red	$610 < \lambda < 760$	$1.63 < \Delta V < 2.03$	AlGaAs, GaAsP, AlGaInP, GaP
Orange	$590 < \lambda < 610$	$2.03 < \Delta V < 2.10$	AlGaInP, AlGaInP, GaP
Yellow	$570 < \lambda < 590$	$2.10 < \Delta V < 2.18$	GaAsP, AlGaInP, GaP
Green	$500 < \lambda < 570$	$1.9 < \Delta V < 4.0$	InGaN/GaN, GaP, AlGaInP, AlGaP
Blue	$450 < \lambda < 500$	$2.48 < \Delta V < 3.7$	ZnSe, InGaN, SiC
Violet	$400 < \lambda < 450$	$2.76 < \Delta V < 4.0$	InGaN Red/blue + violet luminophore
Ultraviolet	$\lambda < 400$	$3.1 < \Delta V < 4.4$	Diamond (λ = 235 nm), Boron Nitride (λ = 215 nm), AlN (λ = 210 nm), AlGaN, AlGaInN (λ = 375–395 nm)
White	Full spectrum	$\Delta V = 3.5$	Blue/ultraviolet + yellow luminophore

Green LEDs are manufactured from the InGaN-GaN system. These diodes are more efficient and brighter than LEDs made from other materials. Today, blue LEDs are made of wide-band gallium nitride (GaN) and indium gallium nitride (InGaN). Previously, they were also made from zinc selenide (ZnSe). The first blue GaN-based diode was manufactured in 1971 (it had very little power). In the 1980s, the optimal way of p-type GaN doping in epitaxial growth was found. Commercially available since the 90's.

The ternary and quaternary compounds AlGaN and AlGaInN began to reach the near UV region (375–395 nm). Diodes (relatively expensive) up to 210 nm are available today.

White LED

The LED function implies that white light cannot be directly emitted. Nowadays, there are several methods of achieving white color.

Three RGB diodes can be used, which emit three basic colors (red, green, blue). By mixing their radiation we get a white (multicolor white LED). Another way is to use a luminophore (yellow fluorescence) that converts the monochromatic radiation of a blue or UV LED to a white broad spectrum (phosphor-based white LED). The same process takes place in fluorescent lamps.

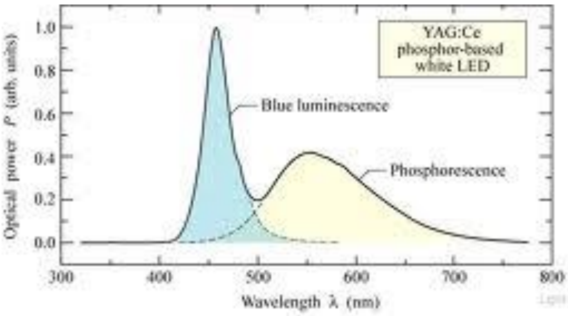


Fig. 21.8. Emission spectrum of a phophor-based white LED manufactured by Nichia Corporation (Anan, Tokushima, Japan).

E.F. Schubert
Light Emitting Diodes (Cambridge Univ. Press)
www.LightEmittingDiodes.org

For phosphor-based white LEDs, usually blue InGaN LEDs are encapsulated in an epoxy-mixed luminophore. Cerium doped yttrium aluminum garnet (YAG:Ce^{3+}) is used as a source of yellow fluorescence. The spectrum of such a "white" LED contains a peak of blue light emitted from an InGaN LED (465 nm) and a broad spectrum caused by a Stokes shift in the range of 500-700 nm. The white diode can also be made from LED emitting near ultraviolet radiation, which is encapsulated in a mixture of a red and blue emitting luminophore doped with europium and a green emitting zinc sulphide doped with copper and aluminum (ZnS: Cu, Al). This method is also used in the production of fluorescent lamps. This method is less efficient than LEDs with YAG:Ce^{3+} due to greater energy-to-heat conversion due to greater Stokes shift, but achieves better spectral characteristics. Due to the higher performance of UV LEDs than blue LEDs, white LEDs achieve comparable luminous intensity. The only problem is the possible UV radiation escape. Another way to obtain a white LED is to homoepitaxial growth of a layer of zinc selenide (ZnSe) on ZnSe substrate. This structure will simultaneously emit blue light from the active P-N junction area and yellow light from the substrate. The new progressive method is based on covering blue LEDs with quantum dots that emit white light when excited by blue LED light. These LEDs emit a warm yellowish white light very similar to conventional light bulbs.

White LED- Luminous efficiency (60–100 W light bulb ~ 15 lm/W, fluorescent lamp up to 100 lm/W

2002	18–22 [lm/W] (Cree Corporation)
2003	65 lm/W at 20 mA
2006	131 lm/W at 20 mA (Seoul Semiconductor)
2008	145 lm/W at 20 mA (Seoul Semiconductor)
2009	150 lm/W at 20 mA (Nichia Corporation)

Unfortunately, the increase in value due to the physical and technological achievement is slowly limiting. Today, the luminous efficiency is in the range of 200-250 lm/W measured in the laboratory. Yet LEDs are clearly well ahead of conventional light bulbs. While the efficiency of energy conversion of light bulbs is from 10 to 20%, highly efficient LEDs currently reach values of 40 and 50%. However, it is still only 50%, so the remaining 50% of the power is lost in the form of heat.

LED lamps

The basis of every LED light bulb is an LED chip, of which there may be more than one in the LED bulb. The total luminous intensity of an LED bulb depends mainly on the luminous intensity and luminous efficiency of the built-in LED chip and on their number. One high-performance LED chip can easily produce more light output than an LED bulb with 100 built-in standard LEDs. Judging the luminous efficiency of LED bulbs only by the number of built-in LEDs, or LED chips, is therefore nonsense. Today's LED bulbs are manufactured in a variety of variants in terms of design and light output. There are both very cheap low-power LED bulbs on the market, which are designed for use to illuminate corridors, etc., and LED bulbs, which replace old incandescent bulbs, to special high-performance systems of LED chips, replacing sodium lamps in the street lamps. LEDs are designed as point or scattering light sources. If it is necessary to illuminate only a specific area, such as a working desk, a spot bulb with, for example, half the light output will be sufficient, as opposed to an LED bulb with a diffused light, which will illuminate everything around it except the working desk. Current LED bulbs normally reach values from 60 lm/W to 170 lm/W. So if an LED bulb has a luminous efficiency of about 100 lm/W, and the power consumption of this LED is 5W, then its total light flux is 500 lm.

LED bulb design

In terms of design, LED bulbs could in principle be divided into single and multiple-chip LED bulbs or LED bulbs with the classic type of LEDs.

1. LED chip bulbs use high-brightness LED modules that produce a relatively large amount of heat over a very small chip area, so the chips need to be cooled. This is usually done by means of a passive heat sink, which forms the aluminum housing of the LED bulb; the need for cooling is, of course, necessary to subject the entire construction of the LED bulb.

2. The second type are LED bulbs fitted with LED diodes in classic cylindrical housings with a diameter of 5 mm. This type of LED light bulbs can usually be purchased as spot LEDs, often

several dozen of which are placed in front of a reflector reflecting light. Or as a 'Corn' type, where the body of the LED bulb forms a cylinder and the individual LEDs are evenly distributed over the entire surface of the housing, except for the part where the socket of the LED bulb is located.

In addition to LEDs (LED chips), there is also an electronic part in the LED bulb, which provides a suitable power supply for the LEDs so that the LEDs can be powered, for example, directly from the 230V electricity grid.

Advantages of LED bulbs

1. High efficiency of LED bulbs, which is currently up to 12 times higher than incandescent bulbs.

2. Very long service life. Today's LED bulbs have a service life of 30 to 100 times the service life of incandescent light bulbs.

3. High mechanical resistance of LED bulbs, i.e. resistance to shocks, impacts, vibrations. If your LED bulb falls to the ground, it is very unlikely that it will be damaged.

4. Simple lighting control. Conventional dimmers for incandescent bulbs can be used for control. In addition, LED bulbs do not change, unlike conventional bulbs, the color of the light according to the intensity of the set light.

5. Low surface temperature of LED bulbs brings increased safety.

6. LED bulbs do not emit ultraviolet and infrared light, they do not contain Hg.

7. Possibility of frequent switching on and off, without affecting the service life of the LED bulb.

8. Immediate lighting to full power, as opposed to energy-saving (fluorescent) bulbs and fluorescent lamps.

The main disadvantage of LED bulbs compared to conventional bulbs is still the higher price. However, even the significantly higher price of an LED bulb will be balanced by 30 to 100 times longer service life and higher efficiency that is conventional light bulb.

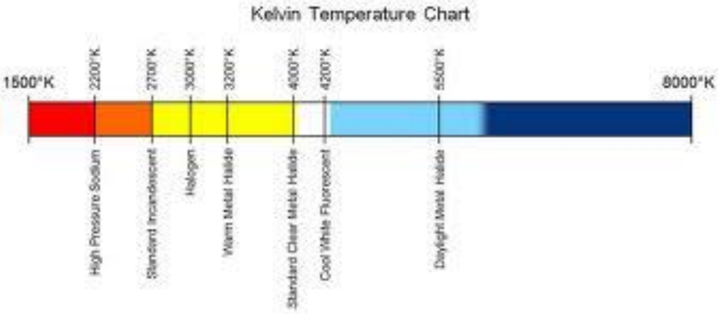
Currently, diodes with an efficiency of about 100-170 lm/W can be found for general lighting, which is better than with fluorescent lamps and some discharge lamps. In addition to high efficiency, the major advantage of LED technology is considered to be long service life. In practice, however, this parameter depends on many circumstances and the responsible manufacturers do not give a number higher than 50.000 hours, but usually much less (e.g. 25.000 hours). The drawback of LED technology is the dependence on temperature, which is crucial for the design of further use, and the drawbacks similar to fluorescent technology: the need for a luminophore for white light and a gradual decrease in luminous flux during the service life. Despite these disadvantages, LED technology is evolving intensively and there are already many functional applications like development of so-called LED lamps, i.e. LED light sources with E14 or E27 threads suitable for replacing ordinary light bulbs. The highest current wattages are 12 W-15 W and in the warm white light variant they correspond to 100 W-150 W of an ordinary light bulb.

Main criteria for selecting a suitable household light source

Purchase price

The purchase price of compact fluorescent lamps and modern, more efficient halogen bulbs is higher than the price of conventional light bulbs and standard halogen bulbs. It is necessary to consider the operation (lighting) regime of the light source, switching frequency and other factors. The return on "investment" in an energy-saving light source can be in the order of several months or several years. However, it usually represents energy savings with a very short return on investment - in contrast to, for example, insulation or window replacement, where in terms of return on investment, this is a long-term investment.

