2. Semiconductor Device Technology

2.1. Introduction

In the previous Chapter, we have reviewed the various techniques used to synthesize semiconductor crystals and thin films. This represented only the first step in the fabrication of semiconductor devices. Several additional steps are necessary before a final product can be obtained, which will be described in this and the following Chapter.

In this Chapter, the discussion will be inspired from the silicon device technology because of its technological predominance and maturity in modern semiconductor industry. We will first describe and model the oxidation process used to realize a silicon oxide film. We will then discuss

M. Razeghi, *Technology of Quantum Devices*, DOI 10.1007/978-1-4419-1056-1_2,

the diffusion and ion implantation of dopant impurities in silicon to achieve controlled doping, and review the methods used to characterize their electrical properties. Although this Chapter discusses silicon, the methods can be equally applied to all types of semiconducting materials.

2.2. Oxidation

The ability to form a chemically stable protective layer of silicon dioxide (SiO_2) at the surface of silicon is one of the main reasons that makes silicon the most widely used semiconductor material. This silicon oxide layer is a high quality electrically insulating layer on the silicon surface, serving as a dielectric in numerous devices, that can also be a preferential masking layer in many steps during device fabrication. In this section, we will first review the experimental process of the formation of a silicon oxide. Then we will develop a mathematical model for it and determine the factors influencing the oxidation. We will then end this section by providing details on how to characterize the thickness of the formed oxide.

2.2.1. Oxidation process

A silicon dioxide layer is often thermally formed in the presence of oxygen compounds at a temperature in the range of 900 to 1300 °C. There exists two basic means of supplying the necessary oxygen into the reaction chamber. The first is in gaseous pure oxygen form (dry oxidation) through the reaction: $Si + O_2 \rightarrow SiO_2$. The second is in the form of water vapor (wet oxidation) through the reaction: $Si + 2H_2O \rightarrow SiO_2 + 2H_2$. For both means of oxidation, the high temperature allows the oxygen to diffuse easily through the silicon dioxide. The silicon is consumed as the oxide grows, and with a total oxide thickness of X, about $0.45X$ lies below the original surface of the Si wafer and $0.55X$ lies above it, as shown in Fig. 2.1. A typical oxidation growth cycle consists of dry-wet-dry oxidations, where most of the oxide is grown in the wet oxidation phase. Dry oxidation is slower and results in more dense, higher quality oxides. This type of oxidation method is used mostly for metal-oxide-semiconductor (MOS) gate oxides. Wet oxidation results in much more rapid growth and is used mostly for thicker masking layers.

Before thermal oxidation, the silicon is usually preceded by a cleaning sequence designed to remove all contaminants. Special care must be taken during this step to guarantee that the wafers do not contact any source of contamination, particularly inadvertent contact with a human person. Humans are a potential source of sodium, the element most often responsible for the failure of devices due to surface leakage. Sodium

contamination can be reduced by incorporating a small percentage of chlorine into the oxidizing gas. Next, the cleaned wafers are dried and loaded into a quartz wafer holder called a boat.

The thermal oxidation process is performed with the wafers sitting in the boat loaded into a furnace where the temperature is carefully controlled. Generally, three or four separate furnaces are used in a stack manner, each with its own set of controls and quartzware. The quartz tube inside each furnace is enclosed around heating coils which are controlled by the amount of electrical current running through. A cross-section of a typical oxidation furnace is shown in Fig. 2.1.

