Chapter 24
Thermal Oxidation—Layer Deposition

24.1 Introduction

High-quality oxide is essential in silicon technology. The most important applications of the oxide are the passivation of the wafer's surface, the isolation between metallizations, the formation of masks for, e.g., diffusion or implantation of dopants, the isolation between devices, and the formation of the gate insulator in MOS devices.

While the oxides used for passivation or isolation between metallization are typically obtained by chemical vapor deposition (Sect. 24.5), the oxide suitable for the other applications listed above is obtained by thermal oxidation. In fact, the extraordinary evolution of the VLSI technology in the last decades is due to a large extent to the excellent electrical properties of the thermally-grown layers of silicon dioxide and to the reliability and control of the growth process.

In crystalline silicon dioxide (quartz), one silicon atom forms chemical bonds with four oxygen atoms, creating a tetrahedral structure. In turn, one oxygen atom forms chemical bonds with two silicon atoms. The tetrahedra are thus connected to form a structure with a stoichiometric ratio $1 : 2$ and density $\rho = 2.65$ g cm^{-3} (Fig. 24.1). In thermally-grown SiO_2 not all tetrahedra are connected, because at the silicon–oxide interface chemical bonds must be created with the pre-existing silicon crystal, whose interatomic distance is different from that of SiO_2. As a consequence, the oxide has a shorter-range order giving rise to a more open (amorphous) structure with density $\rho = 2.20$ g cm^{-3}. Because of this the diffusion of contaminants, in most cases Na and H_2O, is easier than in crystalline silicon dioxide.

Also, the need of adapting to the silicon crystal produces a mechanical stress in the oxide layer closer to the silicon surface, which in turn influences the concentration of substrate defects and the value of some electrical properties in MOS devices (typically, the threshold voltage, Sect. 22.6.2). The properties of the mechanically-stressed layer are influenced by the process temperature T. At relatively low process temperatures, $T < 950\,^\circ$C, the stressed layer is thinner and the mechanical stress in it is stronger; when $T > 950^\circ$C, the stress distributes over a thicker layer and becomes locally weaker.

The growth of a thermal oxide's layer is obtained by inducing a chemical reaction between the silicon atoms belonging to the wafer and an oxidant species that is

M. Rudan, *Physics of Semiconductor Devices*,
DOI 10.1007/978-1-4939-1151-6_24

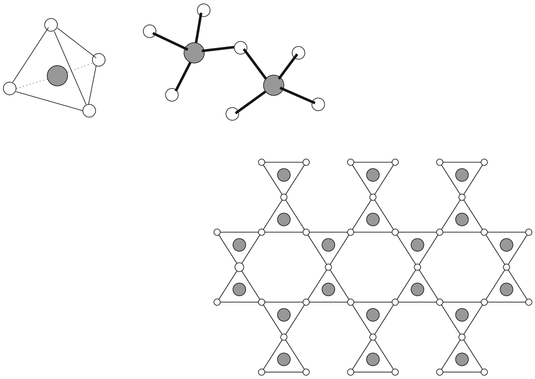


Fig. 24.1 Structure of quartz. Silicon atoms are represented in *gray*, oxygen atoms in *white*. Within the tetrahedron, the distance between two oxygen atoms is about 0.227 nm, that between the silicon atom and an oxygen atom is about 0.160 nm. The schematic representation in two dimensions is shown in the *lower-right part* of the figure

brought into contact with it. As mentioned above, another technique for obtaining an oxide layer is deposition. The latter process has actually a broader scope, in fact it is used for depositing several types of conducting or insulating materials that are necessary in the fabrication of the integrated circuits. Deposition differs from the thermal growth because the chemical reaction may be absent or, if present, it does not involve the species that are in the solid phase. One special type of deposition is epitaxy, that is used to grow a crystalline layer over another crystalline layer.

The chapter illustrates the oxidation of silicon, starting from the description of the chemical reactions involved in it, and deriving the relation between the thickness of the oxide layer and that of the silicon layer consumed in its growth. The kinetics of the oxide growth is analyzed, the *linear–parabolic model* is worked out, and its features are commented. Then, a brief description of the deposition processes is given, followed by the description of the chemical reaction involved in the epitaxial process and by the analysis of the epitaxial kinetics.

In the last part of the chapter a number of complementary issues are discussed. Specific data about the parameters governing the thermal oxidation, deposition, and epitaxial processes in semiconductors are in [46, Chap. 2], [104, Chap. 9], [105, Chap. 3], [71, Chap. 3]. Many carefully-drawn illustrations are found in [75, Sect. 1.2].

24.2 Silicon Oxidation

Silicon exposed to air at room temperature oxidizes spontaneously and forms a shallow layer of SiO_2 of about 1 nm called *native oxide*. As soon as the native oxide is formed, the oxygen molecules of the air can not reach the silicon surface any more and the chemical reaction dies out. In fact, oxidation of silicon is caused by the inward motion of the oxidant. To activate the reaction and grow a layer of the desired thickness it is necessary to place the wafers at atmospheric pressure in a furnace (Fig. 24.2) kept at a temperature in the range $800°\mathrm{C} \le T \le 1200\,°\mathrm{C}$. This