Light color, Color temperature



The light color of a light source is determined by comparison with the wavelength of the absolute black body radiation at a given temperature. The temperature is conventionally expressed in kelvins (K) at which the radiation of an absolute black body corresponds to the color of the light source is called the color temperature. In conventional light bulbs, the color temperature is more or less the same as the temperature of the tungsten filament. With other technologies, such as compact fluorescent lamps, this relationship is more complex and the color is assessed in terms of human perception of light quality (the so-called color replacement temperature is given). Color temperatures over 4500 K are called "cool light" (bluish), while lower color temperatures (2500–3500 K) are called "warm light" (yellowish). The range of 2500–3500 K is also typical of standard incandescent or halogen bulbs. The cool light is typical for fluorescent lamps and halide lamps. Currently, most compact fluorescent lamps are offered both in the "warm white" version with a color temperature of 2500 K or 2700 K and in the "cold white" version (usually the color temperature of 4000 K) or "Day white" with a blue tinge (usually a chromaticity temperature of 6500 K).

Light source service life and number of switching cycles

Light source service life is a parameter given in hours and depends on the light source technology used. The average service life of a light source is about 1000 hours of lighting for standard incandescent bulbs, 2000-3000 hours for modern halogen bulbs, about 10.000 to 20.000 hours for compact fluorescent lamps and up to 45.000 hours for LED bulbs. However, the service life may also depends on the number of on and off cycles. Most fluorescent lamp sold today, that use the preheat function, can theoretically be switched on and off many times each day for 20-30 years. However, there are still cheaper and lower quality products on the market for which the high number of switching on and off cycles is destructive.

Color rendering

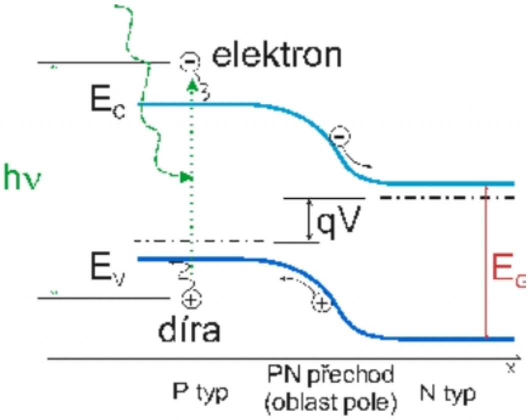
A color rendering index (CRI, Ra) is a quantitative measure of the ability of a light source to reveal the colors of various objects faithfully in comparison with an ideal or natural light source. The highest possible CRI value is 100 and would only be given to a continuous-spectrum light, such as sunlight or light emitted from incandescent light bulbs, allows excellent color reproduction. Compact fluorescent and LED bulbs, which emit light only in defined wavelengths range, show lower values of the color rendering index (70-95) compared to conventional bulbs. A light source that emits light only in a very narrow segment of the color spectrum (e.g., a low-pressure sodium lamp) has a negative color rendering index and does not provide good color resolution.

Additional light source parameters

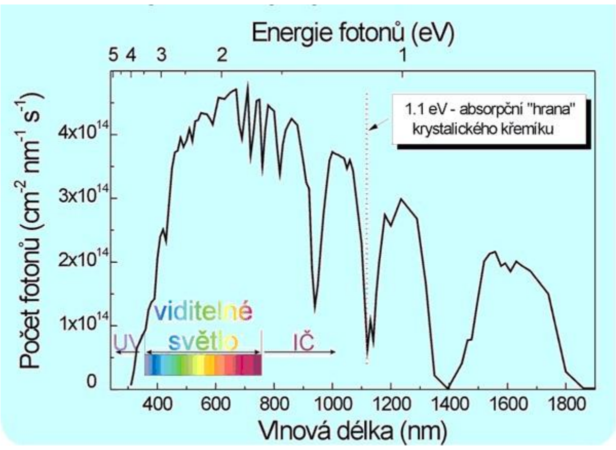
Other important parameters are the warm-up time, the start rate of the source at full power and the dimming option. For each light source, there is also a gradual decrease (degradation) of light flux during service life. The aging factor of the light source indicates the rate of flux at the end of the service life compared to the original flux at the beginning.

6.2. Photovoltaic solar cells

Photovoltaic cells use direct conversion of light energy into electric energy due to an internal photo effect. A solar cell is a large-scale photodiode with at least one p-n junction. Without an external applied voltage (open-circuit), an equilibrium condition due to diffusion of charge carriers across the junction and their recombination is reached. The region where electrons and holes have diffused across the junction, is called the depletion region (the space charge region) because it contains practically no mobile charge carriers. Across this region a potential difference called built-in potential V_{bi} is formed (typically 0.6-0.7 V in silicon). After solar cell irradiation, electron-hole pairs are generated (a negatively charged electron and a positively charged hole), which are separated from each other by an internal electric field of the p-n junction and diffuse towards the respective contact. Charge separation results in a voltage difference in the solar cell contacts (typically 0.6-0.7 V in the case of silicon). The voltage measured is equal to the difference in the quasi Fermi levels of the majority carriers. If we connect an electric device (load) to the contacts, a direct electric current passes through it, which is directly proportional to the solar cell area and the intensity of the incident sunlight. Electrons that are created on the n-type side, or created on the p-type side and "collected" by the p-n junction and swept onto the n-type side, may travel through the wire, power the load, and continue through the wire until they reach the p-type semiconductor-metal contact. Here, they recombine with a hole that was either created as an electron-hole pair on the p-type side of the solar cell, or a hole that was swept across the junction from the n-type side after being created there. Voltage of one cell is too low for further use, so the individual cells are interconnected to obtain the voltage that is already applicable in different types of photovoltaic systems. As a standard, cell sets for nominal operational voltage of 12 V or 24 V are used. Such cell sets, in serial or serial-parallel interconnection, are hermetically sealed into the covering materials creating final solar panel.



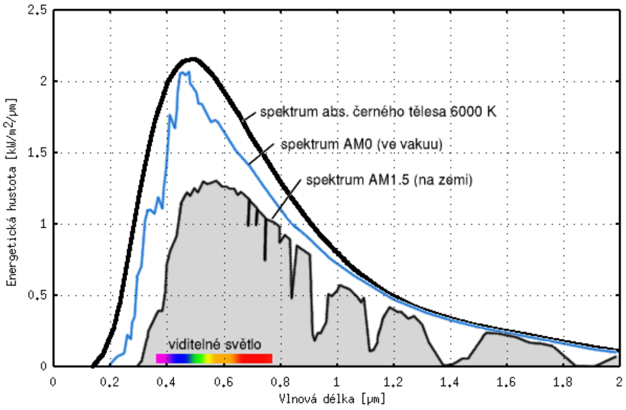
P-N junction- energy bands



Solar spectrum after passing through the atmosphere

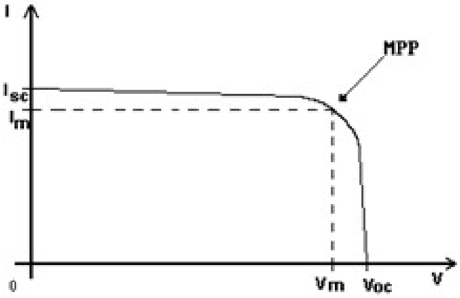
Sunlight incident on the Earth surface (after passing through the atmosphere) consists of photons of different wavelengths and thus different energies. From the whole solar spectrum, only the radiation in the range of 380 nm to 780 nm is visible for the human eye. Solar radiation extends even further into the ultraviolet (UV) region with a shorter wavelengths (greater energies) and significantly into the infrared (IR) region with a greater wavelengths. In the construction of solar cells, the goal is to use such material which would absorb radiation from the entire sun spectrum. If, for instance, a silicon is irradiated by photon with energy smaller than 1.12 eV (this energy corresponds to the width of forbidden band $Eg=Ec-E_V$, and thus the “absorption edge” of silicon), it passes through the silicon and is not absorbed. If its energy is greater than 1.12 eV, then the photon is absorbed and free charge carriers are generated in the semiconductor.

Solar cells efficiency.



Solar radiation is produced by a thermonuclear fusion in the center of the Sun at temperatures of about 15 million Kelvin. On the surface of the sun, the temperature is already around 6.000 K. The luminous power of the whole sun is 3,85x10^{23} kW. Most of this power is emitted into space and only the power of 1,744x10^{14} kW reaches the Earth across the whole irradiated hemisphere. The Earth orbits the Sun at the distance of 150 million km. The energy density of sunlight in this distance is 1367 ± 7 W/m^2 in the vacuum (AM0 - Air Mass Zero). This energy is distributed into the electromagnetic spectrum roughly equivalent to the radiation of an absolute black body at the temperature of 5700 K. When passing through the atmosphere, part of the solar energy is lost. About 300 W/m^2 is absorbed in the atmosphere and about 100 W/m^2 is scattered. Part of the scattered energy contributes to the total illumination as diffuse radiation of the sky (sky blue – Rayleigh elastic scattering). The efficiency of solar cells is measured at the defined illumination of AM 1.5 (Air Mass), which is a model spectrum of sunlight after passing through a cloudless atmosphere. If the sun is right at its zenith (the height of 90 degrees), sunlight passes through the thinnest possible atmosphere layer. Such a spectrum is designated as AM1. However, most of the time, the Sun is not so high, and sunlight has to pass through a thicker layer of atmosphere. For photovoltaic applications, the AM1.5 spectrum, corresponding to the height of about 45° above the horizon is therefore used. In this case, sunlight passes through a layer of air one and a half times thicker. The energy density of this spectrum is 1 kW/m^2; but in reality it strongly depends on the atmosphere's transparency. The total theoretically usable sunlight power corresponds to the grey area under the AM1.5 curve.

Photon energy which is not consumed to create electron-hole pairs is converted into heat. Thus, in a photovoltaic cell, a maximum of 50% of the incident light can be theoretically converted into electricity. Efficiency of a photovoltaic cell $\eta = P_{elm}/P_{rad}$ is determined by the ratio of electric power output $P_{elm} = I_m U_m$ diverted from the cell (I_m and U_m are the current and voltage of the photovoltaic cell to the external load) and the power output P_{rad} of the incident radiation. The cells efficiency may be expressed by a product of fractional efficiencies $\eta = \eta_r \eta_e \eta_p FF$.



where $\eta_r = P_{abs}/P_{rad} = 0{,}70$ is the ratio of the reflected light power and the incident light power (with the average reflectance of silicon of $R = 0{,}30$), $\eta_e = 1\text{-}T/T_s = 0{,}95$ is the efficiency of the Carnot cycle, where $T = 300$ K and $T_s = 6000$ K are the ambient temperature and the temperature of the Sun, $\eta_p = 0{,}42$ is the contribution to efficiency due to silicon lattice constant no matching to the sunlight spectrum (CdTe or possibly amorphous silicon (a-Si) are optimal materials for solar conversion), and *FF* (the Filling factor) is the so-called factor of filling of photovoltaic cells determines the maximum power from a solar cell and is given by the cell's electromagnetic parameters (quality of the contacts, material morphology and resistance of the active semiconductor layer). The *FF* is defined as the ratio of the maximum power from the solar cell to the product of *Voc* and *Isc*. $FF = U_m I_m / U_{oc} I_{sc}$. The image shows an example of an I-U characteristics of a real photovoltaic cell during radiation. I_{sc} is the short-circuit current through the cell, U_{oc} open-circuit voltage, U_m and I_m are current and voltage corresponding to the maximum electric power output P_{elm} (MPP) of a photovoltaic cell. Today, the efficiency of industrially produced cells is about 15%. For experimental laboratory-produced cells, efficiency of up to 40% is achieved. The current thin-film cells reach the efficiency of 8-9%, which, however, decreases over time much faster than in the case of thick-film cells

The power output of the photovoltaic cells and panels is given in the units *Wp* (watt peak). The power output is strongly dependent on illumination and on the angle of the incident light, which is why the power output of cells is measured under defined conditions:

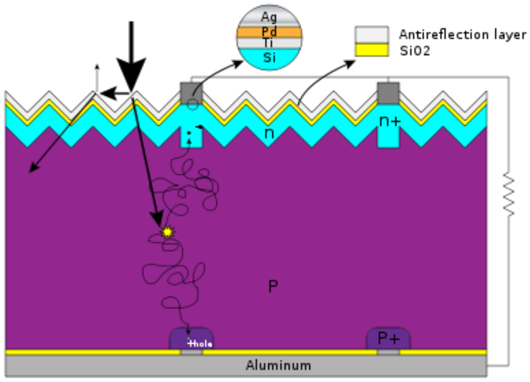
- Power spectral density of sunlight of 1000 W/m^2
- Radiation spectrum of AM1.5
- Solar cell temperature of 25^{o}C.

