The Role of Physics in the Development of Microelectronics (Part 4)

Eugen Ştefan LAKATOŞ*

Abstract

The purpose of this paper is to present the basic contributions of Physics in the development of Microelectronics, a science field found at the edge between Physics and Electronics. We will show the road followed by the 20th century science and technique, beginning with 1900, from the rise of quantum theory, elaborated by Max Plank, continuing with the invention of bipolar transistor, up to the manufacturing of the first microprocessor and of the first integrated memory, this way opening the way to the construction of the first personal computer. In this paper we will present a short review of the main Physics chapters which contributed to the development of the Microelectronics field, of the main theoretical concepts and of physical models which led to the development of the semiconductor devices manufacturing technologies industry. We will present, from the point of view physical phenomena and of modeling, the basic processes in integrated-circuit fabrication: oxidation, epitaxial growth, ion implantation and solid-state diffusion. We will describe the physical models of these processes, models which led to the design of the process flows from Microelectronics industry. These physical models are also used in simulation software of technological processes. For two of the basic processes, oxidation and solid-state diffusion, we will shortly describe the characterization methods, emphasizing once more the role of Physics in Microelectronics field. We will also refer, using examples, to the main process flows used in integrated circuits structures manufacturing that led to the development of Microelectronics.

Keywords: microelectronics, physical model, technological process, process flow

5. Technologies for the execution of *pn* junction semiconductor structures

Figure 5.1. shows the important stages of basic technologies for the manufacture of pn semiconductor junctions. The order of presentation illustrates the timeline of development for these technologies [10].

The alloyed method is a particular case of the general method of heating a semiconductor sample in contact with a contaminating substance. An example of application of this method is illustrated in figure 5.1.a.

A small piece of aluminum is placed on the surface of a wafer of semiconductor material (n-type monocrystalline silicon). The system is heated until eutectic Al-Si system temperature (T=580 °C), under conditions that exclude the oxidation processes (hydrogen atmosphere). At the temperature of alloying, the molten aluminum attacks silicon really fast, forming a liquid solution of silicon and aluminum.

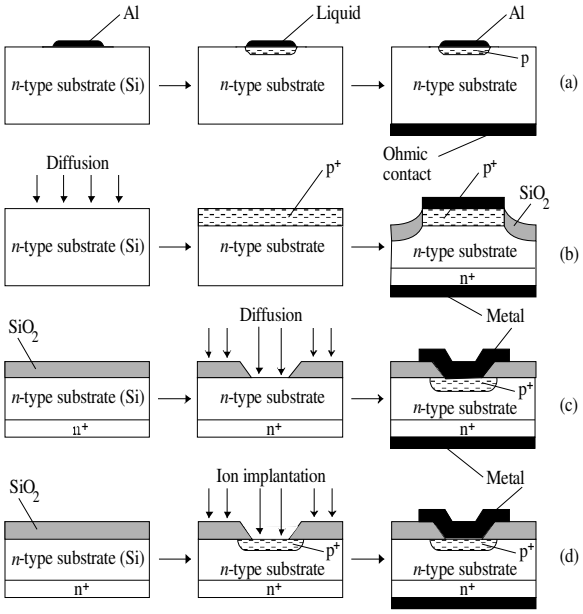

Figure 5.1. Technologies for the execution of the pn junction: (a) allied junction; (b) diffused junction; (c) planar junction; (d) planar junction obtained by ion implantation (adapted after Sze [10])

The silicon continues to be dissolved in the aluminum until the saturation limit is reached to the temperature at which the process takes place. The heating lasts for 1-15 minutes, at a temperature between

*Eugen Şt. LAKATOŞ: Univ.Prof.Doc.Phys., Universitatea „Valahia" Târgovişte, Bd. Carol I, nr. 2, Târgovişte, România; e_lakatos@yahoo.com

600 °C and 850 °C, according to the type of device that is intended to be obtained (typically semiconductor junctions with the breakdown voltage between 3 V and 6 V), then the system is cooled until ambient temperature. During cooling, the aluminum-silicon solution crystallizes in the same type of lattice as the solid silicon. Thus appears a p-type region (the aluminum in silicon is acceptor), on the n-type substrate. The pn junction if realized at the border of the p and n regions. The profile of impurities produced through the alloying method is highly dependent on the parameters of the alloying cycle (temperature, time, speed of the cooling system) and is difficult to control, this being a major disadvantage of this method. Moreover, the method is no longer used today. The method was used in the manufacture of the first germanium based diodes and transistors and continued to be a series production method until the new emerging applications required the performance enhancement of these devices [1] [8]. The method could not be applied when obtaining desired devices performance in terms of breakdown voltage, both in terms of technical reproducibility and in terms of breakdown voltage values.

The diffusion in solids was the technology which has removed the main disadvantage of the alloying method: the imprecision of the control on the impurity profile, therefore of the location of the pn junctions. This technology has been developed since 1956 and led to the development of the method of achieving diffused junctions. The method of MESA type diffused junctions is illustrated in figure 5.1.b.

P-type impurities are diffused in the volume of the semiconductor layer (n-type silicon). After the diffusion, a chemical etching is performed through a metal mask (Ni-Au, for example), a MESA profile structure being produced (concave shape of the edge of the structure).

An improvement in the degree of control regarding the lateral geometry of the diffused junctions, was achieved with the development of planar technology, which can be defined briefly as representing the diffusion through an oxide mask, on one side of the semiconductor wafer (figure 5.1.c).

This technology combines the performances of the diffused junctions with the possibility of growing of the silicon dioxide for the precise definition of the geometry of the device. A layer of silicon dioxide is grown on the face of the silicon wafer (thickness 0,5 μm). By the help of the lithographic techniques (photo-lithography, X-ray lithography, ions beam or electrons beam) windows are cut out in the layer of oxide, obtaining the desired geometric configurations of the structure. Through these windows impurity diffusion occurs (the source being liquid or gaseous). This technology led to the development of integrated circuit manufacturing technology. The benefits of implementing this technology were: increased electric performance of the devices (the sensitivity to external factors decreased while the stability of electric characteristics increased), the accomplish-ment of the MOS structures (metal-oxide-semiconductor) and of the field-effect transistors with MOS structure (MOSFET). A major economic advantage also exists, which led to the development of planar technology, specifically the cost of achieving of an integrated circuit. If an oxide window size of 0,2 μm^2 x 0,2 μm^2 is considered, the number of junctions with this size, having the distance between them of 0,2 μm^2, on a wafer having a diameter of 30,48 cm (12 inches) is of the order of 1012. The simultaneous manufacturing capacity of thousands of devices was really the key of economic advantage of planar technology.