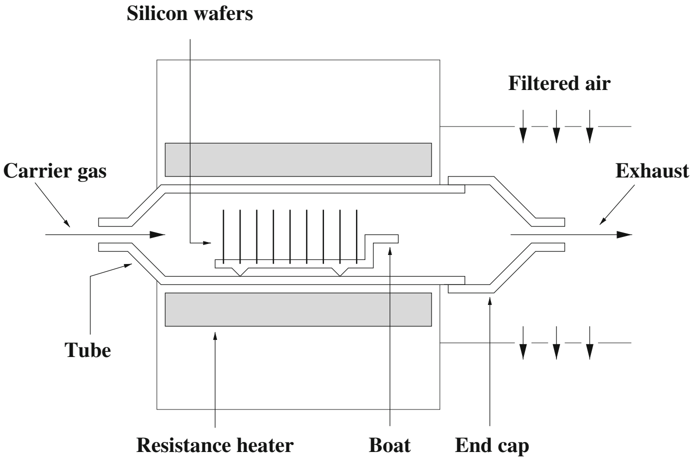


Fig. 24.2 Furnace for silicon oxidation. The intake of the carrier gas (O_2 or H_2O) is on the *left end* of the furnace, the exhaust on the *right end*. The tube, end cap, and boat, are made of fused quartz to avoid contamination

increases the diffusion coefficient of the oxidant. The latter penetrates the already-formed oxide layer and reaches the silicon surface, where new SiO_2 molecules are formed. The furnace is made of a quartz or polycrystalline-silicon tube heated by a resistance or by induction through a radiofrequency coil. To grow the oxide layer in a reproducible way it is necessary to control the temperature inside the furnace within $\pm 1\,^\circ$C. The oxidant is introduced from one end of the furnace after being mixed with a carrier gas (typically, N_2 or Ar).

The chemical reactions involved in the growth of thermal oxide are different depending on the type of oxidant. The latter is either molecular oxygen (O_2) or steam (H_2O). The corresponding thermal growth is called, respectively, *dry oxidation* or *wet (steam) oxidation*. The reactions read

$$\mathrm{Si} + \mathrm{O}_2 \rightleftharpoons \mathrm{SiO}_2, \qquad \mathrm{Si} + 2\mathrm{H}_2\mathrm{O} \rightleftharpoons \mathrm{SiO}_2 + 2\mathrm{H}_2, \tag{24.1}$$

where the hydrogen molecules produced by the second reaction are eliminated by the carrier gas. The formation of SiO_2 molecules is accompanied by a change in volume. In fact, each newly-formed SiO_2 molecule uses up a silicon atom initially belonging to the silicon crystal. On the other hand, the concentration of the silicon atoms in a silicon crystal is about $N_1 = 5.0 \times 10^{22}$ cm^{-3}, while that of the silicon atoms in a thermally-grown SiO_2 layer is about $N_2 = 2.2 \times 10^{22}$ cm^{-3}. Thus the ratio between the volume V of SiO_2 and that of the silicon consumed for its formation is

$$\frac{V(\mathrm{SiO}_2)}{V(\mathrm{Si})} = \frac{N_1}{N_2} \simeq 2.28. \tag{24.2}$$

As the oxide layer is free to expand only in the direction normal to the wafer, (24.2) is actually the ratio between the thickness s of the newly-formed SiO_2 layer

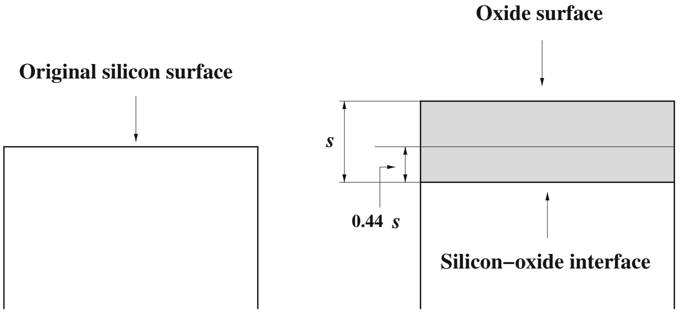


Fig. 24.3 The *left part* of the figure shows the position of the original silicon surface (prior to oxidation). The *right part* shows the position of the oxide's surface and of the silicon–oxide interface after an oxide layer of thickness s has been grown

and that of the silicon layer consumed in the process. It follows

$$\frac{s(\mathrm{Si})}{s(\mathrm{SiO_2})} = \frac{V(\mathrm{Si})/A}{V(\mathrm{SiO_2})/A} = \frac{N_2}{N_1} \simeq 0.44, \tag{24.3}$$

with A the area of the oxidized region. In other terms, when an oxide layer of thickness s is grown, the silicon–oxide interface shifts by $0.44\,s$ with respect to the original position (Fig. 24.3). If the oxidation takes place uniformly over the whole area of the wafer, the shift of the interface is uniform as well. However, in many cases the oxidation involves only portions of the wafer's area. This makes the silicon–oxide interface non planar, because the shift is different from one portion to another.

24.3 Oxide-Growth Kinetics

Growth kinetics is modeled after Deal and Grove [28]. The model describes the succession of steps by which the oxidant, initially in the gaseous phase, comes into contact with silicon and reacts with it. The steps are: the oxidant (i) diffuses from the source region into the already-formed oxide, (ii) crosses the oxide still by diffusion and reaches the silicon–oxide interface, (iii) produces a chemical reaction that forms a new SiO_2 molecule.

The motion of the gas parallel to the wafer surface is not considered. As a consequence, the only non-vanishing component of the oxidant average velocity has the direction x normal to the wafer surface. The corresponding flux density is $F = \mathbf{F} \cdot \mathbf{i}$, with $\mathbf{i}$ the unit vector parallel to x. The oxidant concentration N is assumed uniform over the wafer's surface, this making N and F to depend on x and t only.