In practice, the power output of a solar cell is lower most of the time, as the cell is not oriented exactly towards the Sun, and the light passes through different layers of the atmosphere depending on the time of the day. In addition, the quantity of the incident light is strongly dependent on how cloudy it is.

Materials for solar cells

Materials used for solar cells must absorb a sufficient amount of radiation in the solar spectrum. Some are optimized for the operation on Earth other outside the Earth's atmosphere. Therefore, combinations of different materials are often created to capture most of the sunlight in the most efficient way. Most solar cells are manufactured in the form of bulk about 0.2 mm thick wafers. Further, thin polycrystalline films or organic polymers applied to a solid substrate are used. The third group are nanocrystals (electron-confined nanoparticles), which form quantum dots.

Silicon – Most of the cells (90%) used today are manufactured from a single crystalline or polycrystalline bulk p-type silicon doped by boron. Polycrystalline or multi crystalline silicon (poly-Si or mc-Si) is suitable for lower power outputs. The main advantage is the relatively good efficiency even at lower illumination levels. Single crystalline silicon (c-Si) which is produced by the Czochralski method, is suitable for greater power outputs. With sufficient illumination, the cells have a greater efficiency than polycrystalline silicon cells (the production efficiency is 16% to 19%, special structures up to 24%). A cell from single crystalline silicon with the area of 100 cm^2 is capable of delivering current of 3-4 A. Polycrystalline silicon ingots are produced with a square cross-section suitable for the production of solar cells. Round single crystalline ingots have to be shaped into a pseudo-square cross-section to better utilize the solar panel area. Ingots are cut into thin wafer on which a texture is created by etching (the wafer becomes matt and absorbs light better). The wafer is then doped by phosphorus to form an n-type thin layer and a P-N junction, and is then equipped by an anti-reflex coating layer of nitride (the cell becomes dark blue), and using a conductive paste, metallization on the front and the back side is created. The cell is then sintered creating a conductive connection between the metallization and the silicon. The finished cells are combined into a series (and/or parallelly) and installed into photovoltaic panels.



Silicon is used in the form of a thin multi-crystalline layer, formed directly from the melt. These cells have a lower efficiency than poly-Si; however, but the cutting process is eliminated (lower material losses – they are cheaper).

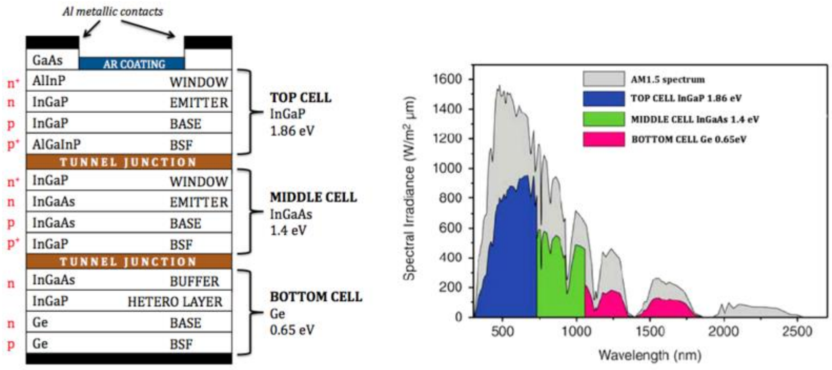
Thin-film Silicon

Thin Si layers are most often formed on a glass substrate by chemical vapor deposition by decomposition of silane with the presence of pure hydrogen (plasma-enhanced chemical vapor deposition PE-CVD). Depending on the growth parameters, Si may be obtained in various forms:

1. Amorphous Si (a-Si or a-Si:H)
2. Polycrystalline Si
3. Nano crystalline (micro crystalline) Si (nc-Si or nc-Si:H)

Thin-film solar cells have a greater absorption capability. Most of the solar spectrum is absorbed in a very thin layer. Already a 1 μm thick layer absorbs 90% of the radiation. Despite this, they usually have a lower efficiency than bulk Si, due to poorer charge transport, but they are much cheaper. Amorphous Si has a greater forbidden band width (1.7 eV) than c-Si or nc-Si, therefore it absorbs the visible part of the solar spectrum more effectively. This is utilized in special structures (tandem cells) consisting of the p-i-n junction structure of amorphous (hydrogenizes) silicon (a-Si:H) and microcrystalline (hydrogenizes) silicon (μc-Si:H). Amorphous silicon has high absorption in the blue, green, and yellow part of the spectrum; monocrystalline silicon then absorbs well in the red and infrared area. Microcrystalline silicon may also be replaced by a compound of silicon and germanium and their optical (absorption) and electrical properties can be modified according to the chosen ratio of both materials. This structure is used commercially for three-layer solar cells production, where the two bottom cells are made with different concentration of Si a Ge.

Gallium Arsenide (GaAs) - The main advantage is higher efficiency - 20%, greater resistance against the cosmic (hard) radiation and the ability to work without reducing efficiency even at temperatures above 100 °C. The drawbacks include a much higher price compared to crystalline silicon. GaAs is the main material for solar cells in space applications.

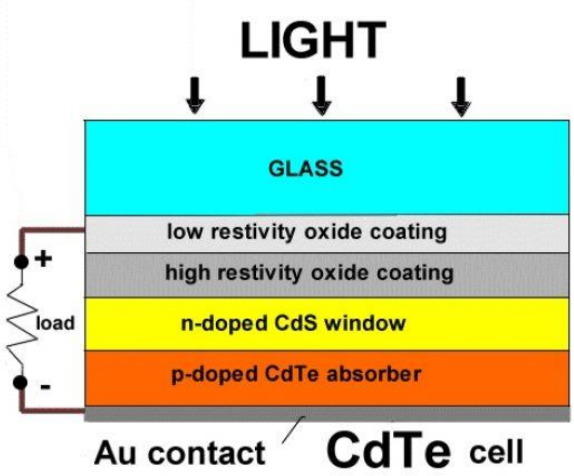


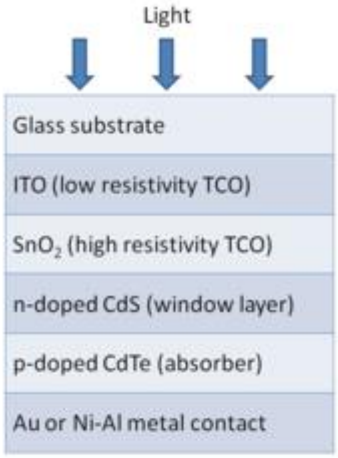
(a) (b)

However, in combination with light concentrators, they are starting to find use in terrestrial systems as well. The highest efficiency is achieved in multijunction cells, which contain heterostructures of different materials produced by the MOVPE method. The structures are connected in series, so the same current flows through them. They are connected in descending order according to the forbidden band width. The materials are selected so as to efficiently absorb most of the solar spectrum. For instance, a three-layer structure containing InGaP, InGaAs and Ge semiconductors is used. Each semiconductor has a different forbidden band width and therefore dominantly absorbs radiation with a different wavelength. InGaP (E_g = 1.8 eV - 1.9 eV), InGaAs (E_g = 1.4 eV) and Ge (E_g = 0.67 eV). All materials must have an addapted lattice constant. These cells are currently most effective, but also most expensive. In 2020, the National Laboratory for Renewable Energy (USA) introduced cells with efficiency of up to 47.10%.

Copper-indium-gallium-selen (CIGS) is a direct compound semiconductor with a forbidden band width in the range of 1.0 eV-1.7 eV depending on the composition (CuInSe-CuGaSe). It is used in the form of thin polycrystalline layers, which are prepared by vacuum evaporating or sputtering, also chemically from a solution. The cells reach efficiency up to 20%.

Cadmium telluride (CdTe) is a direct semiconductor with an optimal forbidden band width of 1.5 eV for single-layer solar cells. It is used in the form of thin polycrystalline layers, which are prepared by sublimation to a glass substrate. So far, it is the only thin-layer material that actually competes with the Si thin films in term of efficiency and price. Today, the cells reach efficiency of about 20%. The cells consist of a back-side metallization layer using a transparent conductive oxide (TCO) formed by the ITO compound – Indium Tin Oxide (90% In_2O_3 and 10% SnO_2 in solid solution), transparent high-resistive layer of SnO_2, input n-type layer of CdS (forbidden band width of 2.42 eV), the absorber of the p-type CdTe layer and front-side metallization layer.





Dye-sensitized cells

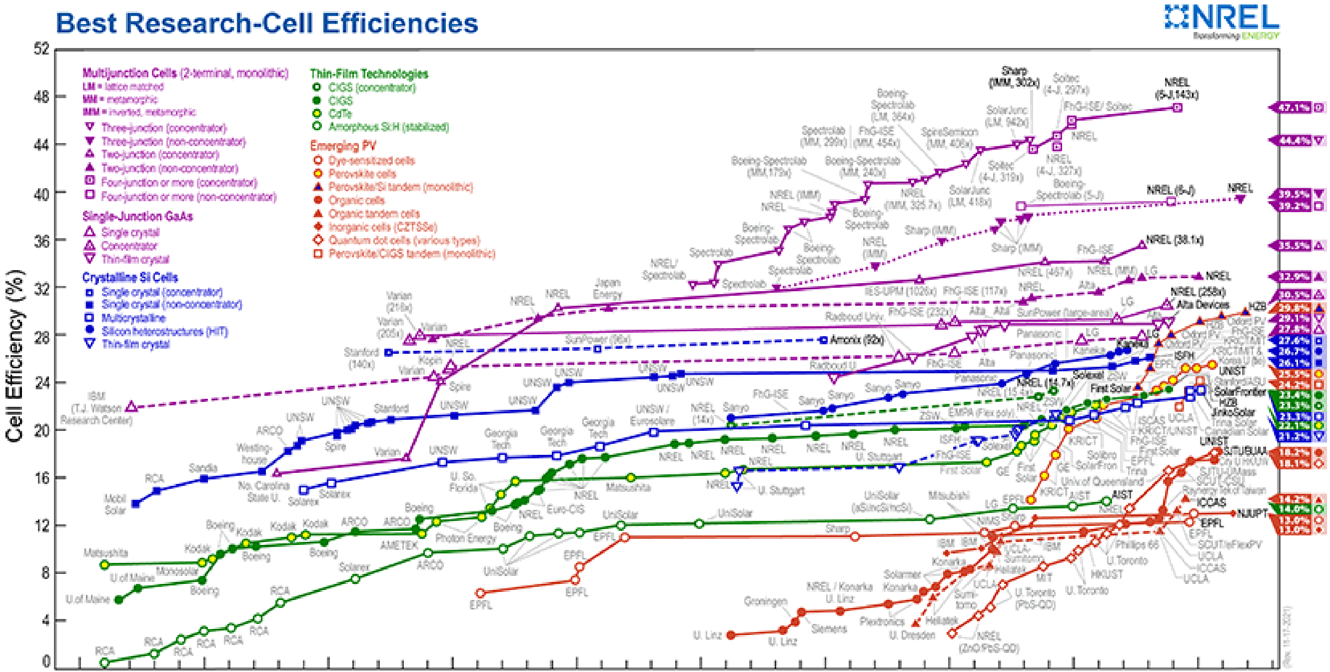
Organic cells

These are photosensitive polymer applied in a thin layer onto a foil. They are light weight, flexible, cheap, and easily shaped (the ability to print on foils similar to an inkjet printer). The disadvantage, however, is the low efficiency (1/3 of c-Si) and temporal photo instability (degradation).