For the achievement of the first planar transistors and of the first integrated circuits in the manufacturing process were used only photo-masking and diffusion steps. The integrated circuits made only through diffusion presented severe limitations compared with the discrete circuit components. For example, for the triple-diffused transistor, the collector region is formed through an n-type diffusion in a p-type wafer (substrate). One of the disadvantages of this structure was given by

[1] The allied junction method was invented by John Saby (General Electric Laboratories) and developed for mass production by RCA. In 1951 the allied junction transistor of pnp type was created, by alloying indium on both sides of a thin layer of n-type germanium [50 Years of Electron Devices, 2002, IEEE].

the high value of the collector series resistance, due to low doping concentration substrate, resulting in a region of high resistivity. Therefore, the substrate under the collector junction was required to have a high concentration of doping. This way, the triple diffused transistor structure was to have a additional substrate, of higher concentration, (the buried layer, (§ 8.1, Part 6), that will become a low-resistance path for the collector current of the transistor. As such a concentration profile cannot be achieved only through diffusion, the epitaxial growth technique was developed [15].

The industrial method to obtain epitaxial layers is the vapor-phase growth or CVD process (Chemical Vapor Deposition), a technology developed since 1957. With this technique may be deposited: metal (e.g., aluminum), insulating (e.g., silicon dioxide) and semiconductor (e.g., silicon) layers. The most important process is the growth of a monocrystalline semiconductor layer on a monocrystalline substrate, having the following properties: the same semi-conductor type (crystal structure of the grown layer is the same as that of the substrate); the same conductivity type or a different type; the same doping or different doping.

Figure 5.1.c illustrates the n-type epitaxial layer, grown on the n+-type substrate. The word epitaxy, derived from the Greek words epi = on and taxis = arrangement, describes the technique of growing a monocrystalline substrate, of another layer, also monocrystalline, having the crystal structure identical to that of the substrate. This ensures monocrystalline continuity at the interface between the two layers.

Basically, the physical-chemical epitaxy process consists in the transport of atoms from a solid, liquid or gaseous phase at the surface of a monocrystalline substrate and their arrangement in ordered positions after the surface diffusion, in such a way that the grown monocrystalline layer continues the crystal structure of the substrate. The concentration of impurities in the epitaxial layer can be controlled independently of the concentration of substrate, having the possibility to be higher or lower. The thickness of epitaxial layers used in the manufacture of integrated circuits range from 1 μm la 20 μm [15]. Epitaxy is used in some CMOS processes and in most BiCMOS processes.

In the case of using silicon as the base semiconductor material, the growth of the epitaxial layers is performed by placing the silicon wafers in an atmosphere of silicon tetrachloride ($SiCl_4$) or silane (SiH_4), at high temperature. A chemical reaction takes place resulting in silicon which is deposited on the surface of the wafer. The basic chemical reaction (reduction in the vapor phase of the silicon tetrachloride), is:

$$SiCl_4 + 2H_2 \rightarrow Si(solid) + 4HCl(gas). \quad (5.1)$$

Typically, the silicon layer is grown at a rate of 1 μm/min, at 1200 °C.

If the conditions during the technological process are appropriately controlled, the grown silicon layer has a crystalline structure with few defects. Such a layer is suitable as starting material for the manufacturing of bipolar transistors.

The use of the epitaxial growth has meant an important step in planar technology, especially in the technology of integrated circuits, the epitaxial layer being the layer in which the entire technological process is achieved. The quality of epitaxial layers is, usually, better than the quality of the diffused layers. The epitaxial growth technology has been developed since 1960.

Figure 5.1.d shows the pn junction formed by ion implantation technique. Used mainly for obtaining shallow junctions (from less than 0,1 μm to about 0,6 μm), with reduced side scattering, this technique has imposed itself as a method which gives the most precise control of the impurity profile in the integrated circuit technology, especially large scale integrated MOS circuits (precise control over total dose, depth profile and area uniformity). In some cases, ion implantation technique is indispensable if compared with the other methods of obtaining the pn junction (the diffusion and the doping during the epitaxial growth). The ion implantation is a technique that allows the direct introduction of impurity atoms in the volume of semiconductor wafers.

Ion implantation is an alternative to the solid diffusion process for controlled doping of a semiconductor substrate. The method

consists in accelerating the ions (from an ions source) to a velocity high enough (equivalent to a kinetic energy) to enable the penetration of the ionized atoms in the volume of the semiconductor substrate, to certain depths. Under certain controlled conditions, the ions that penetrated the volume of the semiconductor occupy positions in the lattice so that they can act as donor or acceptor impurities. Very important is the fact that the process takes place at ambient temperature and is a very clean process (it is carried out in a vacuum atmosphere). The lattice structural defects induced by the process, are removed by a subsequent heat treatment conducted at a temperature of 700 °C or lower. The method of doping semiconductors by ion implantation not being thermal process, has a number of advantages. The subsequently thermal treatments applied use temperatures much lower than those used in the doping by diffusion. Thus the impurity concentration profiles obtained from previous diffusion processes are not affected. An important feature of doping the semiconductor with this method is that the introduction of a certain concentration of impurities is not dependent on the properties of the substrate, but on the external system (the nature and energy of the incident ions). The concentration of impurities introduced by the ion implantation method is not dependent on their solubility in the semiconductor to be doped.

In the following sections are presented the physical models for the main technological processes that make up the manufacturing technological flows of microelectronic devices and integrated circuits. The process of diffusion will be described in two separate sections (§ 6, 7, part 5).

5.1. Thermal oxidation

The development of the thermal oxidation process has a twofold significance: on the one hand, the possibility of developing planar technology, on the other hand, the possibility of using silicon as the basic semiconductor material for semiconductor devices, instead of germa-nium, due to easiness of the process of making silicon dioxide (SiO_2). In planar technology, the silicon dioxide is used for several purposes: to build the required masks necessary during operations of diffusion or implantation, that result in *windows* for the doping species crossing, in order to achieve *pn* junctions (the oxide layer acts as a screen in the path of dopant species); to prevent impurity exdiffusion process during the heating processes; as a mask for the selective etching of some layers; for the purpose of passivation and protection of structures; as the basic building block of the *MOS* devices; for the dielectric insulation of a circuit components [16].