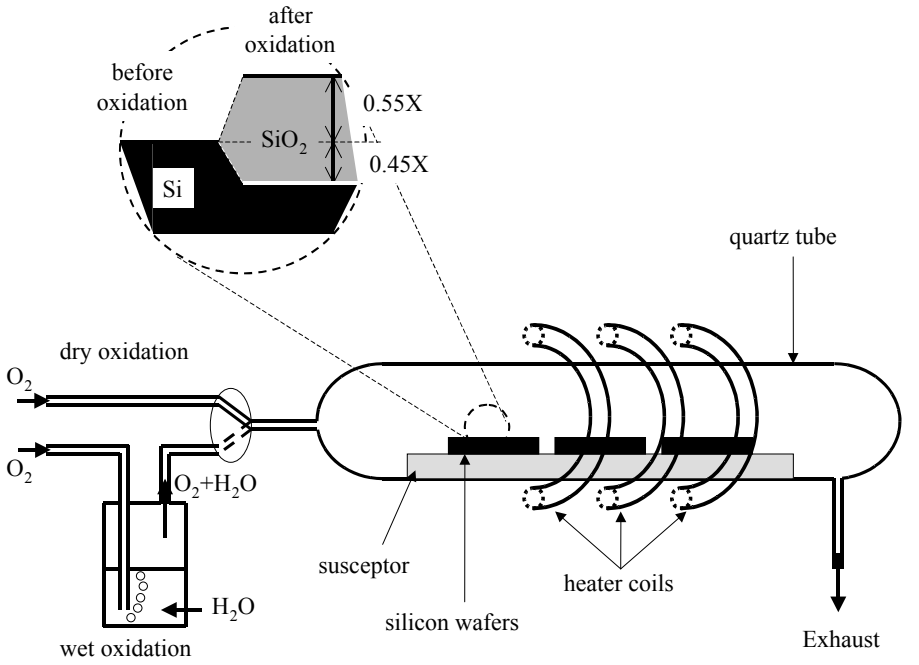

Fig. 2.1. Cross-section of an oxidation furnace: a quartz tube, heated by coils surrounding it, contains the silicon wafers in which either dry oxygen gas or water vapor can be introduced to provide the oxidizing gas. On the top left, the cross-section at the surface of a silicon wafer before and after oxidation is shown.

The furnace is suitable for either dry or wet oxidation film growth by turning a control valve. In the dry oxidation method, oxygen gas is sent into the quartz tube. High-purity gas is used to ensure that no unwanted impurities are incorporated in the layer of oxide as it forms. The oxygen gas can also be mixed with pure nitrogen gas in order to decrease the total cost of running the oxidation process, as nitrogen gas is less expensive than oxygen. In the wet oxidation method, the water vapor introduced into the furnace system is created by flowing a carrier gas into a container or bubbler

filled with ultra pure water and maintained at a constant temperature below its boiling point (100 °C). The carrier gas can be either nitrogen or oxygen, and both result in equivalent oxide thickness growth rates. As the gas bubbles through the water, it becomes saturated with the water vapor. The distance to the quartz oxidation tube must be short enough to prevent condensation of the water vapor. The bubblers used in the wet oxidation process are simple and quite reproducible, but they have two disadvantages associated with the fact that they must be refilled when the water level falls too low: an improper handling of the container can result in the contamination of the water prior or during filling, and the bubbler cannot be filled during an oxidation cycle.

2.2.2. Modeling of oxidation

Using radioactive tracer experiments, the oxygen or water molecules in a dry oxidation process were found to move through the oxide film and react with the silicon atoms at the interface between the oxide film and silicon. As the oxide grows, the growth rate of the oxide layer decreases because the oxygen must pass through more oxide to reach and combine with the silicon. This is schematically illustrated in Fig. 2.2. The movement of these molecules through the forming oxide layer can be mathematically modeled using Fick's first law of diffusion:

Eq. (2.1) $$\Phi_e^{diff} = -D_n \frac{dn}{dx}$$

The objective of the following mathematical model is to determine the growth rate of the oxide layer, that is how fast the oxide layer forms. In this model, we consider that there is a flow of a gas containing oxygen, called the oxidant, onto the sample surface, which we assume diffuses through the existing oxide layer and reacts with the underlying silicon. We will consider three different fluxes (units of particles per $cm^2 \cdot s^{-1}$) of oxidant, each governed by a different physical mechanism. These fluxes are shown in Fig. 2.3.

The objective of the following mathematical model is to determine the growth rate of the oxide layer, that is how fast the oxide layer forms. In this model, we consider that there is a flow of a gas containing oxygen, called the oxidant, onto the sample surface, which we assume diffuses through the existing oxide layer and reacts with the underlying silicon. We will consider three different fluxes (units of particles per $cm^2 \cdot s^{-1}$) of oxidant, each governed by a different physical mechanism. These fluxes are shown in Fig. 2.3.

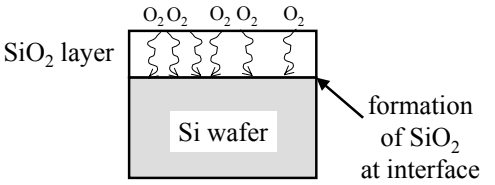

Fig. 2.2. Formation of SiO_2 in a dry oxidation process. The oxygen molecules diffuse through the existing oxide film until they reach the oxide-silicon interface where they react with silicon atoms to continue to form an oxide.

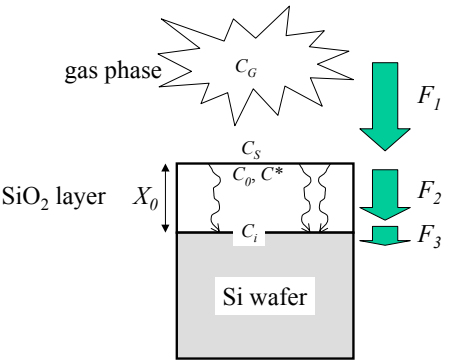

Fig. 2.3. Model for the thermal oxidation of silicon. F_1 represents the flux of oxidant from the bulk gas phase onto the sample surface, F_2 represents the flux of oxidant diffusing through the existing oxide, and F_3 represents the flux of oxidant which reaches the oxide-silicon interface and is consumed through chemical reaction with the silicon.