Perovskites

Summary

The most common solar cells based on c-Si reach the efficiency of 16-19% in global production (special structures of up to 24%). The production is relatively expensive (expensive production of crystalline silicon); despite this, they will dominate the market in the next few years.



https://www.nrel.gov/pv/cell-efficiency.html

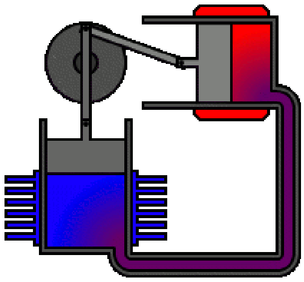
Alternatives

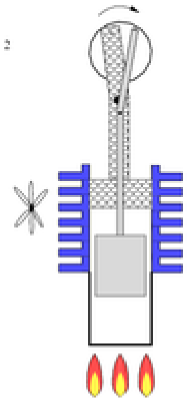
Concentrated solar thermal power plant - CSP

The CSP plant using movable parabolic mirrors or lenses to concentrate a large area of sunlight onto a chamber-absorber with a special liquid located in the focal point of mirrors. By solar radiation, steam evolves from the liquid, and drives a heat engine (usually a steam turbine) connected to an electrical power generator. The parabolic mirror positions are controlled by a central computer, which follow the Sun's movement in the sky throughout the day. The energy supply from CSP is unstable, because it operates only during the day, however, the power output is much more regular and predictable than in case of wind power plants, for example. Using an energy storage device will make CSP a regulated source. Another way to improve the possibility to regulate and the reliability of CSP is to build a hybrid power plant using another fuel to heat the steam (such as natural gas). Such power plant with the peak power of 377 MW is built in Ivanpah in the Mojave Desert in the US, currently the largest solar thermal power plant in the world.

Stirling engine.

The Stirling engine has two workspaces between which gas can freely flow (there is virtually the same pressure). One of the spaces is cold, the other one is hot. This is achieved either by direct heating and cooling of chambers or, more often, by an external heater and cooler. A heat-accumulating regenerator is usually included between the heater and the cooler accumulating the heat of the gas passing from the heater to the cooler or opposite. Two pistons move in the working spaces simultaneously with opposite phase. By compression at low working gas temperature and expansion at high working gas temperature, heat energy is transformed into mechanical work.





The pistons may be in separate cylinders, or in a single cylinder, with one common piston operating in dual mode. One piston is leak tight while the other is semi-permeable. Air was previously used as a working gas. Nowadays, helium or hydrogen are used thanks to a higher heat conductivity and smaller turbulent losses.

The main advantage of this engine is the fact that it can work with various sources of external thermal energy, from geothermal or solar to fossil fuels or biomass. The thermal efficiency of engines with the power output between 1 kW and 25 kW ranges between 25% and 33%. The energy efficiency is in the range of 18% and 22%. Other benefits include silent operation, durability, and minimal possibility of failure. The drawbacks are a poor regulation and lower readiness for operation. It also needs a relatively large cooler with a powerful fan, and had to work with high pressure gas to achieve high efficiency.

In 2008, a solar system in Albuquerque, New Mexico that consist of Stirling engines placed in the focal point of a parabolic solar-radiation concentrator and connected to generators produces an alternating voltage managed to achieve a net efficiency rate of 31.25% in electricity generation. The stated efficiency is the total efficiency, i.e. solar-to-grid efficiency, and includes conversion directly to alternating current. This differs from the efficiency reported for photovoltaic panels, which includes only conversion of sunlight to a direct current and then, electricity needs to be converted from DC to AC in the systems supplying electricity to the electrical grid (the efficiency of this conversion is around 93 %).

Solar thermal collector for preparation of hot water and additional heating

Solar thermal collector supplies heat by absorbing sunlight, for the purpose of either direct heating or indirect electrical power generation from the heat. Solar collectors operate with conversion efficiency several times higher than solar cells. If they are used to produce hot water in the home or water in a pool, their application is clearly the most efficient of all solar systems in our climatic conditions. Recently, systems in combination with additional heating have been becoming more widespread. An optimally designed solar system may cover about 60-70% of consumption for heating of water in our climate conditions or to assist in additional heating up to 30%.

Electricity production

Electricity production in the Czech Republic is dominantly ensured by the coal-fired power plants, which are located close to coal deposits, and nuclear power plants. Coal power plants produce about 37% of the total energy supply in ČR. Two nuclear power plants, one in Temelín in south Bohemia and the other in Dukovany in south Moravia are another important source of electricity. Their share the electricity production amounts to 34%. Natural gas is currently 11% of the energy mix. Construction of several other steam-gas power plants is being prepared (Počerady, Mělník).

Renewable energy sources

In the recent years, the share of renewable energy sources has been increasing (13% in 2020). Of this, solar systems (2.8%) and the energy biomass combustion (3.1%) account for a significant proportion. The Czech Republic has good conditions for biomass production and can be burned in the modified boilers of coal-fired power stations either together with coal or entirely on its own.

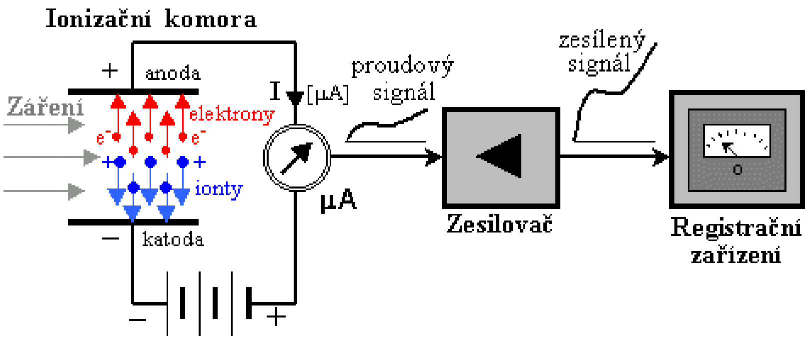
Biogas is another important source (3.2%), which emerges, in the processing of biodegradable waste. There are many hydro power plants in the CR; however, due to the nature of Czech rivers (longer and only gently flowing rivers), only artificially built large dams may be used for the dominant production of electricity. The small hydro-plants reach 500-1000 kW only. Today, hydro power plants supply 2.6% of electricity and they are used mainly for compensating power outputs in the consumption peak.

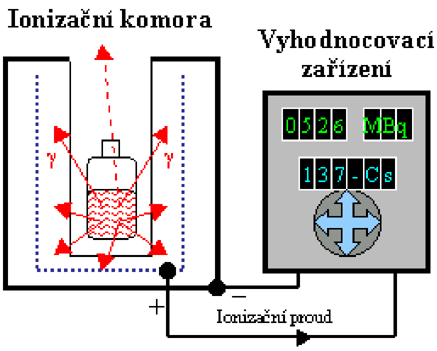
6.3. Radiation detectors

We can divide radiation detectors into several categories:

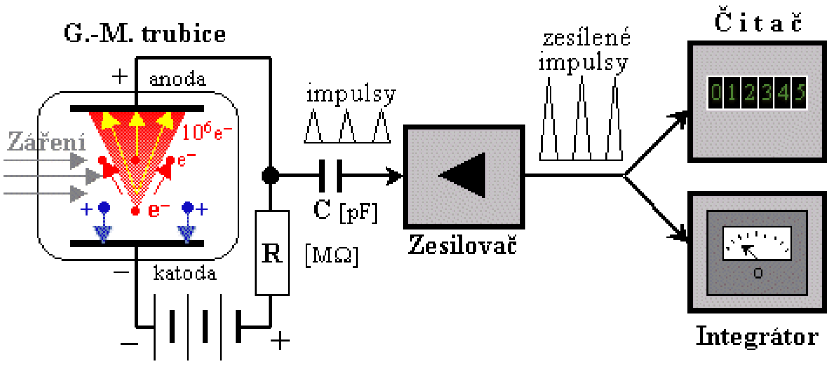
- Emulsion detectors   oldest detectors. Detector - photographic emulsion.
- Gas detectors   Ionization chamber, Geiger-Miller counter, Cloud chamber, Bubble chamber.

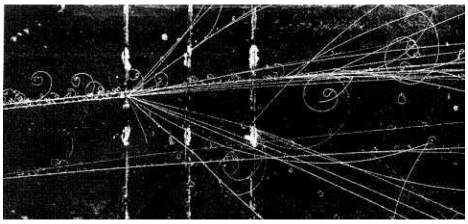
Ionization chamber is a closed gas-filled chamber fitted with two metal electrodes (anode and cathode) placed in a gaseous environment to which voltage of several hundred volts is applied. It is widely used for the detection and measurement of various types of ionizing radiation; X-rays, gamma rays or beta particles. Without the presence of radiation, no current passing through the circuit, because the gas between the electrodes is non-conductive. But, when ionizing radiation enters into the space between the electrodes, electrons are released from their originally neutral position in the gas atoms and change atoms into positive ions. Negative electrons drift in the applied electric field toward the positive anode while positive ions to the negative cathode and weak electric current, which is proportional to the intensity of the ionizing radiation, starts to flow through the circuit caused by ionic conductivity of the ionized gas between the electrodes. Electric current is amplified and measured by an electrometer circuit. The electrometer must be capable of measuring the very small output current which is in the range of femtoamperes to picoamperes, depending on the chamber design, radiation dose and applied voltage. Using different design, we are able to measure radioactive liquids, as well.