The concentration of the oxidant in the bulk of the gaseous phase, N_G, is a known boundary condition because it is regulated by the microprocessors controlling the furnace. In the gaseous region, at the gas–oxide interface, and in the oxide region, no generation or destruction of oxidant molecules occurs. The flux density is given by

$$F = -D_S \frac{\partial N}{\partial x}, \qquad F = -D_O \frac{\partial N}{\partial x}, \tag{24.4}$$

respectively in the source and oxide region. In (24.4), the symbol D_S (D_O) indicates the diffusion coefficient of the oxidant in the source (oxide) region. Each diffusion coefficient is taken independent of time and spatially constant in its own region. The matching conditions at the source–oxide interface are the same as in (23.20), namely,

$$N_O = k N_S, \qquad D_O \left(\frac{\partial N}{\partial x}\right)_O = D_S \left(\frac{\partial N}{\partial x}\right)_S, \tag{24.5}$$

where k is the gas–oxide segregation coefficient, while the index S (O) attached to the concentration or its derivative indicates that the function is calculated at the source–oxide interface on the side of the source (oxide). As one of the two phases is gaseous, it is $D_S \gg D_O$ whence $|(\partial N/\partial x)_S| \ll |(\partial N/\partial x)_O|$. The situation here is similar to that illustrated in Fig. 23.3. It follows that the interface concentration of the source region, N_S, is practically equal to the boundary condition N_G. The first of (24.5) then yields $N_O = k N_G$.

To proceed one observes that, due to the thinness of the oxide layer, the oxidant concentration in it can be described by a linear approximation. Due to this, the flux density in the oxide layer (the second equation in (24.4)) becomes

$$F = -D_O \frac{N_I - N_O}{s} = D_O \frac{k N_G - N_I}{s}, \tag{24.6}$$

where N_I is the oxidant concentration on the oxide side of the silicon–oxide interface, and s the oxide thickness. Note that the flux density in (24.6) is constant with respect to x, whereas it is time dependent because s increases with time.

When the oxidant reaches the silicon–oxide interface it reacts with silicon, so that there is no flux density of the oxidant on the semiconductor's side of this interface. In fact, the oxidant's molecules are destroyed at the interface to form molecules of SiO_2. The flux density F_I entering the silicon–oxide interface gives the number of oxidant molecules destroyed per unit area and time which, to a first approximation, is taken proportional to the concentration N_I. It follows

$$F_I = v_r \, N_I, \tag{24.7}$$

where the constant v_r (m s^{-1}) is called *reaction velocity*. As F_I is just another symbol to denote the spatially-constant flux density within the oxide, one combines (24.7) with (24.6) to obtain

$$N_I = \frac{D_O \, k N_G}{v_r s + D_O}. \tag{24.8}$$

At a given instant Eq. (24.8) expresses N_I in terms of the boundary condition N_G, the process parameters k, D_O, v_r, and the oxide thickness s.

24.4 Linear–Parabolic Model of the Oxide Growth

The relation between the oxidant's flux density F and the growth velocity $\mathrm{d}s/\mathrm{d}t$ of the oxide layer is found as follows. Letting A be the area of the oxidized region, the product $A\,F$ provides the number of oxidant molecules reaching the silicon–oxide interface per unit time. Each molecule, in turn, makes the volume V of the oxide layer to increase by a finite amount w. As a consequence, the volume increase per unit time of the oxide layer is $\mathrm{d}V/\mathrm{d}t = wA\,F$. As shown in Sect. 24.2, the oxide layer is free to expand only in the direction normal to the wafer, so that $\mathrm{d}V/\mathrm{d}t = A\,\mathrm{d}s/\mathrm{d}t$. Combining the above relations and using (24.7, 24.8) yields a differential equation in the unknown s:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = wF_I = wv_r N_I = v_r D_O\,\frac{w\,kN_G}{v_r\,s + D_O}. \tag{24.9}$$

The above is readily separated as $(s/D_O + 1/v_r)\,\mathrm{d}s = wkN_G\,\mathrm{d}t$ and integrated from $t = 0$, to yield

$$\frac{1}{c_p}\,\left(s^2 - s_i^2\right) + \frac{1}{c_l}\,(s - s_i) = t, \qquad \begin{cases} c_p = 2w\,kN_G D_O \\ c_l = w\,kN_G\,v_r \end{cases}, \tag{24.10}$$

with $s_i = s(t = 0)$. The relation between s and t given by (24.10) is called *linear–parabolic model* of the oxide growth. The quantities c_p ($\mathrm{m}^2\ \mathrm{s}^{-1}$) and c_l ($\mathrm{m\ s}^{-1}$) are the *parabolic coefficient* and *linear coefficient*, respectively. The model is recast as

$$\frac{1}{c_p}\,s^2 + \frac{1}{c_l}\,s = t + \tau, \qquad \tau = \frac{1}{c_p}\,s_i^2 + \frac{1}{c_l}\,s_i. \tag{24.11}$$

Using (24.11) and the definitions (24.10) of c_p and c_l one finds two limiting cases of the $s(t)$ relation. Specifically, it is $s \simeq c_l\,\,(t + \tau)$ when the oxide thickness is such that $v_r s \ll 2D$, while it is $s \simeq [c_p\,\,(t + \tau)]^{1/2}$ when the oxide thickness is such that $v_r s \gg 2D$. Due to the form of c_p and c_l, in the first limiting case the oxide growth does not depend on the diffusion coefficient D_O, whereas in the second limiting case it does not depend on the reaction velocity v_r. This is easily understood if one considers that the concentration $N_O = kN_G$ is prescribed. As a consequence, as long as the oxide thickness is small the derivative of the concentration (hence the flux density) is limited essentially by the flux density entering the silicon–oxide interface, (24.7); on the contrary, when the oxide thickness becomes large the flux density is limited essentially by the diffusion across the oxide because the value of the concentration N_I at the silicon–oxide interface becomes less important.