There are three methods for obtaining silicon dioxide: *vapor-phase deposition*, *electrochemical oxidation* (anodization) and *thermal oxidation*. The most used in industrial practice is the thermal oxidation, due to of the good properties of silicon dioxide obtained by using this method. For proper functioning of electronic devices and circuits the following conditions have to be met: precise control of the *thickness* of the SiO_2 film, the SiO_2 purity control, the knowledge of the *oxidation process kinetics*. Also, the phenomena which occur in the oxidation process are important to determine the electric characteristics of the planar silicon devices [8].

Increasing silicon dioxide films on silicon wafer surfaces is an important step of the planar technology and represents the main feature of this technology. Typically, any stage of creation of the doped layers in a controlled manner, by one of two methods, diffusion or ion implantation, is preceded by an oxidation step and a step for opening windows into the oxide, through photogravure method. The thermal oxidation method has two embodiments: *dry* or *wet* thermal oxidation. Basic chemical reactions are:

$$\mathrm{Si(solid)} + \mathrm{O_2(dry\ oxygen)} \rightarrow \mathrm{SiO_2(solid)}, \quad (5.2)$$

or

$$\mathrm{Si(solid)} + 2\mathrm{H_2O(steam)} \rightarrow \mathrm{SiO_2(solid)} + 2\mathrm{H_2}, \quad (5.3)$$

The thermal oxidation process consists in exposing the silicon wafer to an oxidizing agent (dry oxygen or water vapors). The principle of the thermal oxidation process is the formation of a covalent bond between oxygen and silicon. Using experimental methods it was established that the

oxidation occurs by the diffusion of the oxidizing species through the oxide to the Si-SiO_2 interface, where one of the chemical reactions shown above takes place. When a silicon dioxide film of thickness *d* grows over the silicon surface, a layer of silicon of thickness 0,45·*d* is consumed (figure 5.2) [16].

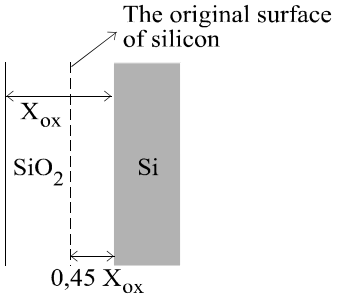

Figure 5.2. The oxide layer grows by the consumption of a silicon layer (adapted after Bârsan [16])

5.1.1. The linear-parabolic model of the oxide growth

The linear-parabolic model (Deal-Grove model [8]) is valid for process temperatures higher than 700 °C, the pressure of oxidizing agent between 0,2 atm and 1 atm, the oxide thickness greater than 300 Å [16].

There are three stages of oxide growth kinetics (figure 5.3.) [16]:

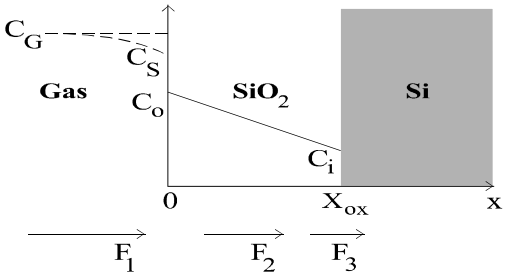

Figure 5.3. Model for thermal oxidation of silicon (adapted after Bârsan [16])

1) *the transport of the oxidizing agent* from the gas volume at the oxide-silicon interface, with the flow F_1:

$$F_1 = h_G(C_G - C_S), \tag{5.4}$$

where: h_G is *the gas phase mass transfer coefficient*, C_G is the oxidant concentration in the volume of the gas (O_2 for dry oxidation or O_2 in water vapors, for wet oxidation), C_S is the concentration of oxidant at the surface of the oxide (the oxide-gas interface);

2) *the diffusion of the oxidizing agent* through the *existing* oxide layer, with the flow F_2. According to *Fick's first law* is obtained:

$$F_2 = -D\frac{\mathrm{d}C}{\mathrm{d}x} = D\frac{C_o - C_i}{X_{ox}}; \tag{5.5}$$

where: *D* is *the diffusion coefficient* of the oxidant into the oxide layer, C_o is the concentration of the oxidant at the outer surface of the oxide (oxide-gas interface), C_i is the concentration of the oxidant at the interface oxide-silicon, X_{ox} is the oxide thickness.

3) *the reaction of the oxidizing agent* with the silicon, at the SiO_2-Si interface, occurs at a rate proportional with the concentration of oxidant:

$$F_3 = k_S C_i, \tag{5.6}$$

where k_S is the constant of the reaction rate at the Si-SiO_2 interface.

Considering valid Henry's law, which states that, at equilibrium, the concentration of a species in a solid is proportional to the partial pressure of the same species in the environment, the following relations are obtained:

$$C_o = Hp_S, \quad C^* = Hp_G, \tag{5.7}$$

where: *H* is the Henry's law constant, p_S is the partial pressure of the oxidant at the surface of the oxide, C^* is the equilibrium concentration in the oxide layer, corresponding to the partial pressure of the oxidant in the gas volume, p_G.

Because, according to the perfect gas law:

$$C_G = \frac{p_G}{kT}, \quad C_S = \frac{p_S}{kT}, \tag{5.8}$$

relation (5.4) becomes:

$$F_1 = \frac{h_G}{HkT}(C^* - C_o) \equiv h(C^* - C_o), \tag{5.9}$$

where *h* is the coefficient of mass transfer in the gas phase *depending on the concentrations from solid* [8].

Imposing the steady state condition, $F_1 = F_2 = F_3$, from the relations (5.5), (5.6) and (5.9) the following expressions are obtained:

$$C_i = \frac{C^*}{1+\frac{k_S}{h}+\frac{k_S X_{ox}}{D}}, \quad (5.10)$$

$$C_o = \left(1+\frac{k_S X_{ox}}{D}\right) C_i. \quad (5.11)$$

The limit forms of the two concentrations are:

$$\lim_{D\to 0} C_i = 0, \lim_{D\to 0} C_o = C^*$$

$$\lim_{D\to \infty} C_i = \lim_{D\to \infty} C_o = \frac{C^*}{1+\frac{k_S}{h}}, \quad (5.12)$$

and correspond to the case *controlled by diffusion* (*D* has a low value, $D << k_S X_{ox}$; the oxidation rate depends on the flow of oxidant at the SiO_2-Si interface), respectively to the *reaction controlled* case (*D* has high value, $D >> k_S X_{ox}$; the oxidation rate depends on the reaction rate at the SiO_2-Si interface). The distributions of the concentration of the oxidizing agent in the oxide layer, for the two limit cases are illustrated in figure 5.4. [8].