The Geiger-Miller counter (G-M) is also an ionization chamber. The device consists of a tube (Geiger tube) filled with an inert gas - helium, neon or argon with small portion of methyl alcohol or bromine vapours, for example, working as a so-called dimmer. The dimmer terminates the resulting discharge preventing the formation of a steady current, which would make further measurements impossible and could damage the electrodes or the gas charge. The dimmer is used to limit the duration of discharge to a few microseconds. The G-M counter uses impact avalanche ionization. Upon collision with gas atoms, the gas is ionized. The released electrons are accelerated to the anode, the positive ions to the cathode. When these primary accelerated electrons impact other atoms, secondary electrons are kick-out from the non-ionized atoms, which can kick-out additional electrons after acceleration. This process is called the avalanche effect. Simultaneously, as free charge carriers of both polarities are formed, they also disappear by mutual recombination of electrons and methyl alcohol or bromine cations. It is then essential for the formation of the discharge that the formation of free charge carriers by ionization overlaps their recombination. During the discharge, a strong electrical pulse passes through the circuit and a high voltage pulse is generated on the load resistor R, which is processed through the isolation capacitor C. It allows the detection of a single quantum of the respective ionizing radiation. The counter cannot differentiate between radiation types and cannot detect two particles coming immediately after each other, because the detector is insensitive for a short time after ionization by one particle - this time is called the dead time. The generation of the voltage pulse is also converted into an audio signal to produce audible clicks representing the number of detected ionization events. The measured signal is displayed by the processing electronics.





Wilson cloud chamber. It consists of a sealed glass cylinder filled with gas (air or argon) with saturated vapours of suitable liquids – water vapor is used with an addition of several organic substances, most often alcohol. The cylinder is equipped with a piston or a membrane on one side, the movement of which enables a quick change of volume and pressure inside the cylinder. If the chamber's working space is expanded rapidly, the adiabatic expansion of the gas in the cylinder causes the temperature drop below the dew point and the present saturated vapours become oversaturated due to cooling. If a charged particles pass through such a workspace either just before or during expansion, they ionize atoms along their path and create a number of ions that attract and thus locally concentrate the vapor molecules. On these ions, as condensation nuclei, supersaturated vapors condensate – the trajectory of the particle is covered by a series of microscopic droplets. The disadvantage of the Wilson cloud chamber is the short sensitive period when the particles are captured during the working cycle.

Bubble chamber – to make the ionization traces visible, the process of creation of bubbles of gas (or vapor) in an overheated liquid along the particle's ionization trajectory is used. If a charged particle passes through the liquid during the unstable state of overheated liquid before the boiling starts, it creates a number of ions along its path. Microscopic bubbles of vapor begin to form on these ions which, if the liquid is sufficiently overheated, can grow into macroscopic proportion – a series of visible tiny bubbles is created along the trajectory. Unlike cloud chambers, in which the gas is diluted, bubble chambers have an advantage of greater density of the liquid, with which the high-energy particles can interact more easily and of the greater speed of the working cycle.

- Thermal detectors – Absorbed radiation changes the temperature of detector and detected signal is thus characterized by a change of some of the absorbing material characteristic which depends on temperature; i.e., we register e.g. a change of the electrical resistance (bolometer), thermoelectric voltage (thermal cell), or change in the pressure of heated gas (pneumatic detectors-Golay). These detectors are independent on the wavelength of the incident radiation;

their response time is long due to the temperature change. They are suitable for calibration purposes.

- Photon (quantum) detectors – Absorbed radiation causes electron transitions, during which free charge carriers are generated. In the applied electric field, these carriers move and create electric current. The output signal depends on the wavelength of the incident radiation, the absorption coefficient and the value of the generated charge.

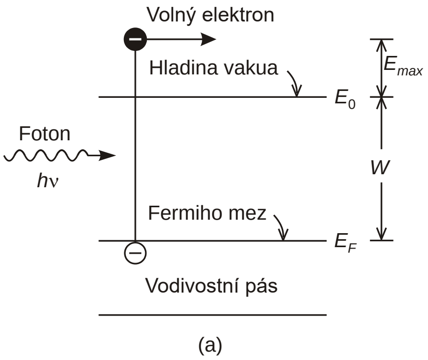
6.3.1 Photon detectors

6.3.1.1. External photoelectric effect

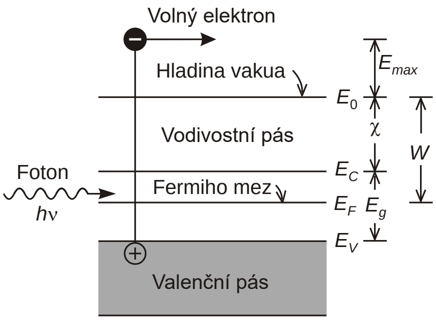
Photon detectors are based on the principle of external or internal photoelectric effect. If the energy of the incident radiation is large enough, excited electrons can cross the surface barrier and they are emitted from the detector surface into space. The image shows the energy diagram of the of the photoelectron emission process from a metal and from a semiconductor. The photon with energy $h\nu$ kicks electrons out of the surface of the metal with the final energy

$E_{max} = h\nu - W$ (1),

where W is the metal work function which expresses the energy difference between the energy of the electron in the vacuum (E_o) and the energy of electron at the Fermi energy level. The proportionality constant h is the Planck constant. If the electron lies below the Fermi energy level, the resulting energy of the emitted electron is lower. Equation (1) is a known Einstein relationship for photoemission. The lowest metal work function is about 2 eV for Cesium, therefore it is only possible to use these detectors in the visible and ultraviolet region. In the case of semiconductors, electrons are usually released from a filled valence band. For photoemission, a similar relationship as for (1) applies



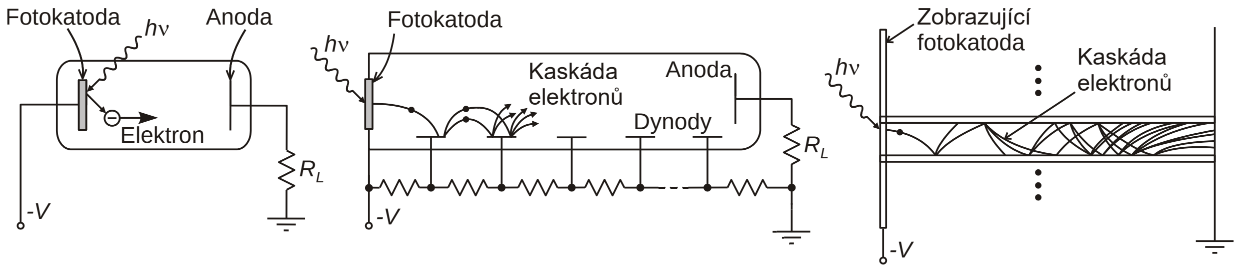
(a)



(b)

$E_{max} = h\nu - (E_g + \chi)$, (2)

where E_g is the forbidden band width and χ is the semiconductors electron affinity, which expresses the energy difference between the electron energy in the vacuum (E_o) and the bottom of the conduction band. Because the energy $E_g + \chi$ is lower than W(Cs) for certain semiconductors, these detectors can also work in the infrared region. The external photoelectric effect is used in **phototubes** (photoelectric cells)**, photomultipliers** and in **microchannel detectors.** In photocells (fig.a), electrons release from a photocathode due to the impact of photons and move towards the anode in the applied accelerating voltage. If a photocell is filled with gas, electrons ionize gas molecules along their trajectory, so that a greater number of electrons reach the anode than have been released from the cathode. In photomultipliers (Fig.b), the photons fall again on the photocathode from which they emit electrons. These are accelerated and directed by the applied electric field on the first amplifying electrode, called dynode, where the number of photoelectrons is multiplied by secondary emissions. Each subsequent dynode impact releases further electrons, and so there is a current amplifying effect at each dynode stage. Each stage is at a higher potential than the previous to provide the accelerating field. There are 10 or even more of such dynodes in the photomultiplier and the resulting multiplication effect can be even several millions.

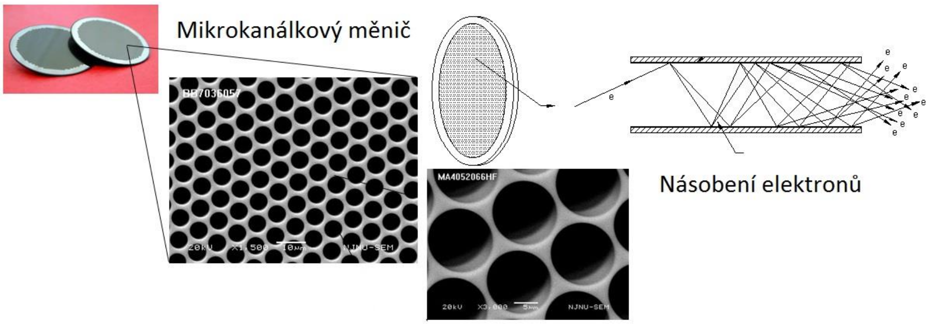


(a) (b) (c)

A special application of the external photoelectric effect is t**he night vision device** (**NVD**), also known as **night-vision goggles** (**NVG**), which is a device magnifying residual light. It is mainly used for military purposes or for guarding important objects. The night vision device may be part of a vehicle’s surveillance equipment, but there are also personal night vision devices for individuals. NVDs operate in the infrared spectrum close to visible light. NVD amplifies radiation of wavelengths between 700 nm to 1000 nm and transfers it to the visible part of the spectrum. The image is usually shown on a green phosphorus screen. The users thus see a monochromatic green image. They can therefore captured a negligible amount of the thermal radiation of usually warm objects.

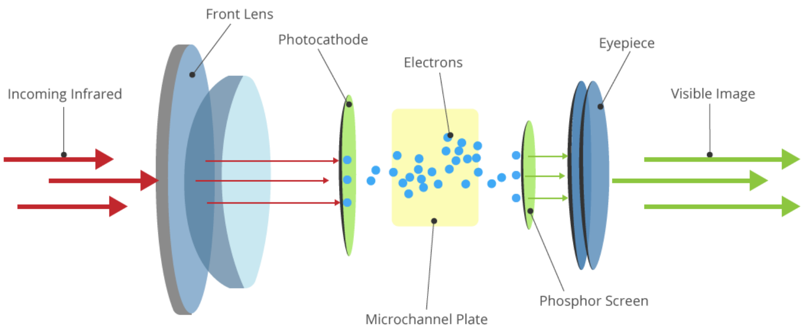
Today’s ***night vision devices of 3rd and 4th generation*** use to enhance image brightness a GaAs photocathode and a microchannel plate as a detector, which is equipped with a layer of ionic barrier (a thin film of sintered aluminum oxide) to eliminate interference. The advantage of GaAs photocathodes is not only their high sensitivity in the spectral range of 450 nm – 950 nm (partially in the visible spectrum and partially in the near infrared), but also the excellent properties at both low and high level of illumination, as well as negligible image distortion. They have been used since the 1990s.

The microchannel plate contains a high number of channels with a wall of certain thickness. The resolution (number of channels per cm^2) largely affects the night vision resolution. The thickness of the walls then results in some distortion (undetectable to the human eye in today's technologies).



Microchannel plate - MCP

The multiplied electrons eject from the multichannel plate and hit the phosphorus screen, which transforms electrons back into visible photons (a human eye-visible component of light). The color of the phosphor screen also determines the color the observer sees through night vision device.



Most often, it is monochromatic green or white. Night vision is further supplemented by a set of optical lenses which are used to direct photons and also to focus on the phosphor screen. Thanks to these lenses, the image is sharp for the observer. The whole process is described on the following figure.