The differential form of (24.11),

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{2}{c_p}\,s + \frac{1}{c_l}, \tag{24.12}$$

is a linear relation between $\mathrm{d}t/\mathrm{d}s$ and s. Such quantities can be measured independently from each other, this providing a method for measuring c_p and c_l. Repeating the measurement at different temperatures shows that the temperature dependence of the parabolic and linear coefficients is given by

$$c_p = c_{p0} \exp\left[-E_{ap}/(k_B T)\right], \qquad c_l = c_{l0} \exp\left[-E_{al}/(k_B T)\right]. \tag{24.13}$$

The form of (24.13) is due to the temperature dependence of $D \propto \exp[-E_{ap}/(k_B T)]$ and, respectively, $v_r \propto \exp[-E_{al}/(k_B T)]$. In fact, the parameters w, N_G that appear in the definitions (24.10) are independent of temperature, whereas the temperature dependence of the segregation coefficient k, that can be measured independently, is shown to be relatively weak.

The measurement of c_p and c_l allows one to determine also other properties of the oxidation process; for instance, the effect of carrying out a steam or dry oxidation, and the influence of the substrate orientation. As for the first issue one finds

$$k(\text{Steam}) > k(\text{Dry}), \qquad D_O(\text{Steam}) > D_O(\text{Dry}). \tag{24.14}$$

The Arrhenius plots of c_p and c_l are shown in Figs. 24.4 and 24.5, respectively. In each plot the upper (lower) continuous curve refers to the steam (dry) oxidation. As for the effect of the crystal orientation of the silicon wafer, one observes that the number of chemical reactions per unit time involved in the formation of SiO_2 molecules must depend on the surface density of silicon atoms at the silicon–oxide interface. Due to this, the reaction velocity is expected to depend on the orientation of the interface. The crystal planes that are typically used in the silicon technology are the (111) one, whose surface density is 11.8×10^{14} cm^{-2}, and those equivalent to the (100) one,[1] whose surface density is 6.8×10^{14} cm^{-2}. In fact the experiments show that

$$\frac{v_r[(111)]}{v_r[(100)]} = 1.68 \simeq \frac{11.8 \times 10^{14}\ \text{cm}^{-2}}{6.8 \times 10^{14}\ \text{cm}^{-2}}. \tag{24.15}$$

The effect on c_l of the crystal orientation is shown by the dotted curves in the Fig. 24.5.

24.5 Layer Deposition and Selective Oxide Growth

The deposition of films of different materials is necessary at several steps of the integrated-circuit fabrication. Conducting materials provide the electrical connections among the individual devices of the integrated circuit, while insulating materials provide the electrical insulation between the metal layers, and the protection from the environment. The majority of the deposition processes take place in the vapor phase under reduced-pressure or vacuum conditions. One exception is the deposition of resist, which is carried out in the liquid phase.

---

[1] The definitions of the crystal planes are given in Sect. 17.8.1.

Fig. 24.4 Parabolic coefficient c_p as a function of $1{,}000/T$. The units are $\mu m^2\ h^{-1}$. The activation energy of the steam case is 0.71 eV, that of the dry case is 1.24 eV

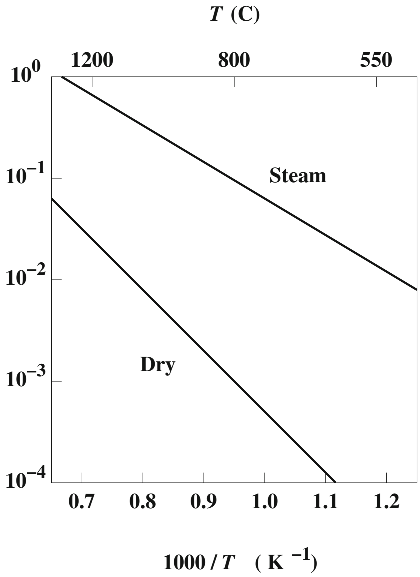
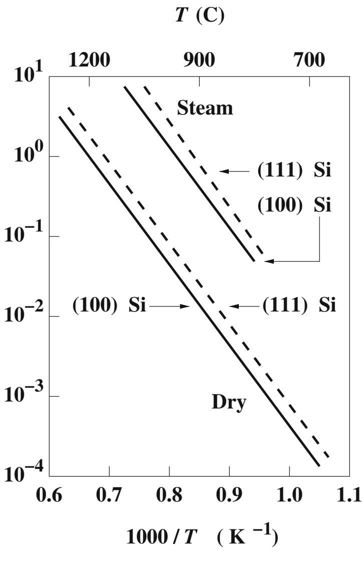


Fig. 24.5 Linear coefficient c_l as a function of $1{,}000/T$. The units are $\mu m\ h^{-1}$. The activation energy of the steam case is 2.05 eV, that of the dry case is 2.0 eV



When the material to be deposited does not react chemically with other substances, the process is called *physical vapor deposition* (PVD). An example of PVD is the deposition of a metal by evaporation *in vacuo*. When the material to be deposited is the product of a chemical reaction that takes place over the wafer surface or in its vicinity, the process is called *chemical vapor deposition* (CVD).

Table 24.1 Examples of CVD reactions

Product	Reaction[a]		Deposition temperature (°C)
Polysilicon	SiH_4	$\rightarrow Si + 2H_2$	575–650
Silicon dioxide	$SiH_4 + O_2$	$\rightarrow SiO_2 + 2H_2$	400–450
Silicon nitride	$3SiH_4 + 4NH_3$	$\rightarrow Si_3N_4 + 12H_2$	700–900

[a] SiH_4 and NH_3 are called *silane* and *ammonia*, respectively

The materials that are most widely used in CVD processes are polycrystalline silicon (also termed *polysilicon*), silicon dioxide (SiO_2), and silicon nitride (Si_3N_4). Examples of CVD reactions are given in Table 24.1. More examples are found in [105, Sect. 6.2].

The structure of the deposited layer depends on the substrate's properties and deposition conditions. In the manufacturing of integrated circuits the substrate is crystalline, that is, it has long-range order extending throughout the entire volume (Chap. 17). If the material to be deposited on a crystalline substrate is the same as that of the substrate, by means of a carefully-controlled process it is possible to obtain a deposited layer that replicates the substrate's structure. Such a process is called *epitaxy* and, with reference to silicon, is described in Sect. 24.6.