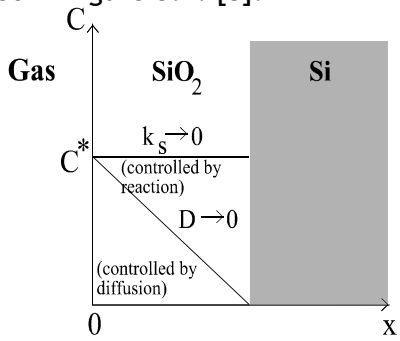

Figure 5.4. Distribution of oxidizing species in the oxide layer for the two limit cases oxidation model (adapted after Grove [8])

In order to determine the dependency relationship of the oxide layer thickness, depending on the time and on the growth constants, N_1 is defined as being the number of oxidizing molecules needed in order to form a volume oxide unit. The flow of oxidant from the oxide-silicon interface is written, according to the relationships (5.6) and (5.10):

$$F_3 = \frac{k_S C^*}{1+\frac{k_S}{h}+\frac{k_S X_{ox}}{D}} = N_1 \frac{\mathrm{d}X_{ox}}{\mathrm{d}t}. \quad (5.13)$$

The solution of the differential equation (5.13), with the initial condition $X_{ox}(t=0)=X_{in}$ (X_{in} representing the existing oxide layer thickness on the silicon wafer, grown at an earlier stage of oxidation), is of the form [16]:

$$X_{ox}^2 - X_{in}^2 + A(X_{ox} - X_{in}) = Bt, \quad (5.14)$$

where:

$$A = 2D\left(\frac{1}{k_S}+\frac{1}{h}\right), \quad B = \frac{2DC^*}{N_1}. \quad (5.15)$$

Solving equation (5.14), the following result is obtained:

$$X_{ox}(t) = \frac{A}{2}\left[\sqrt{1+\frac{t+\tau}{A^2/4B}}-1\right], \quad (5.16)$$

where τ is the time required growth of the oxide layer thickness X_{in}:

$$\tau = \frac{X_{in}^2 + AX_{in}}{B}. \quad (5.17)$$

The limit case of the control through *diffusion* corresponds to the high oxidation times (higher thickness of the oxide layer), so that $t >> A^2/4B$ ($t >> \tau$). In this case we obtain a *parabolic dependence*:

$$X_{ox}^2 = Bt, \quad (5.18)$$

where *B* is called *the parabolic growth constant*.

The opposed limit case, of the control through reaction, corresponds to the lower oxidation times, so that $(t + \tau) << A^2/4B$. In this case, a linear dependence is obtained:

$$X_{ox}(t) = \frac{B}{A}(t+\tau). \quad (5.19)$$

where the ratio *B*/*A* is called *the constant of linear growth*.

The constants *B* and *B*/*A* depend on: the oxidizing atmosphere, the pressure of the oxidant, the temperature at which the oxidation process takes place, the crystallographic orientation of the mono-crystalline silicon, the doping of the silicon substrate, the presence of foreign matters. In the specialized literature there are graphs showing the dependence of the two constants on temperature, for the two variants of thermal oxidation used, wet or

dry. Particularly, it had been found that the amount of B is proportional to the partial pressure of the oxidant into the gas, which indicates the use of Henry's Law to be justified. The constant of the linear increase B/A depends on the crystallographic orientation of the silicon, which at its turn depends on the speed at which silicon atoms are incorporated into the oxide. This speed depends on concentration of silicon atoms at the surface, which is depending on the crystallographic orientation (the orientation substrate <111> oxidizes faster than the substrate orientation <100>).

The dependencies graphs B, B/A=f(t, T), existing in the specialized textbooks, can be used as an indication only, because each manufacturer sets its own technology, by determining the process parameters and keeping them at constant values (*the calibration* process, a required stage after a technological process simulation).

The described Deal-Grove model is well verified experimentally. For this reason it is used in the technology of semiconductor devices and integrated circuits based on monocrystalline silicon.

Experimentally it was found that, in the case of oxidation of a wafer that was initially clean (oxide-free), $X_{in} \cong 200$ Å, for the dry oxidation and $X_{in} = 0$ for the wet oxidation.

The B and B/A constants verify experimental relationships of the type [16]:

$$B = C_1 \exp\left(-\frac{E_1}{kT}\right), \quad B/A = C_2 \exp\left(-\frac{E_2}{kT}\right). \qquad (5.20)$$

For low-doped monocrystalline silicon, with a crystallographic orientation <111>, the parameters of the relationships above have the values:

- for dry oxidation:

$C_1 = 7{,}72 \cdot 10^2$ μm^2/h, $C_2 = 6{,}23 \cdot 10^6$ μm/h, $E_1 = 1{,}23$ eV, $E_2 = 2$ eV;

- for wet oxidation:

$C_1 = 3{,}86 \cdot 10^2$ μm^2/h, $C_2 = 1{,}63 \cdot 10^8$ μm/h, $E_1 = 0{,}78$ eV, $E_2 = 2{,}05$ eV.

For the whole range of temperatures used in the process of oxidation, it is noted that wet oxidation is faster than dry oxidation.

5.1.2. Methods for measuring the thickness of the oxide layer

Generally, the methods for the estimation and measurement of the thickness of the oxide layer are based on optical phenomena, the polarization and the interference of light.

The most simple and at the same time, the most rapid *estimation* of the thickness of the oxide layer is *the method of colors*. Due to the interference of light waves, the oxidized wafer, viewed perpendicularly, in a white light, has a certain color, depending on oxide thickness. As the colors are repeated, in approximately identical groups, at about 2500 Å, the thickness of the oxide is determined by comparing the color of the wafer with those of some *yardstick* wafers, to which the oxide thickness was measured using another, more precise method. To use this method, it is required to know the area in which the oxide thickness is to be measured.

The drop of hydrofluoric acid (HF) *method* is used as an alternative to the method of colors. A drop of hydrofluoric acid placed on the surface of the oxide, attacks the oxide, forming concentric rings of different colors (depending on the thickness of the oxide), the result being the appearance of a sequence of colors corresponding to the color spectrum of the white light, from the original surface of the grown oxide, to the surface of silicon (figure 5.5).

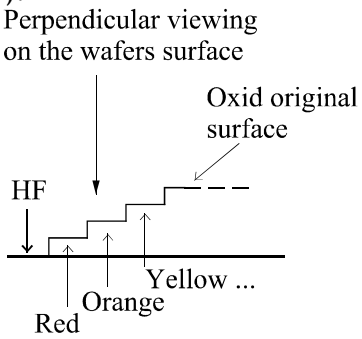

Figure 5.5. The HF drop method for measuring the thickness of the oxide layer

The profilometer method consists in measuring a step, performed in the oxide layer (by hydrofluoric acid etching, through a photoresist mask), by the movement of a calibrated probe (figure 5.6). This method it is used for measuring thicker layers.