Night vision device parts

1 – Objective lens – lens for night vision collects residual light and also light that cannot be seen with the naked eye. The beams of the photons are directed and focused toward the photocathode.

2 – Photocathode – Photons hit the photocathode and generates electrons due to the external photoelectric effects. These electrons are then attracted by the electric field to the phosphor screen.

3 – Microchannel plate (MCP) – electrons pass through the microchannel plate, which multiplies their number many times.

4 - Voltage supply– generates the electric field in which electrons accelerate towards the phosphorus screen.

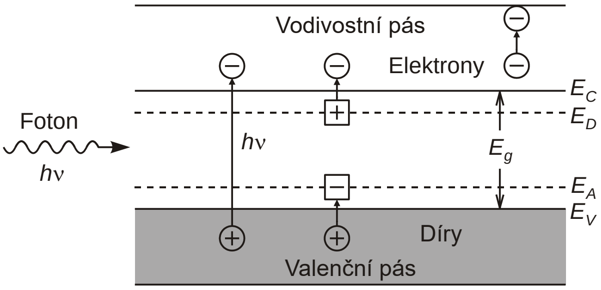
5 – Phosphor screen – When the electron image hits the surface of the phosphor screen, it causes the display to emits visible light (conversion of electrons into photons). Since the phosphor screen emits this light in exactly the same pattern and contrast as it was collected by the lens, the image seen through the night vision is identical with the image of the observed scene.

6 – Ocular lens (eyepieces) – Ocular lens is used to direct the image displayed on the phosphor screen to the eye of the observer, i.e., it is used for focusing.

7 – Exchanger case – high-quality, durable material so that the device can withstand even recoil of weapons.

For detection of greater amount of thermal radiation, which, at normal body temperatures, radiates most brightly at wavelengths from about 1 μm to about 14 μm), **thermographic camera** (also called an **infrared camera** or **thermal imaging camera**, **thermal camera** or **thermal imager**) is used, which are working based on the principle of internal photoelectric effect. Materials for cooled infrared detecting sensors represent narrow band-gap semiconductors including indium antimonide (3 μm-5 μm), indium arsenide, mercury cadmium telluride (MCT) (1 μm-2 μm, 3 μm-5 μm, 8 μm-12 μm), lead sulfide or lead selenide.

6.3.1.2 Internal photoelectric effect



The absorbed radiation causes internal electron transitions where photoexcited carriers (electrons or holes) remain inside the substance. If we apply an electric field to a semiconductor, these transitions are manifested by an increase in conductivity, which is called photoconductivity. We divide electron transitions into (fig.)

a) *Intrinsic* transition- excitation of electrons from the valence to the conduction band of a semiconductor - <u>intrinsic conductivity</u>

b) *Impurity* transition - excitation of electrons from the valence band to the acceptor level or excitation of electrons from the donor level to the conduction band - <u>impurity conductivity</u>

c) *Interaction with free carriers* - excitation of electrons between the levels in the conduction band

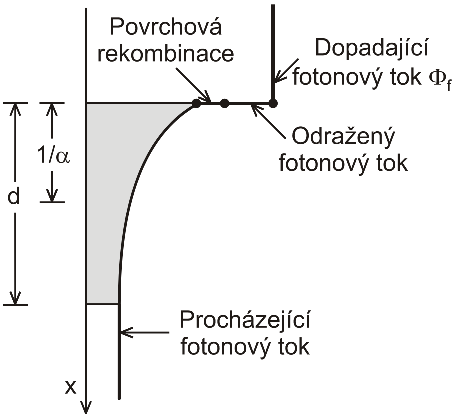
ad a) long-wavelength limit (cutoff wavelength) of sensitivity of intrinsic detector λ_g: $\lambda < \lambda_g = hc/E_g$

ad b) long-wavelength limit (cutoff wavelength) of sensitivity of impurity detector λ_i: $\lambda < \lambda_i = hc/E_i$, where *Ei* is the ionization energy of the impurity. The main disadvantage of impurity detectors is the large dark current at room temperature. Detectors must be operated at low temperatures (4 K – 30 K) to avoid thermal ionization of the centers (competes with optical ionization). Another reason is the low absorption in the impurity region - it requires a greater detector thickness. Ionized levels have a large effective cross-section for recapturing carriers. The *lifetime of the carriers* and thus the *detector sensitivity* decreases with increasing concentration of these levels.

6.3.1.3 Quantum efficiency η: (0 < η < 1)

It is defined as the probability that an incident photon will generate one electron-hole (e–h) pair.

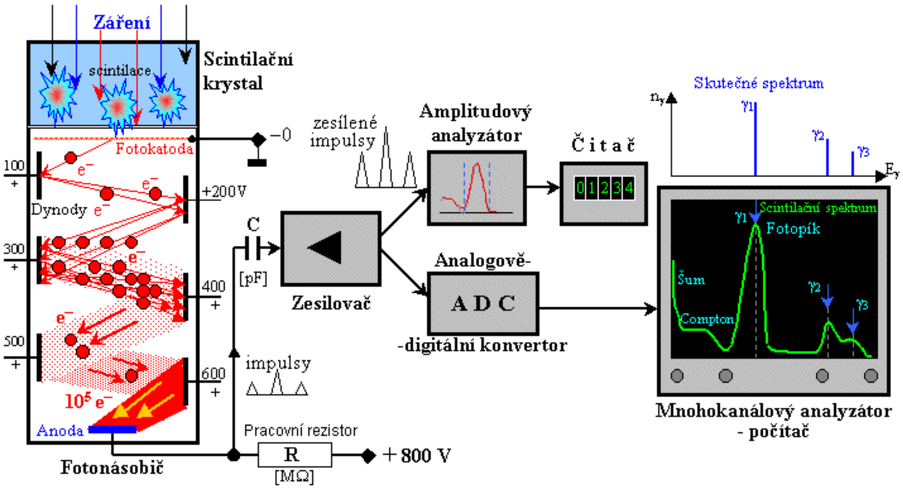
$$\eta = \Phi_{e-h} / \Phi_f = (1 - R_\lambda)s\beta,\ \text{kde}\ \ \beta = [1 - \exp(-\alpha d)] \qquad (3)$$



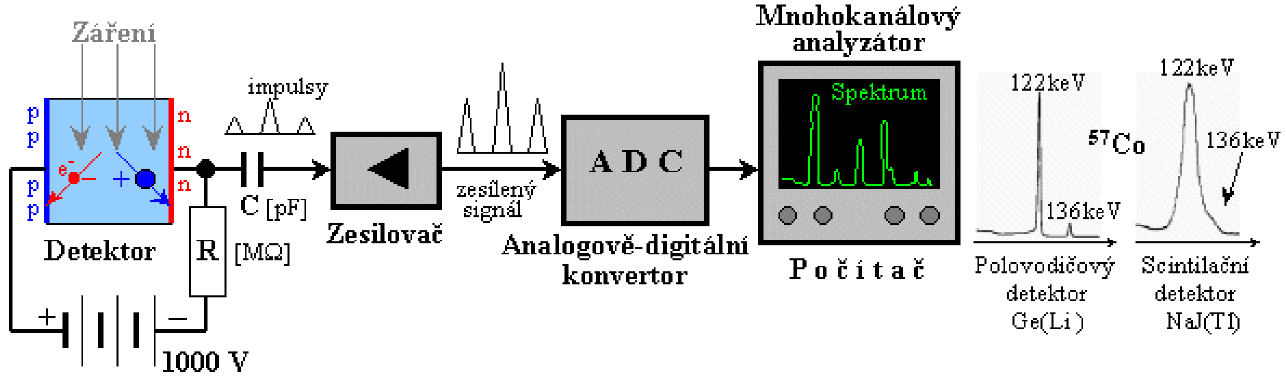
where Φ_f is the flux of incident photons (number of photons per second), Φ_{e-h} is the flux of generated e-h pairs, (1 - R_λ) is the term representing the reflection on the surface (reduction of reflectance - anti-reflection layers), s is the term representing the surface recombination of e–h pairs (reduction of surface states - surface passivation), α is the absorption coefficient and d is the thickness of the detector. The thickness of the detector must be large enough to absorb most of the radiation in the detector. The quantum efficiency η depends on the wavelength λ, because the absorption coefficient α depends on λ. For large λ [$\lambda > \lambda_g = hc/E_g$, je $\lambda_g(\mu m) = 1.24/E_g(eV)$] there is no absorption because the photon energy is not enough to overcome the band gap. λg represents here the wavelength of the absorption edge. For small λ, quantum efficiency decreases because most photons are absorbed at the surface, where due to increased surface recombination, the photogenerated carriers recombine before being diverted to the circuit.

6.3.1.4 Scintillation detectors

These detectors are based on the principle of excitation of an electron into a higher energy state by incident radiation. Return of the electron to the basic state acting as a flash of visible light. The emitted photons impact the photocathode of a photomultiplier tube (PMT) (sometimes a charge-coupled device (CCD) camera or a photodiode is used), where photons emit electrons due to an external photoelectrical effect (Einstein). PMT consists of a photocathode and secondary electrodes – dynodes. Primary electrons generated from the photocathode are accelerated and focused by an electrical potential so that they strike the first dynode of the PMT, where they generate a number of secondary electrons. Each subsequent dynode impact releases further electrons, and so there is a current amplifying effect at each dynode stage. The output signal at the anode is fed to a charge amplifier, which integrates the energy information, and an output pulse is obtained which is proportional to the energy of the particle exciting the scintillator. Scintillation detectors have a good quantum efficiency and can measure both the intensity and the energy of incident radiation. Thallium-activated sodium iodide - **NaI(Tl)**, in the form of a single crystal, is most often used today as a scintillator.



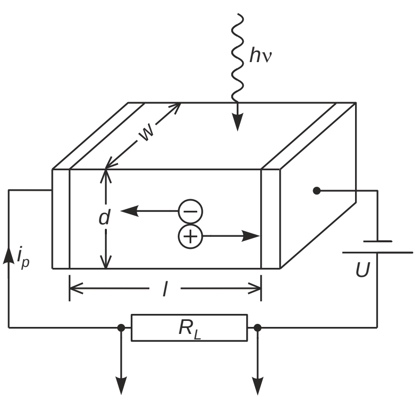
6.3.1.5 Semiconductor detectors

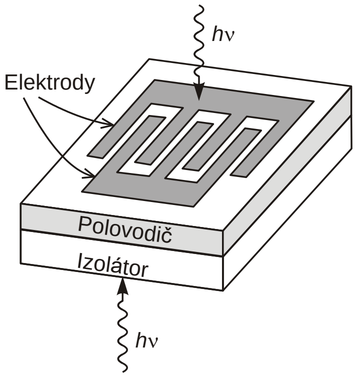


Types of semiconductor detectors

a. Photoresistor

A photoresistor (also known as a light-dependent resistor, LDR, or photo-conductive cell) is a passive component that decreases resistance with respect to the radiation (illumination) absorbed on the sensitive surface. Photoresistors are usually in the form of a plate or a thin layer. They are most often produced by vapor evaporation or sintering of a thin layer of suitable semiconductor material (e.g. CdS, CdSe for visible radiation) on a ceramic substrate. In order to achieve a larger detector area with a small transit distance between the electrodes, both electrodes has a meander shape. If the trapping centers are present in the material, the gain of the photoresistor is z >1, i.e. there is amplification of the photocurrent!