The structure of silicon deposited on a different material is polycrystalline, that is, it has a long-range order only within small volumes. Such volumes, called *grains*, have an average diameter of about 1 μm and are oriented randomly with respect to each other. Polycrystalline silicon is used for fabricating the gate electrodes in MOS devices, for obtaining ohmic contacts to shallow junctions, and for producing resistors. To increase the gate's conductivity, a layer of metal or metal silicide (like tungsten or tantalum silicide) may be deposited over the polycrystalline silicon.

The structure of deposited SiO_2 or Si_3N_4 is amorphous, that is, it has a short-range order only. The applications of SiO_2 have been illustrated in Sect. 24.1. Silicon nitride Si_3N_4 provides a strong barrier to the diffusion of water, that corrodes the metallizations, and of other contaminants, like sodium, that make the devices unstable by changing their threshold voltage. In addition, Si_3N_4 is resistant to high temperatures and oxidizes slowly. For these reasons it is used for passivating the wafer and for producing the masks that are necessary for the selective oxidation of silicon. The latter process, also called *local oxidation* (LOCOS), consists in depositing and patterning a Si_3N_4 layer over the areas where the substrate's oxidation must be prevented. As oxidation is isotropic, a lateral penetration of the oxidized region occurs under the edge of Si_3N_4. This produces a characteristic profile of the oxide layer called *bird's beak*. To compensate for the effect of the lateral penetration, the Si_3N_4 mask must be larger than the area whose oxidation is to be prevented.

A layer replicates the topography of the surface onto which it is deposited. For this reason it is important to avoid, or reduce, the formation of steps on the substrate. In fact, over a step the layer's thickness is smaller than on a flat surface which, in turn, may cause reliability problems in the final circuit. For instance, the non-uniform thickness of a metal line causes a non-uniform distribution of the current density. This may induce metal migration and the eventual breakdown of the metal connection.

24.6 Epitaxy

Epitaxy (from the Greek verb *epitàsso*, "to deploy side by side") is used to grow a monocrystalline layer over another monocrystalline layer. Most epitaxial processes use the CVD method. When the epitaxial layer is made of the same material as the substrate, e.g., silicon over silicon, the term *homoepitaxy* is also used, while the term *heteroepitaxy* is reserved to the case where the materials are different. Heteroepitaxy is possible when the difference between the lattice constants[2] of the two materials is small. An example of heteroepitaxy is the silicon-on-sapphire (SOS) process, that belongs to the silicon-on-insulator (SOI) technological family and consists in growing a thin layer of silicon (about 0.5 μm) on a wafer made of a sapphire crystal (Al_2O_3). Another application of heteroepitaxy is the fabrication of the heterojunctions that are necessary in optoelectronic devices.

In the silicon technology, epitaxy originated from the need of producing high-resistance layers in bipolar technology. This type of layers is necessary, e.g., for realizing the collector region of the bipolar junction transistor, whose dopant concentration must be substantially lower than that of the base region. Due to the high temperature of the CVD process (about 1200 °C), during an epitaxy a diffusion occurs of the substrate dopant into the epitaxial layer and of the epitaxial-layer's dopant into the substrate. This effect must be accounted for, and compensated, at the design stage of the process.

The fundamental reaction of epitaxy combines silicon tetrachloride $SiCl_4$ with molecular hydrogen in the vapor phase to obtain silicon in solid phase, while the hydrochloric acid HCl remains in the vapor phase and is eliminated:

$$SiCl_4 + 2H_2 \rightleftharpoons Si + 4HCl. \tag{24.16}$$

Reaction (24.16) is reversible: an excess of HCl removes silicon atoms from the wafer's surface and releases $SiCl_4$ and $2H_2$ in the vapor phases. This reaction is used in the first stages of the process to the purpose of cleaning the wafer's surface. Besides (24.16), a secondary reaction takes place as well, namely,

$$SiCl_4 + Si \rightarrow 2SiCl_2. \tag{24.17}$$

Reaction (24.17) removes silicon from the wafer's surface and releases silicon dichloride $SiCl_2$ in the vapor phase. For this reason, reactions (24.16) and (24.17) compete with each other. When the vapor concentration of $SiCl_4$ is sufficiently low the first reaction prevails and the thickness of the epitaxial layer increases with time. In contrast, at higher $SiCl_4$ concentrations the second reaction prevails and silicon is etched.

The epitaxial layer is doped by introducing hydrides of the dopants into the vapor phase. The hydride, e.g., *arsine* (AsH_3), *phosphine* (PH_3), or *diborane* (B_2H_6), is

---

[2] The definition of lattice constant is given in Sect. 17.6.4.

absorbed on the surface, decomposes, and is incorporated in the growing layer, e.g.,

$$2\mathrm{AsH_3} \rightarrow 2\mathrm{As} + 3\mathrm{H_2}. \tag{24.18}$$

24.7 Kinetics of Epitaxy

As in the case of the oxide-growth kinetics (Sect. 24.3), the motion of the vapor parallel to the wafer surface is not considered. As a consequence, the only non-vanishing component of the average velocity of the $SiCl_4$ molecules has the direction x normal to the wafer surface. The corresponding flux density is $F = \mathbf{F} \cdot \mathbf{i}$, with $\mathbf{i}$ the unit vector parallel to x. The $SiCl_4$ concentration N is assumed uniform over the wafer's surface, this making N and F to depend on x and t only.