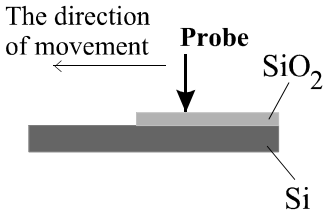

Figure 5.6. Profilometer method for measuring the thickness of the oxide layer

The optical interference method is based on the phenomenon of interference of light on a blade parallel plane faces. A light beam with a fixed wavelength, undergoes reflection and refraction phenomena at the encounter of a plane parallel blade (oxide layer in this case). The difference of *optical path*[2] between the reflected beam and refracted beam, depends on the thickness and index of refraction of the oxide layer. Figure 5.7 illustrates the principle of optical interference method on the parallel plane blade.

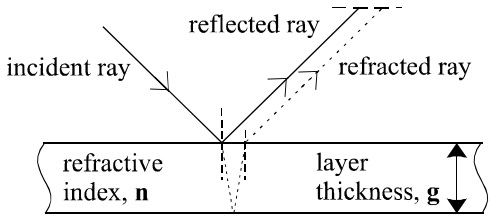

Figure 5.7. The principle of optical interference on the blade parallel plane faces to measure the thickness of the oxide layer

Ellipsometry is the method used for measuring thin oxide layers. The principle of the method is the change of the polarization of a laser beam by reflection on the surface of the oxide layer. For the implementation of this method, it is necessary to know the size order of the thickness of the layer to be measured.

5.2. Epitaxy

The epitaxial growth is the formation method of semiconductor layers on semiconductor substrates of the same type (silicon-on-silicon, for example). As shown, the layer is called *epitaxial* if crystalline lattice of the grown layer continues the crystalline lattice of the substrate.

The technology used to build the first planar bipolar transistors presented a major drawback: the series collector resistance was high, due to the low concentration of impurities in the collector layer. The solution was provided by the development of epitaxial process, by which the collector layer can be produced with a suitable concentration and thickness, depending on requirements of the respectively integrated circuit. Basically, the whole structure of a bipolar transistor is produced based on a basic semiconductor substrate, having a high doping concentration, which determines a low series collector resistance (figure 5.8).

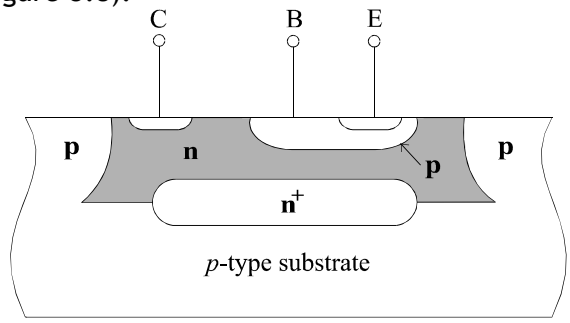

Figure 5.8. Section through the structure of a bipolar transistor made in an epitaxial layer (adapted after Gray and Meyer [15])

Epitaxy is used for: making *pn* junctions (between the substrate and the epitaxial layer), making *n* (epi) – n^+ (substrate) layers or *p* (epi) – p^+ (substrate) layers, making *abrupt profiles*, approximated by the approximation of the one-side abrupt junction (the real distribution of the electric charge is very close to this approximation). With the help of this technological process a strict control of the concentration of impurities in the epitaxial layer is achieved (independent of the concentration of impurities in the substrate), by controlling the concentration of impurities in the *process gas* (vapors from the process chamber). For carrying out the epitaxial growth process, two methods are used.

The first refers to *the reduction* from vapor phase of *the silicon tetrachloride* ($SiCl_4$):

$$SiCl_4(gaz)+2H_2(gaz)\leftrightarrow Si(solid)+4HCl(gaz), \quad (5.21)$$

which has industrial uses. The chemical reaction is reversible, so that, under certain conditions corrosion of the silicon may occur. The process takes place at relatively high temperatures: between 1150 °C and 1200 °C (generally higher than 1000 °C), for

[2] Optical path, (l): the product of the geometric path, l, traversed by a beam of light through a transparent medium and environmental refractive index, n.

which the deposition speed is high. The high temperature is necessary in order to obtain a layer having similar characteristics to the crystalline substrate. A drop in temperature has the effect of reducing the mobility of the deposited atoms, resulting in an increase of defects in the grown layer, the layer running the risk to lose the quality of *crystalline* layer.

The $SiCl_4$ concentration influences the deposition conditions. If the concentration is very high, the corrosion of the silicon substrate interferes, even if the concentration of hydrochloric acid (HCl) in the gas entering the reactor is low. Normally, the growth takes place at low concentrations, typically at a rate of 1 μm/min, the rate of increase approximately linearly dependent on *the* $SiCl_4$ *molar fraction* in the gas mixture [8].

The second method uses the chemical reaction of the decomposition of *silane* (SiH_4):

$$SiH_4(gaz) \rightarrow Si(solid) + 2H_2(gaz), \quad (5.22)$$

at temperatures between 1050 °C and 1150 °C.

The doping of epitaxial layers is made *in situ* by the introduction of hydrogenated components of the doping element in the process. For example, for doping with phosphorus, *the phosphine* (PH_3), is used, which is absorbed at the surface of the epitaxial layer, decomposes, the phosphorus being incorporated into the epitaxial layer during growth [16].

The reactors used in the epitaxial growth are of two types: vertical and horizontal. The vertical reactors have the advantage of uniform thickness of the layers, due to the rotation of *the susceptor*(the support on which silicon wafers are placed, figure 5.9).

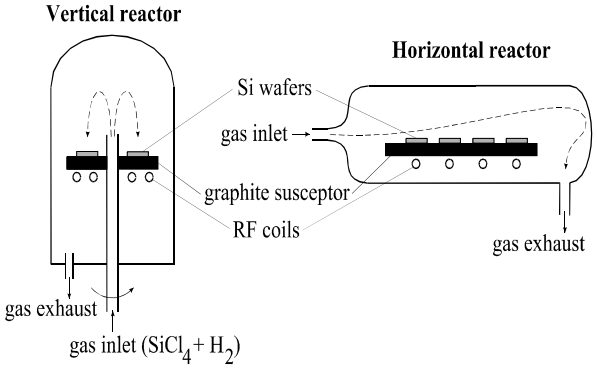

Figure 5.9. The scheme of the horizontal and vertical reactors for the epitaxial growth

Even without the susceptor's rotation, due to uniform distribution of gases flow, the thickness' uniformity of the deposited layer is better than the one in the horizontal reactors.