(a)

(b)

Photoresistor materials

CdS a *CdSe*: (Eg = 2,42 eV a 1,74 eV), easily accessible, high gain z = 10^3–10^4, response time is in the order of 10 ms and depends on the intensity of incident light.
CdS_xSe_{1-x}: semiconductor with variable width of forbidden gap by changing the composition Eg = 1.74 V-2, 42 eV.
PbS (in the form of thin films): (Eg = 0,37 eV), near IR detector at room temperature for wavelengths of 1 µm −2,5 µm; when cooled to −30°C up to 2 µm −4 µm. In the wavelength range around 2 µm, it is one of the most sensitive photodetectors. Response time about 200 µs; relatively high resistance (1MΩ).
InSb: (Eg = 0.36 eV), low resistivity, detection of radiation up to 7 µm, response time 50 ns. To reduce noise, it is convenient to cool the detector to 77 K, which, however, limits the range of sensitivity to 5 µm (due to Eg increasing).

$Cd_xHg_{1-x}Te$ (CMT) a $Pb_xSn_{1-x}Te$ (LTT). By changing the composition, the width of the forbidden gap can be changed continuously. CMT: Eg (77K) changes in the interval (−0,2 eV, 1,5 eV) (Eg = 0 eV for x = 0,14). This material is used to make detectors that cover the wide infrared wavelength range of 2 μm−30 μm covering three atmospheric windows. The greatest interest is in the region of 8 μm–14 μm (one of the three atmospheric windows). The great advantage of this material is the high mobility of carriers, the low intrinsic concentration and the low lifetime of the minor carriers (only a few μs).

In the past, impurity-doped detectors were used for the wavelength range of > 8μm, mainly from Ge and Si doped with various impurities. However, these detectors had to be cooled to a liquid He (4.2K) temperature.

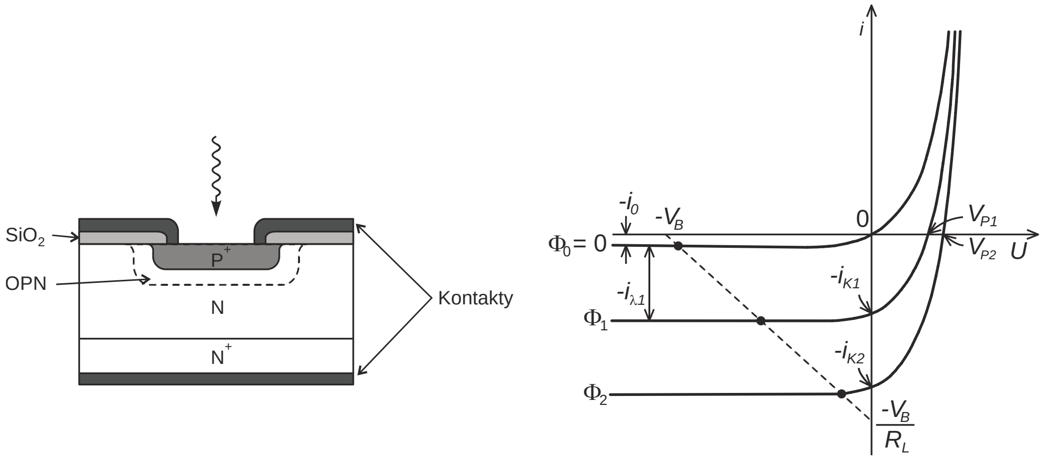
Advantages (+) and disadvantages (-) of photoresistors compared to other photosensitive components:

+ Considerable sensitivity
+ Easy to use
+ Low price
+ Possibility of application for DC and AC circuits.

- Long response time
- Considerable temperature dependence of resistance
- Time degradation of photoresistor

b. Photodiode

A photodiode is a regular planar semiconductor diode with P-N junction that allows radiation (illumination) to penetrate into the P-N junction area. If the junction is not irradiated, the volt-ampere characteristics has the same profile as the regular diode. The effect of the junction irradiation is expressed by a linear increase of the anode current with increased photon flux, when the diode is polarized in the reverse direction. So diode acts as a passive component, whose electrical resistance in the reverse direction depends on the photon flux. The photodiode responds very quickly to changes in the photon flux, in the order of $10^{-9} - 10^{-6}$ s.



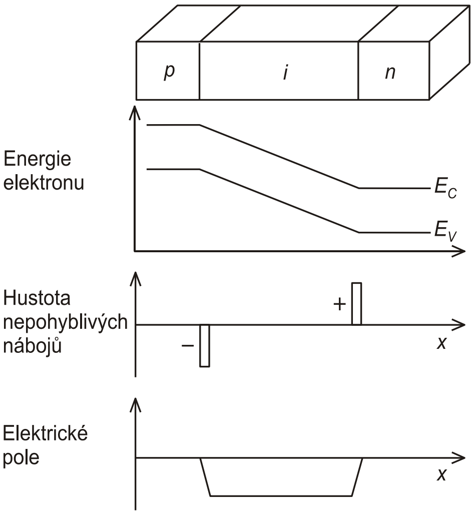
(a) (b)

The photodiode can be connected in three regimes: *no-load* (no external voltage - photovoltaic PV connection), *short-circuit* (U = 0) and with *voltage applied in the reverse direction* (photoconductivity PC connection). In the photovoltaic mode radiation generates in the depleted layer of the photodiode e − h pairs. The result is an increase of the electric field, which leads to the formation of photovoltage V_P. The advantage of this connection is that we cannot necessary used the power supply, but the detector is slower. Solar cells also work in this mode. Applying a voltage to the diode in the reverse direction, a wide space charge region (OPN) is created, where there are no free carriers. The electric field is concentrated only in OPN. After irradiation, e − h pairs are separated in OPN and both types of carriers are injected (drifted) due to the internal electric field toward the electrodes. If the trapping centers are not present in the material, the gain of the photodiode is z <1, i.e. there is no amplification of the photocurrent. At U = 0 (short circuit) regime, $J = J_k$.

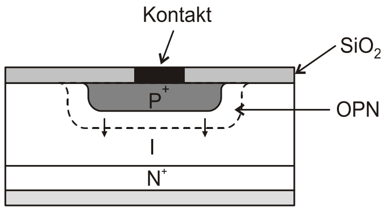
c. PIN photodiode

PIN photodiode is a diode in which the highly doped p-type and n-type regions are separated by a wide low-doped “intrinsic” region I, which is denoted by ν for the n-type region with low n-type conductivity and π for the region with low p-type conductivity. After applying reverse bias a depleted space charge region spreads over the entire intrinsic area. The PIN diodes improve the frequency characteristics and sensitivity in the long-wavelength region of the spectrum.

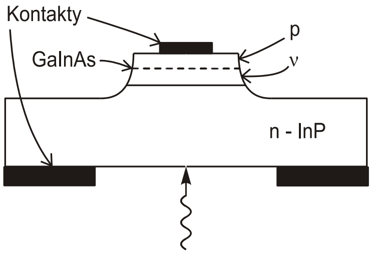
PIN photodiodes have the following advantages:
- OPN extension - photosensitive area increases, p-n junction capacitance reduces and hence the RC time constant (improving time response)
- OPN extension - the ratio between drift length and diffusion length increases. For the wavelength region of 0.4 μm – 1 μm, a high-resistance Si substrate ($10^3 - 10^4$ cm) is most often used, on which a N^+ Si layer and one side and a thin P^+ Si layer on the other side are epitaxially grown. The surface is protected by natural oxide (SiO_2), in which contact windows are formed. A sufficiently large voltage applied in the reverse direction empties the entire intrinsic area and the resulting electric field accelerates the carriers to the limit speed $10^5\ \mathrm{ms}^{-1}$, which gives a drift time $t_{\mathrm{drift}} = 10$ ps through 1 μm wide intrinsic area. Increasing the intrinsic area increases the response time and reduces the junction capacitance (~ 1 pF). GaInAs material is used for the wavelength range 1 μm – 1.6 μm. A 3 μm–4 μm $Ga_xIn_{1-x}As$ thin layer with a concentration of $10^{21}\mathrm{m}^{-3}$ and a 1 μm thin p-type conductive layer are epitaxially grown on the InP substrate. GaInAs has a high absorption coefficient ($10^6\mathrm{m}^{-1}$), therefore a thin layer is sufficient for absorption of the most radiation - high response speed.



(a)



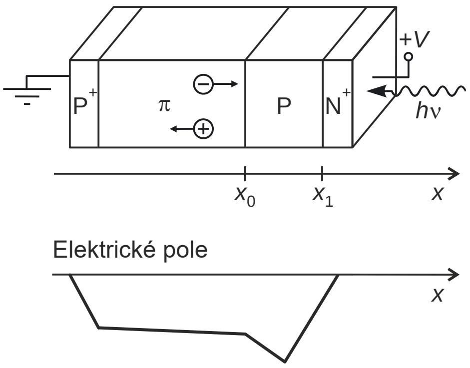
(b)



(c)

d. Avalanche photodiode - APD

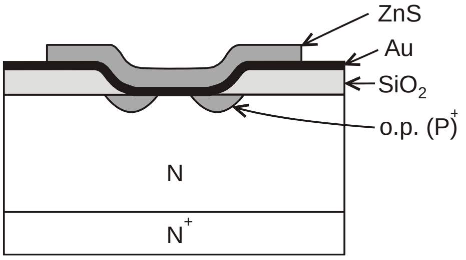
If the sensitivity of the photodiodes or PIN photodiodes is not sufficient, avalanche PIN photodiodes that show self-amplification may be used (they have a gain >1). In these components, there is a region with a high electric field, which accelerates the generated charge carriers enough that other (secondary) electrons are knocked out. Avalanche



multiplication of current carriers occurs here in which the primary current generated by the radiation increases by 2 to 3 orders of magnitude, which increases the sensitivity of the diode. The advantage of the self-gain is, however, connected with more demanding design of the photodiode. The operating voltage is high (Ge approx. 30 V, Si 300 V) and must be stabilized. This requires a high-quality stabilized voltage source, with the required voltage value varying for each diode (even from the same material). In the avalanche diode, the radiation absorption area is separated from the avalanche multiplication area. At the applied voltage, the largest electric field is in the p-type material ($x_0 < x < x_1$), where the avalanche multiplication occurs. If the radiation is absorbed in this area, both types of carriers are generated here. To the left of this region (for $x < x_0$), electrons are injected into the region of multiplication; holes are injected during absorption in the surface layer (for $x > x_1$). The process of avalanche generation of carriers is described by ionizing coefficients α_n and α_p, which indicate how many free pairs are generated by an electron resp. hole in the unit path of its drift in the crystal. The coefficients α_n, α_p depend on the electric field and their inverse values indicate the mean free path between two consecutive collisions. Rate of ionization α_n/α_p has a great importance; the bigger the difference, the better. For the correct diode function, only the injection of the one type of carrier with a higher ionization coefficient is optimal (mostly electrons).

e. Schottky photodiode (named after the German physicist W. Schottky).