The $SiCl_4$ concentration in the bulk of the vapor phase, N_G, is a known boundary condition because it is regulated by the microprocessors controlling the furnace. The flux density is given by

$$F = -D\,\frac{\partial N}{\partial x} \simeq v_G\,(N_G - N_I), \tag{24.19}$$

where D is the diffusion coefficient of $SiCl_4$ in the vapor phase and N_I the $SiCl_4$ concentration at the wafer's surface. The diffusion coefficient is taken independent of time and spatially constant. The form of the right hand side of (24.4), where the parameter v_G (m s^{-1}) is called *gas-phase mass-transfer coefficient*, is due to the observation that D is very large because the diffusion takes place in the vapor phase. As a consequence, the derivative $\partial N/\partial x$ is so small that a linear approximation for N is acceptable (the situation here is similar to that illustrated for the region on the left of the origin in Fig. 23.3). Note that the flux density in the vapor phase (24.19) is constant with respect to x. In principle it depends on time because the extension of the vapor phase decreases due to the growth of the epitaxial layer. However, this time dependence can be disregarded because the relative variation in the vapor-phase extension is negligible.

The flux density F_I entering the silicon surface gives the number of $SiCl_4$ molecules destroyed per unit area and time which, to a first approximation, is taken proportional to the concentration N_I. It follows

$$F_I = v_r\,N_I, \tag{24.20}$$

where the constant v_r (m s^{-1}), as in the case of the oxide-growth kinetics, is called *reaction velocity*. As F_I is just another symbol to denote the spatially-constant flux density, one combines (24.20) with (24.19) to obtain

$$N_I = \frac{v_G}{v_r + v_G}\,N_G. \tag{24.21}$$

At a given instant Eq. (24.21) expresses N_I in terms of the boundary condition N_G and process parameters v_G, v_r.

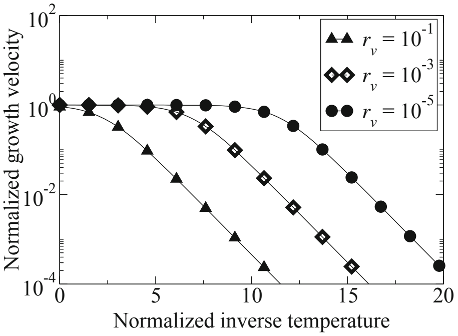


Fig. 24.6 Normalized growth velocity as a function of the normalized inverse temperature, as given by (24.23) and (24.24), at different values of the $r_v = v_G/v_{r0}$ ratio

The relation between the flux density F of $SiCl_4$ and the growth velocity $\mathrm{d}s/\mathrm{d}t$ of the epitaxial layer is found by the same reasoning as that used in Sect. 24.4 for the growth velocity of SiO_2. From (24.20, 24.21) it follows

$$\frac{\mathrm{d}s}{\mathrm{d}t} = w F_I = w v_r N_I = w \,\frac{v_r v_G}{v_r + v_G}\, N_G \tag{24.22}$$

whence, observing that $s(t = 0) = 0$,

$$s = c_l\, t, \qquad c_l = w\, \frac{v_r v_G}{v_r + v_G}\, N_G. \tag{24.23}$$

The $s(t)$ relation (24.23) is linear with respect to time. The growth velocity c_l of the epitaxial layer depends on the concentration N_G of $SiCl_4$ at the boundary and on the process parameters w, v_r, and v_G. The temperature dependence of the gas-phase mass-transfer coefficient v_G is weak. As w and N_G are independent of temperature, the temperature dependence of c_l is to be ascribed to v_r. It is found

$$v_r = v_{r0}\, \exp\left[-\, E_{al}/(k_B T)\right]. \tag{24.24}$$

When the temperature is such that $v_r \ll v_G$, which typically happens for $T <$ 1,150 °C, the second of (24.23) yields the limiting case $c_l \simeq w\, N_G\, v_r$, whence $c_l \propto \exp[-\, E_{al}/(k_B T)]$; when, instead, it is $v_r \gg v_G$, which typically happens for $T >$ 1,200°C, the limiting case is $c_l \simeq w\, N_G\, v_G = \mathrm{const}$. An Arrhenius plot of the normalized growth velocity $c_l/(w\, N_G\, v_G)$ as a function of the normalized inverse temperature $E_{al}/(k_B T)$ is shown in Fig. 24.6 for different values of the $r_v = v_G/v_{r0}$ ratio.

24.8 Complements

24.8.1 An Apparent Contradiction

In commenting (24.6) it was noted that the flux density F in the oxidation process is constant with respect to x, whereas it depends on time due to the time dependence of

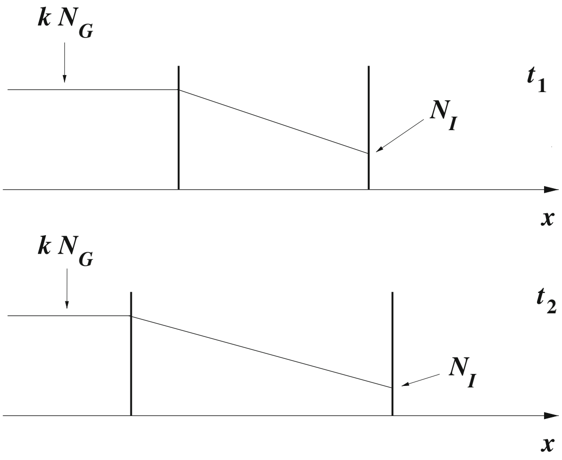


Fig. 24.7 Oxidant concentration within the oxide at two different instants, t_1 and $t_2 > t_1$

the oxide thickness s. This seems to bring about a contradiction. In fact, as the one-dimensional form of the continuity Eq. (23.3) with $W = 0$ yields $\partial N/\partial t + \partial F/\partial x = 0$, the constancy of F makes N independent of time. However, N does depend on time. This is demonstrated by Fig. 24.7, that shows the linear approximation of the oxidant concentration within the oxide at two different instants, t_1 and $t_2 > t_1$. The value $k\,N_G$ at the source–oxide interface is kept constant by the boundary condition as explained in Sect. 24.3, while the value N_I at the silicon–oxide interface changes with time due to (24.8), and the oxide's thickness changes as well.