An important component of the epitaxial growth system is represented by *the gas distributor*, by means of which the control of the concentration of the reaction products is carried out, by controlling the gases flow (figure 5.10).

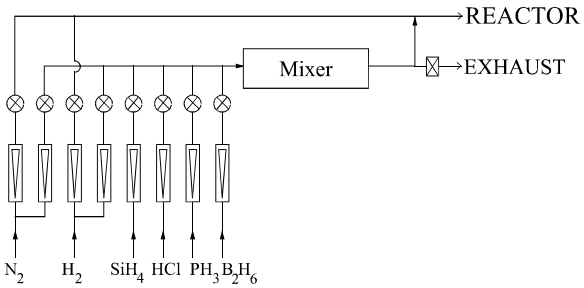

Figure 5.10. The scheme of the gas distributor

5.2.1. Epitaxial growth kinetics

Epitaxial growth kinetics is modeled very simply on the Grove model [8]. The epitaxial layer growth kinetics stages are as follows (figure 5.11):

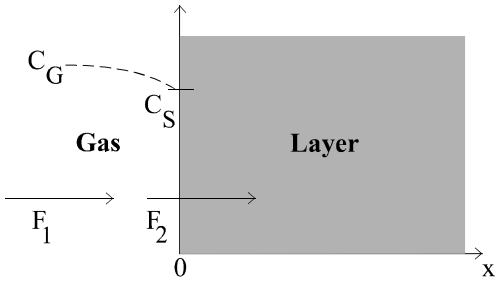

Figure 5.11. The stages of epitaxial layer growth kinetics (adapted after Grove [8])

- *the transport of the silicon tetrachloride* from the gas volume at the interface layer, with the flow F_1:

$$F_1 = h_G(C_G - C_S), \quad (5.23)$$

where: h_G represents the gas phase mass transfer coefficient, C_G is the concentration of silicon tetrachloride in the gas volume and C_S, the concentration of the silicon tetrachloride on the surface of the grown layer;

- *the growth chemical reaction of the epitaxial layer*, defined by the F_2 flow relationship:

$$F_2 = k_S C_S, \quad (5.24)$$

where k_S is *the chemical reaction constant rate at the surface*.

The steady state condition is expressed by the relation $F_1 = F_2$, the following expression are obtained:

$$C_S = \frac{C_G}{1 + \frac{k_S}{h_G}}. \qquad (5.25)$$

Two limit cases can be discerned:

- $k_S >> h_G$, when $C_S \to 0$, a condition called *control by mass transfer*;

- $k_S << h_G$, when $C_S \to C_G$, a condition called *control by reaction at the surface*.

The growth rate of the epitaxial layer (of silicon) can be expressed by the relationship:

$$V = \frac{F_2}{N_1} = \frac{k_S h_G}{k_S + h_G}\frac{C_G}{N_1} = \frac{k_S h_G}{k_S + h_G}\frac{C_T}{N_1}Y, \qquad (5.26)$$

where: N_1 defines the number of silicon atoms/unit volume of the increased layer, C_T is the total number of molecules in the gas, and Y, the molar fraction of the reactive species.

The growth rate, for a given Y, is determined by the lowest of the coefficients k_S and h_G. In the two extreme situations stated above, the growth rate is expressed:

- for kS >>hG:

$$V = h_G \frac{C_T}{N_1}Y, \qquad (5.27)$$

- for $k_S << h_G$:

$$V = k_S \frac{C_T}{N_1}Y, \qquad (5.28)$$

The dependence of the growth rate directly proportional to the molar fraction, corresponds to experimental observations, for small values of Y ($Y \leq 0{,}01$). In this context, the linear approximation of the reaction from the surface, defined by the flow F_2 is also valid.

The growth rate depends also on the temperature. Thus, at low temperatures, the dependence is given by the exponential of $V \sim \exp(-E_a/kT)$, *the activation energy* E_a having a value of 1,9 eV or, according to some authors, 1,6 eV. At high temperatures, the growth rate is saturated and becomes relatively independent of temperature.

The discussed model is a simple model, which explains the general features of the experimental results shown in [17]. According to these data, it appears that the best fit with the dependence relation calculated on the basis of (5.26) is obtained for values of h_G=5÷10 cm/s, and $k_S = 10^7 \cdot \exp(-1{,}9/kT)$ cm/s.

This simplified model does not consider the reaction product flow (HCl). Also, the model neglects the existing reactor temperature gradient, between the heated substrate and the reactor walls. In conclusion, we can say that the model discussed describes some general properties observed experimentally [8].

A good-quality epitaxial layer (monocrystalline) is obtained at a higher value growth temperature (from 1000 °C to 1200 °C) [16].

5.2.2. Mass transfer in the gas phase

Generally, the conditions of the industrial epitaxial growth are given by the approximation of *the control by mass transfer*. According to relation (5.23), the flow of gas volume to the solid surface (layer) is directly proportional with the coefficient h_G. To determine the factors on which this coefficient depends on, *the stationary layer model* is considered [8], which is the simplest model for the mass transfer. In this model, the gas phase is divided into two parts: a stationary layer, adjacent to the solid surface, of a δ thickness and a well-mixed gas volume which flows along the surface of the solid (figure 5.12).

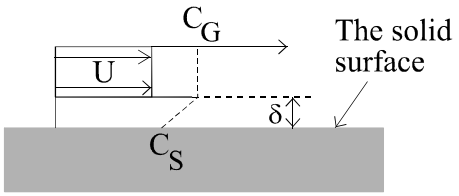

Figure 5.12. The stationary layer model (adapted after Grove [8])

Since the active species transport through the stationary layer is achieved by diffusion, the F_1 flow is expressed by the relationship:

$$F_1 = D_G \frac{C_G - C_S}{\delta} \equiv h_G (C_G - C_S), \qquad (5.29)$$

where D_G represents the diffusion coefficient of silicon tetrachloride, in the gas. For the h_G coefficient, the following relationship results:

$$h_G = \frac{D_G}{\delta}. \tag{5.30}$$

The thickness of the growth layer must be determined experimentally for each separate case.