It uses the rectifying effects of the metal-semiconductor interface. The semiconductor is commonly n-type silicon or GaAs, the metal gold or aluminum. Schottky diodes are most often made by planar-epitaxial technology, where a thin semitransparent metal layer forms a rectifying contact.



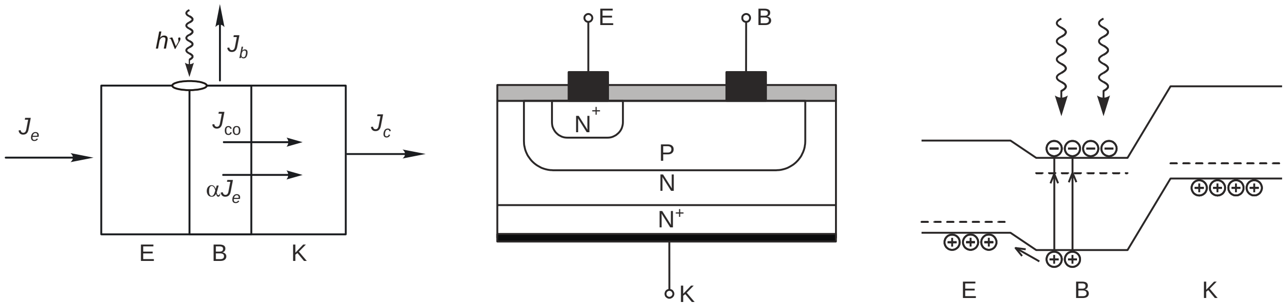
Compared to diodes with a P-N junction, the Schottky diode shows a lower forward voltage, at which the current begins to increase significantly (Si Schottky diode 0.2 V, Si diode with a P-N junction 0.6 V) and a higher reverse current (up to hundreds of nA). Only the majority carriers participate in the current conductivity in the Schottky diode. During diffusion, the minority carriers do not accumulate at the edges of the gate layer, so the time between the formation and extinction of the gate layer is considerably shorter. Schottky diodes can be switched from forward to reverse direction (from conductive to non-conductive state) in less than 1 ns. For this reason, Schottky diodes can be used at higher frequencies (up to 100 GHz.

Advantages of Schottky diodes:
+ Simplicity of production (even on materials, where the P-N junction cannot be prepared).
+ Possibility to produce diodes with a very small junction area.
+ Metal layer contributes little to the series resistance (compared to P+ layer in the P+-N junction).
+ OPN reaches the surface! The effect of surface recombination is reduced.
+ High sensitivity and speed (response time – Schottky diodes reach the parameters of PIN diodes.
+ It allows to detect He-Ne laser radiation (λ = 633 nm).
- The thickness of OPN is small! Limitation of diode sensitivity range (for Si diodes $\lambda < 1\,\mu$m).

The above-mentioned parameters predetermine to use the Schottky diode in extremely fast switching circuits in computer technology, radar devices or for rectifying small voltages with a frequency of up to hundreds of GHz.

f. Phototransistor

Phototransistor is a light-sensitive transistor and is used as an optical receiver, where the intensity of the incident radiation affects the magnitude of the collector current. The phototransistor, like an avalanche photodiode, is an element with internal amplification. It represents a bipolar transistor, whose emitter-base (E-B) junction is connected in the forward direction and base-collector (K) junction in the reverse direction. There is a transparent window at the base-emitter junction, in which a lens is inserted and through which the E-B junction is illuminated. The base electrode is generally not connected and the transistor is controlled (switched on / off) by the intensity of the light radiation. The external source is connected between the collector and the emitter so that the collector junction is polarized in the reverse direction.

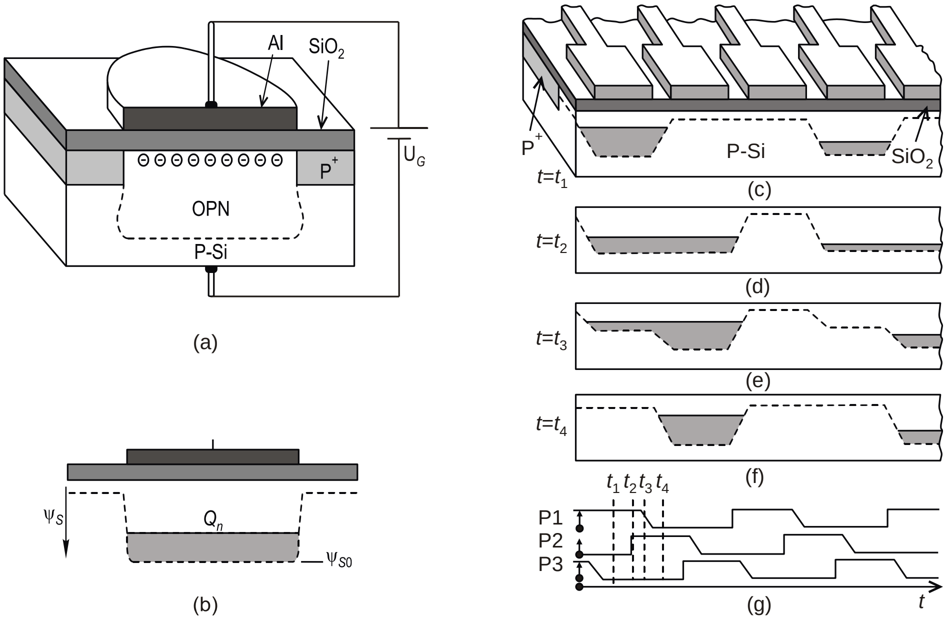


(a) (b) (c)

Since the E-B junction is in the forward direction, the majority of the charge carriers (here electrons) pass from the emitter to the base. The electrons move in the base by diffusion create the gradient of the electron concentration in the base with the highest value just close to the E-B junction. Since the base is very thin, only a small part of the electrons recombine. Due to the small thickness of the base, the most of electrons reaches the B-K junction. This junction is polarized in the reverse direction for the majority holes, but forward for the minority electrons that pass into the collector. By illuminating the E-B junction, there is an injection of minority carriers into the base region like in the conventional transistor. This injection increases the current at the B-K junction (in the circuit). After the generation of e − h pairs, the minority carriers are drained by the electric field of the junction and together with the majority carriers they cause the amplification of the photocurrent. The sensitivity of phototransistors is usually higher than that of the PIN diode, but the response time is longer because the carriers diffuse through the base area. Therefore, the effort to make the base thin. However, the base must not be too thin to limit the ability of the base to absorb radiation. For low radiation intensities, the phototransistor shows a decrease in gain, which is caused by recombination of the generated carriers in the base region or in the E-B junction.

g. Charge Coupled Device (CCD)

Charge Coupled Device is a charge-coupled structure in which the charge generated by incident radiation is deposited into surface potential wells formed under closely arranged MIS (metal-insulator-semiconductor) gates. The basic element of the CCD chain is the MIS structure consisting of a p-type Si plate coated with an insulating layer of natural oxide SiO_2 and with both-side metal contacts. If a positive voltage is applied to the MIS structure, a depleted layer forms under the gate contact. With a suitably arranged sequence of voltage pulses applied to the chain of gates, the charges are shifted along the chain and output is converted to serial sequence of pulses (shift register). This allows to monitor, transmit and process both digital and sampled analog signals.



When incident light in form of photons impacts on the surface of the CCD structure, a charge in the potential wells is created and stored. Potential wells prevent the free lateral movement of electrons, so the charge is accumulated in the well. Each such potential well represents one pixel of a CCD sensor.

The magnitude of the captured charge is mainly affected by the intensity of the incident light and the time we leave the CCD chip exposed to light. The captured charge must be replaced after a period time and converted into an electrical signal, otherwise the potential well could overflow. The CCD image sensors therefore contain a matrix of pixels (potential wells), in which the charge is gradually shifted from one well to the next one toward the edge of the chip, where the charge is converted to a voltage signal (see Fig. CCD2).

Problem: The CCD chip must not be irradiated while the signal is shifted from individual lines, so it should be shielded to avoid charge interference.

From this point of view, there are 3 principles:

"FF - Full Frame" – The whole CCD chip is exposed to light, i.e. all available pixels. A mechanical shutter covering the chip is required to shift / move the charge.

"FT - Frame Transfer" - The chip is divided into an area still exposed to incident light (Imaging Area) and a permanently shielded area (Storage Area) at the boundary of the chip. At a given moment, the whole charge created in the CCD chip matrix is very quickly move into the Storage area, and from it is already possible to slowly move the charge for digital processing throughout the scanning time of the next image in the imaging area of the chip (approx. tens of ms). The principle is also referred to as an electronic shutter and is one of the cheapest but qualitatively worst solutions.

"IT - Interline Transfer" - the principle is similar to FT. This is also an electronic shutter, but differently made. Next to each light-sensitive column in the CCD matrix, there is a neighborhood non-light-sensitive (shielded) column that holds and shifts the charge to convert to a signal during the next image scan.

Miniature lenses focusing light are implemented on the surface of the CCD sensor to avoid as much as possible the loss of incident light caused by the photons impacting into the Storage area. Lenses refract light and direct it only to the Sensitive area. These principles fully valid for monochromatic sensors, you could say, black and white CCD sensors. Today, we mostly need to scan the image in color. Therefore, color filters in the form of so-called Bayer masks are placed above the image cells of CCD sensors. Bayer masks consist of RGBG filters, where the G - green filter - is doubled due to the higher sensitivity of the human eye to this color. Each of the filters transmits light only by specific wavelength (color). Individual pixels then detect only one color from the RGB or CMY color model and the final color is then interpolated - calculated from several surrounding pixels. Typically, the color is calculated from the four pixels in the 2 × 2 matrix, or from the eight pixels surrounding the current pixel in the 3 × 3 matrix. So the resulting photo is stored in the RGB color space with 256 brightness levels – which will give, after multiplying the known theoretical 16.7 million shades of 8-bit JPEG color scheme. It is currently the predominant design.

Complementary Metal Oxide Semiconductor (CMOS) technology, which uses planar technology (photolithography) for the production of semiconductor components controlled by an electric (field-effect transistor - FET), is also used in the production of the light-sensitive CCD chips using in digital cameras. The CCD chip transmits the photo generated signal from pixel to pixel and transforms it into a voltage in the reading circuits. CMOS chips convert the signal to a voltage right inside each pixel. Only one supply voltage is needed for their operation and the electrical consumption of these sensors is very low. The technology itself is relatively cheap and well mastered. Most computer integrated circuits, including processors, are also manufactured by this technology. CMOS chips are divided into passive (Passive Pixel Sensor) and active (Active Pixel Sensor). Passive chips are made up only of photodiodes and the signal is not modified in any way. Today's widely used active chips have in addition an amplifier and a noise-cancelling circuit for each cell. They have less noise than passive chips, on the other hand, each of these amplifiers can amplify a little differently, thus achieving higher noise than a CCD.

The claim that CMOS sensors are more suitable for cheaper cameras due to their lower quality and sensitivity, is no longer true. Technological progress in this area is incredibly fast and currently CMOS technology dominate the market for sensors for digital cameras and mobile phones.