The contradiction is eliminated by observing that the continuity Eq. (23.3) has been derived for the case where the boundary is fixed, whereas the growth of thermal oxide is a moving-boundary process. The motion of the boundary is not a rigid one (otherwise the problem could be fixed by moving the reference accordingly), because the oxide volume is actually expanding. In conclusion, Eq. (23.3) must not be used. In fact, the derivation of the linear–parabolic model of the oxide growth (24.10) is based solely on the definition of the flux density.

24.8.2 Elementary Contributions to the Layer's Volume

The relation $\mathrm{d}s/\mathrm{d}t = wF_I$ was used to connect, in Sect. 24.4, the growth velocity of the oxide layer to the flux density of the oxidant and, in Sect. 24.7, the growth velocity of the epitaxial layer to the flux density of $SiCl_4$. The coefficient w is the amount by which one oxidant or $SiCl_4$ molecule makes the volume of the layer to increase. To specify w for the oxidation process one must distinguish between the dry and steam cases. In the first one, each molecule of the oxidant produces one SiO_2 molecule. As a consequence, w is the volume of the SiO_2 molecule. In the steam case, two H_2O molecules are necessary for producing one SiO_2 molecule, hence w is half the volume of the latter. By the same token, in the epitaxial process w is the volume of a Si atom.

24.8.3 Features of the Oxide Growth and Epitaxial Growth

The quadratic term in the left hand side of (24.10) becomes dominant at larger oxide thicknesses. This in turn slows down the growth rate, as shown by (24.9). A qualitative explanation of the phenomenon is easily obtained by considering that, in order to reach the silicon–oxide interface, the oxidant must diffuse across the already-formed oxide. The slope of the oxidant concentration, hence its flux density, decreases with time because the thickness of the oxide region increases, while the value $k\,N_G$ at the source–oxide interface is kept constant by the boundary condition. The decrease in the oxidant concentration N_I at the silicon–oxide interface, shown by (24.8), is not sufficient to contrast the decrease in the concentration's slope. The reasoning above does not apply to the epitaxial growth; in fact, in this case the chemical reaction occurs at the vapor–silicon interface and there is no intermediate layer to be crossed. As a consequence, the corresponding model (24.23) has no quadratic term.

In the analysis of the oxide-growth kinetics carried out in Sect. 24.3 it is assumed that the interface concentration in the source region, N_S, is practically equal to the boundary condition N_G. The simplification is used in the expression (24.6) of the flux density in the oxide layer. The calculation then proceeds by considering only the oxidant diffusion across the already-formed layer and the chemical reaction at the silicon–oxide interface. In this respect, the assumption $N_S = N_G$ has the mere effect of introducing a negligible change in (24.6). A similar approximation would not be possible in the analysis of the epitaxial growth. In fact, letting $N_I = N_G$ in (24.19) would set the flux density to zero. The difference between the two cases is that in the epitaxial growth the flux density exists only in the vapor phase, while in the oxide growth it exists both in the gaseous and solid phases. However, as $D_S \gg D_O$, only the diffusion in the solid phase plays a significant role in determining the kinetics of the oxidation process.

24.8.4 Reaction Velocity

The reaction velocity v_r is among the parameters used in the analysis of the oxide-growth kinetics and epitaxial kinetics. This parameter controls the flux density through (24.7) or (24.20), and is found to depend also on the concentration N_{dop} of dopant atoms in the silicon lattice. The dependence is negligible as long as $N_{\mathrm{dop}} \le n_i(T)$, where n_i (called *intrinsic concentration*, Sect. 18.3) is calculated at the process temperature. When $N_{\mathrm{dop}} > n_i(T)$, the reaction velocity increases with N_{dop}. It should be noted that $n_i \simeq 10^{18}$ cm^{-3} at $T = 1000\,^{\circ}\mathrm{C}$. As a consequence, the dependence of v_r on N_{dop} becomes important only at relatively high dopant concentrations.

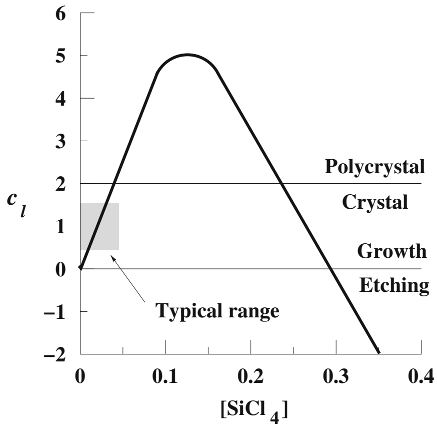


Fig. 24.8 Typical growth velocity c_l of an epitaxial process, expressed in microns per minute, as a function of the mole fraction of tetrachloride. The *shaded area* shows the typical operating range

24.8.5 Molecular Beam Epitaxy

Epitaxy can also be obtained by a process different from CVD, that is called *molecular beam epitaxy* (MBE) and is based on evaporation. The main advantages of MBE are the low-temperature processing and the fine control of the dopant distribution throughout the epitaxial layer. On the other side, MBE has a low throughput and a higher cost. As a consequence, CVD is used is in the majority of cases for growing epitaxial layers in the silicon technology [105, Sect. 6.3].

24.8.6 Secondary Reaction in the Epitaxial Growth

The analysis of the epitaxial kinetics carried out in Sect. 24.7 is based on the hypothesis that only the fundamental reaction (24.16) is present. However, as mentioned in Sect. 24.7, the secondary reaction (24.17) also takes place, which removes silicon atoms from the wafer's surface and, therefore, competes with (24.16). At low concentrations of $SiCl_4$ the effect of the secondary reaction is negligible and the theory of Sect. 24.7 holds; in particular, as shown by (24.23), the growth velocity c_l is proportional to the concentration N_G of $SiCl_4$ in the bulk of the vapor phase. At larger tetrachloride concentrations the dependence of c_l on N_G becomes sublinear as shown in Fig. 24.8. Further increases in N_G make c_l to continuously decrease, and to eventually vanish when the secondary reaction balances the fundamental one. Further on, the secondary reaction prevails and c_l becomes negative, that is, silicon is etched.