5.3. Ion implantation

The ion implantation method consists in the acceleration of ions (from an ion source), to a velocity high enough to penetrate the semiconductor sample. Under conditions that can be controlled, the ions that penetrated in the volume of a semiconductor occupy positions in the lattice so that they can act as impurity atoms, donors or acceptors [12].

The wafers are placed in a vacuum chamber and the desired species impurity ions are sent to the wafer at high velocity (the depth of penetration of impurity ions depends on the velocity, in other words on the kinetic energy that ions have at the entrance into the volume of the semiconductor wafer). Finally, the wafers are heat treated at a moderate temperature (for example 10 minutes at 800 °C). In this way the implanted ions are rearranged in the crystal lattice and, also, the crystal lattice, which has suffered because of implantation, is restored [15].

The semiconductor doping by ion implantation isn't, practically, a thermal process (heat treatments subsequent to the implantation process occur at lower temperatures than those at which the doping by diffusion method is achieved), resulting in a significant technological advantage, namely the possibility of adjusting the value of the threshold voltage of the MOS transistors, to a value different from the designed one. This adjustment is made in the technological process, by implanting impurities at the surface where the induced channel is formed. Thus, for a MOS transistor of *n-channel enhancement mode*, the threshold voltage can be increased by implanting *p*-type impurities; on the contrary, by implanting of *n*-type impurities in the channel region, a MOS transistor structure is obtained with *n-channel depletion mode* (conducting channel at voltage V_{GS} = 0 [15]). The major advantage is the fact that this adjustment does not interfere with the junction profiles realized through the diffusion method.

One important and useful feature of ion implanted layers is that the maximum of the impurities concentration profile can be placed below the surface of the semiconductor wafer (as opposed to the diffused layer, to which this maximum is, generally, at the surface of the semiconductor substrate). This feature allows the fabrication of bipolar and junction field-effect transistor structures, with better features than those of diffused devices [15].

Other important features of this method are: the insertion of a certain impurities concentration does not depend on the properties of the substrate but on the external system (the nature and the energy of incident ions); the concentration of impurities introduced does not depend on their solubility into the semiconductor material to be doped.

The equipment for carrying out ion implantation consists of an ion accelerator, equipped with a mass analysis and dosimeter system (*ion implanter*). The main components of the ion implanter are: the high voltage generator, the ion sources, the ions accelerator (ions energy provided is from 30 keV to 200 keV), the mass separator (to select ions of a desired mass), collimating and ions beam deflection systems. The ion sources used are *fluorides* of the chemical elements used for the controlled impurification: BF_3, for the boron (B) doping, PF_5 for the phosphor (F) doping, and AsF_3 for arsenic (As) doping.

The implantation dose is expressed as a function of the window area which limits the ions beam (A), the time of the implantation process (t) and the ions beam current (I_F) [16]:

$$\Phi = \frac{1}{qA}\int_0^t I_F\,dt = \frac{I_F t}{qA}, \tag{5.31}$$

if current I_F is constant.

Each implanted ion suffer collisions with atoms and electrons of the crystalline target which gradually reduce its energy till a complete rest [18]. The implanted ions

creates punctual and linear defects or amorphous micro zones (figure 5.13).

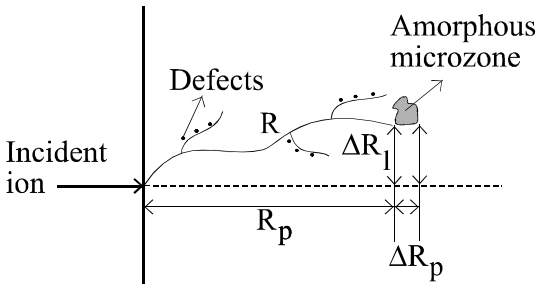

Figure 5.13. The trajectory of the implanted ions in a solid target (adapted after Barsan [16])

The length of the trajectory traveled by the ions into the solid target, R, depends on the braking properties of the solid. There are two mechanisms by which the implanted ions lose their initial energy. Thus, the ions collide with the target nuclei, causing the deflection of ions and the change in position of the nuclei with which they collided, resulting in a *nuclear braking force* $S_n = (\mathrm{d}E/\mathrm{d}x)_n$. Into the crystalline substrate of the target faults occur because of these collisions. Therefore, the ions collide with bound and free electrons, causing transient generation of electron-hole pairs, resulting in an *electronic braking force* $S_e = (\mathrm{d}E/\mathrm{d}x)_e$. In terms of distance, the defects generated by nuclear braking are higher at larger distances towards the surface of the target (at a distance approximately equal to R_p).

If N is the density of the respective target atoms, the total average loss of energy is [16]:

$$-\mathrm{d}E/\mathrm{d}x = N[S_n(E) + S_e(E)]. \qquad (5.32)$$

The expression of R follows:

$$R = \frac{1}{N}\int_0^E \frac{\mathrm{d}E}{S_n(E) + S_e(E)}. \qquad (5.33)$$

Due to the different number of ions collisions on the same distance in the target (lower or higher than the average number of collisions), for the probability of locating of the impurity atoms, a statistical distribution is obtained, the Gaussian distribution being the much simpler model. The distribution parameters are: the most probable depth R_p, corresponding to the position of the distribution's maximum (R_p - the projection on the incident direction of the length of the path traveled by the ions), the square deviation ΔR_p and the medium square lateral deviation ΔR_l (figure 5.13).

The concentration profile of the implanted atoms based on the depth of penetration, from the surface of the target, is expressed [16]:

$$C_i(x) = C_i(R_p)\exp\left[-\frac{(x-R_p)^2}{2(\Delta R_p)^2}\right]. \qquad (5.34)$$

and the implanted dose has the relationship:

$$\Phi = \int_0^\infty C_i(x)\mathrm{d}x = \sqrt{2\pi}C_i(R_p)\cdot\Delta R_p. \qquad (5.35)$$

Substituting (5.35) into (5.34) the distribution of the implanted impurities is obtained [16]:

$$C_i(x) = \frac{\Phi}{\sqrt{2\pi}\Delta R_p}\exp\left[-\frac{(x-R_p)^2}{2(\Delta R_p)^2}\right]. \qquad (5.36)$$

The relation (5.36) is consistent with the peak profile (for the distribution of implanted atoms in the silicon), the correlation being lower on the flanks of distribution, in particular for high-energy implantations, due to the effects of the asymmetric scattering of the implanted ions on the target atoms.

As in the case of diffusion, the theoretical condition of the junction formation is expressed by the relationship:

$$C_i(x_j) \equiv C_B = C_i(R_p)\exp\left[-\frac{(x_j-R_p)^2}{2(\Delta R_p)^2}\right], \qquad (5.37)$$

which can be solved according to x_j (figures 5.14, 5.15).