The growth velocity also influences the structure of the epitaxial layer. When c_l is low the deposited silicon atoms match the preexisting crystalline structure, so that the newly-formed layer is crystalline as well. Higher values of c_l make the matching more and more difficult. Beyond a critical value of c_l the epitaxial layer is polycrystalline, as sketched in Fig. 24.8. To avoid the growth of a polycrystal it is necessary to keep the $SiCl_4$ concentration in the range shown in the figure.

Problems

24.1 A silicon wafer covered with an $s_i = 0.21$ μm-thick thermal oxide undergoes a second thermal oxidation whose duration is 136 min. Using the values $c_p = 4.43 \times 10^{-2}$ μm^2 h^{-1} for the parabolic coefficient and $c_l = 8.86 \times 10^{-1}$ μm h^{-1} for the linear coefficient, calculate the silicon thickness consumed during the second thermal oxidation.

24.2 Consider a thermal-oxidation process of silicon where the parabolic and linear coefficients are, respectively, 4.43×10^{-2} μm^2 h^{-1} and 8.86×10^{-1} μm h^{-1}. At some instant the oxidant concentration N_I at the silicon–oxide interface is 1×10^{12} cm^{-3}. Find the gradient of the oxidant concentration within the oxide layer at the same instant, expressed in 10^{17} cm^{-4} units.

24.3 A silicon wafer covered with an $s_i = 105$ nm-thick layer of thermal oxide undergoes a second thermal-oxidation process that consumes a 100 nm-thick layer of silicon. Letting the parabolic and linear coefficients be $c_p = 4.43 \times 10^{-2}$ μm^2 h^{-1} and $c_l = 8.86 \times 10^{-1}$ μm h^{-1}, respectively, determine the duration in minutes of the second oxidation process.

24.4 Consider a thermal-oxidation process of silicon where the parabolic and linear coefficients are, respectively, 0.12 μm^2 h^{-1} and 3 μm h^{-1}. At some instant the oxide thickness is 20 nm, and the oxidant concentration in the oxide at the source–oxide interface is $N_O = 3 \times 10^{12}$ cm^{-3}. Find the oxidant concentration at the silicon–oxide interface, expressed in 10^{11} cm^{-3} units.

24.5 A silicon wafer covered with an $s_i = 5$ nm-thick layer of thermal oxide undergoes a thermal-oxidation process that grows a $\Delta s_1 = 7$ nm-thick oxide layer and a successive thermal-oxidation process that grows a $\Delta s_2 = 20$ nm-thick oxide layer. In both processes the parabolic and linear coefficients are, respectively, 4.1×10^{-2} μm^2 h^{-1} and 8.5×10^{-1} μm h^{-1}. Determine the total duration in seconds of the two processes.

24.6 A silicon wafer covered with an $s_i = 80$ nm-thick layer of thermal oxide undergoes a thermal-oxidation process whose linear coefficient is 1 μm h^{-1}. The ratio between the diffusion coefficient of the oxidant within the oxide and the reaction velocity at the silicon–oxide interface is $r = D_O/v_r = 50$ nm. Find how many minutes are necessary to reach a final oxide thickness equal to 150 nm.

24.7 A silicon wafer covered with an $s_i = 40$ nm-thick layer of thermal oxide undergoes a thermal-oxidation process where the oxidant's diffusion coefficient in the oxide is $D_O = 4.5 \times 10^{-6}$ cm^2 s^{-1}, the reaction velocity is $v_r = 4$ cm s^{-1}, and the product of the parabolic coefficient and the process duration is $c_p t_P = 5 \times 10^{-11}$ cm^2. Calculate the final thickness of the oxide in nm.

24.8 A silicon wafer undergoes an epitaxial growth that produces a 12 μm-thick silicon layer. At the end of the process the wafer's weight has increased by 907 mg. Using $p_{\mathrm{Si}} = 2.33$ g cm^{-3} for the specific weight of silicon, determine the wafer's diameter in inches (1 in. = 2.54 cm).

24.9 A 1.2 μm-thick epitaxial layer of silicon is grown by a 1 min-long process in which the reaction velocity is $v_r = 10$ cm s^{-1}. Find the concentration of $\mathrm{SiCl_4}$ at the silicon surface expressed in 10^{16} cm^{-3} units.

24.10 The flux density of $\mathrm{SiCl_4}$ in an epitaxial process in silicon is 8.33×10^{16} cm^{-2}s^{-1}. Remembering that the concentration of the silicon atoms in the crystal lattice is 5×10^{22} cm^{-3}, determine how many minutes are necessary to grow a 2 μm-thick epitaxial layer.

24.11 Determine the reaction velocity v_r (in cm min^{-1}) of an epitaxial process in silicon that in 5 min grows an $s = 2$ μm-thick layer. For the surface concentration of the silicon tetrachloride and the atomic volume of silicon use, respectively, the values 10^{16} cm^{-3} and $(5 \times 10^{22})^{-1}$ cm^3.

24.12 In an epitaxial process the ratio between the $\mathrm{SiCl_4}$ concentration in the bulk of the vapor phase and at the wafer's surface is $a = N_G/N_I = 2$, while the ratio between the reaction velocity and growth velocity is $b = v_r/c_l = 4.87 \times 10^5$. Remembering that the concentration of the silicon atoms in the crystal lattice is 5×10^{22} cm^{-3}, determine the value of N_G in cm^{-3}.