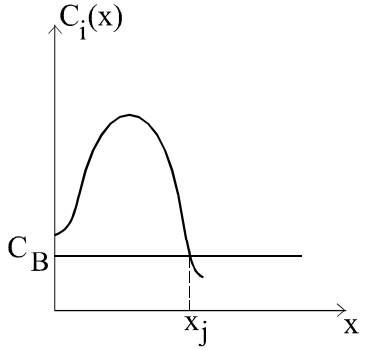

Figure 5.14. The junction formation in the case of shallow ion implantation

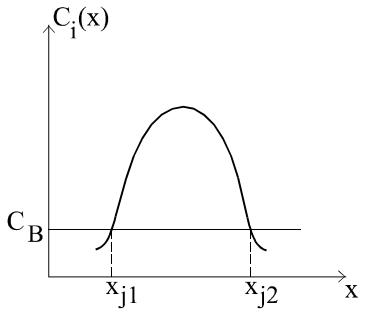

Figure 5.15. The junction formation in the case of deep ion implantation

The lateral dispersion of the ion implantation has a practical importance because based on it, *the effective length* of the MOS channel transistors can be calculated (with source and drain junctions made by ion implantation), this one being inferior to the lateral diffusion.

The bi-dimensional profile of the impurities is expressed by the relationship

$$C_i(x,y)=\frac{C_i(x)}{2}\left[\operatorname{erfc}\left(\frac{y-a}{\sqrt{2}\Delta R_l}\right)-\operatorname{erfc}\left(\frac{y+a}{\sqrt{2}\Delta R_l}\right)\right], \qquad (5.38)$$

where $C_i(x)$ is one-dimensional concentration away from the edges of the window.

The lateral distribution of impurities (in the *y* direction) after the implantation through a 2*a* window width (infinite length) is illustrated in figure 5.16 ($a >> \Delta R_l$).

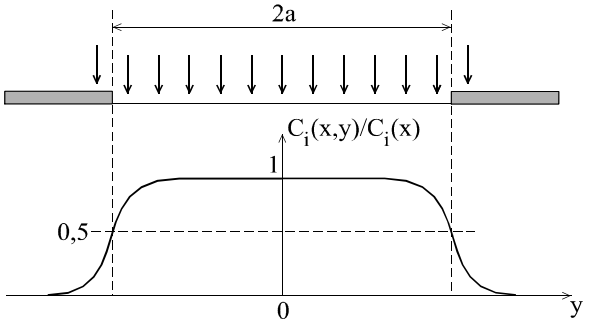

Figure 5.16. Side profile of impurities implanted through a 2a window width and infinite length (adapted after Barsan [16])

Any monocrystal (silicon) acts as an amorphous target if the ions beam is misaligned compared to the crystallographic directions, with 6 ÷ 7°. Otherwise, the implanted profile is strongly modified due to the effect of sewage (ions entering the empty space between the rows of atoms - channels) [16]. The result if the guidance of the ions through the center of the channels due to action of the atomic forces, the ions covering long distances.

The minimum angle of tilt of the target to beam, in order to minimize channeling, is called critical angle and depends on: the type of implanted ions, their energy, the monocrystalline orientation. The misalignment angle is an important parameter because a small misalignment causes a deeper penetration of the implanted ions.

The ion implantation to adjust the threshold voltage of MOS transistors offers the possibility of initial choice of substrate doping without taking into account the desired threshold voltage (only the optimum performance of the structure - current-voltage characteristic - is taken into consideration, the threshold voltage value remaining to be adjusted later). The calculation of the threshold voltage variation, depending on the implanted dose, is (figures 5.17. [19]):

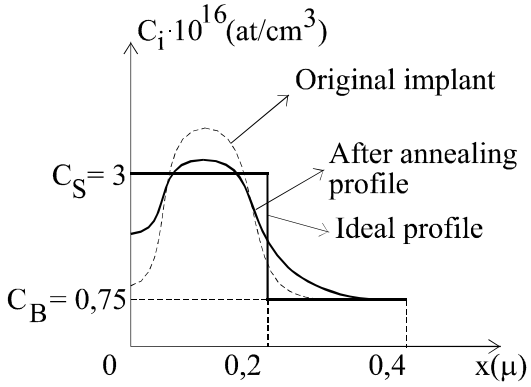

Figure 5.17. The profile of the implanted impurities below the oxide layer of the gate. The ideal profile is used to estimate the threshold voltage variation obtained by ion implantation (adapted after Wolf [19])

$$\Delta V_T=\frac{qC_i}{C_{ox}}. \qquad (5.39)$$

In the approximation of the relation (5.39), the implantation depth is ignored, approximation accepted in the design practice.

6. References (for Part 4)

[8] Grove, A.S., *Fizica şi tehnologia dispozitivelor semiconductoare* [Physics and technology of semiconductor devices], Editura Ştiinţifică şi Enciclopedică, Bucureşti 1977.

[10] Sze, S.M., *Physics of Semiconductor Devices*, Publishing John Wiley & Sons, Inc. 1981.

[12] Dolocan, V., *Fizica dispozitivelor cu corp solid* [Physics of solid state devices], Editura Academiei Române, Bucureşti 1978.

[15] Gray, P.R., Meyer, R. G., *Circuite integrate analogice. Analiză și proiectare* [Analysis and design of analog integrated circuits], Editura Tehnică, București 1997.
[16] Bârsan, R. M., *Fizica și tehnologia circuitelor MOS integrate pe scară mare* [Physics and technology of MOS large-scale integrated circuits], Editura Academiei Române, București 1989.
[17] Shepherd, W.H., *J. Electrochem. Soc.* **112**, p.988, 1965.
[18] Lindhardt, J., Sharff, M., Schiott, H., *Mat.-Fys. Med. Dan.* Vidensk. Selsk. 33, p.1, 1963.
[19] Wolf, S., *Silicon Processing for the VLSI Era*, vol. 2 - *Process Integration*, Lattice Press 1990.

7. Biography

Eugen Șt. LAKATOS was born in Arad in 1954, August, 31.
He graduated the University of Bucharest, Faculty of Physics in 1979.
He is Doctor in Physics in 1996. He is professor at the "Valahia" University of Targoviște.
His research interests concern: modelling and simulation of microelectronic process and active semiconductor devices, modelling and SPICE simulation of active microelectronic semiconductor devices and analog integrated circuits.