The first one is the flux of oxidant from the bulk gas phase onto the sample surface, denoted F_1. This flux is proportional to the difference in concentration of oxidant between the bulk gas phase and at the surface of the forming oxide:

Eq. (2.2) $$F_1 = h_G\left(C_G - C_s\right)$$

where h_G is the vapor phase mass transfer coefficient, C_G denotes the oxidant concentration in the bulk gas phase, and C_S denotes that at the surface of the forming oxide. These concentrations are generally different because some oxidant is consumed in the oxidation process. These concentrations are directly related to the partial pressures of the oxidant gas in the bulk gas phase, P_G, and at the oxide surface, P_S, through the ideal gas law:

Eq. (2.3)
$$\begin{cases} C_G = \dfrac{P_G}{k_b T} \\ C_S = \dfrac{P_S}{k_b T} \end{cases}$$

where k_b is the Boltzmann constant and T is the absolute temperature.

We can relate the oxidant concentration in the gas with the oxidant concentration in the solid phase, i.e. the oxide layer, near the surface through Henry's law:

Eq. (2.4)
$$C_0 = K_H P_S$$

where C_0 is the oxidant concentration inside the oxide layer just below its surface, K_H is Henry's law constant and P_S is the partial pressure of the oxidant in the gas phase at the oxide surface.

It will be convenient to introduce the equilibrium value of C_0, which will be denoted C^*. This concentration is related to the partial pressure in the bulk of the gas P_G through:

Eq. (2.5)
$$C^* = K_H P_G$$

Combining Eq. (2.3), Eq. (2.4) and Eq. (2.5), Eq. (2.2) can then be successively written as:

Eq. (2.6)
$$F_1 = h_G \left[\frac{P_G}{k_b T} - \frac{P_S}{k_b T} \right] = \frac{h_G}{k_b T} \left[\frac{C^*}{K_H} - \frac{C_0}{K_H} \right] = h\left[C^* - C_0 \right]$$

where we have defined:

Eq. (2.7)
$$h = \frac{h_G}{k_b K_H T}$$

The second flux, denoted F_2, to consider is that of the oxidant diffusing through the oxide layer already present which can be expressed as:

Eq. (2.8)
$$F_2 = \frac{D}{X_0}\left(C_0 - C_i \right)$$

where D is the diffusion coefficient of the oxidant through the oxide, C_i is the oxidant concentration at the oxide-silicon interface, and X_0 is the thickness of the oxide.

The third flux, denoted F_3, corresponds to the incorporation of oxidant molecules which reach the oxide-silicon interface and react chemically to expand the oxide. This can be expressed as:

Eq. (2.9) $$F_3 = k_s C_i$$

where k_S is the chemical reaction constant for the formation of oxide.

Under steady-state conditions, these three fluxes must be equal:

Eq. (2.10) $$F_1=F_2=F_3=F$$

This gives us three equations, for the three unknowns: C_0, C^* and C_i. Using Eq. (2.8) and Eq. (2.9) to equate F_2 and F_3, we get:

Eq. (2.11) $$C_0 = \left(1 + \frac{k_S X_0}{D}\right) C_i$$

Now, using Eq. (2.6) and Eq. (2.9) to equate F_1 and F_3, we get:

Eq. (2.12) $$C^* = C_0 + \frac{k_s}{h} C_i$$

which, after considering Eq. (2.11), becomes:

Eq. (2.13) $$C^* = \left(1 + \frac{k_S X_0}{D} + \frac{k_S}{h}\right) C_i$$

It is convenient to rearrange these relations to express C_0 and C_i as a function of C^*:

Eq. (2.14) $$C_i = \frac{1}{\left(1 + \frac{k_S X_0}{D} + \frac{k_S}{h}\right)} C^*$$

Eq. (2.15) $$C_0 = \frac{\left(1 + \frac{k_S X_0}{D}\right)}{\left(1 + \frac{k_S X_0}{D} + \frac{k_S}{h}\right)} C^*$$

We can now consider a particular case. If we assume that $h >> k_s$, i.e. the oxidation reaction at the oxide-silicon interface is much slower than the arrival of oxidant at the oxide surface, the oxidation process is then said to be interfacial reaction controlled. The Eq. (2.14) and Eq. (2.15) can then be simplified into:

Eq. (2.16) $$\begin{cases} C_i \approx \dfrac{1}{\left(1 + \dfrac{k_s X_0}{D}\right)} C^* \\ C_0 \approx C^* \end{cases}$$

Combining Eq. (2.9) and Eq. (2.16) to eliminate C_i, we can express the flux F as a function of C_0:

Eq. (2.17) $$F = \frac{k_s}{\left(1 + \frac{k_s X_0}{D}\right)} C_0$$

The rate at which the oxide layer grows is then given by the flux divided by the number N of oxidant molecules that can be incorporated into a unit volume of oxide:

Eq. (2.18) $$\frac{dX_o}{dt} = \frac{F}{N} = \frac{1}{N} \frac{k_s C_0}{\left(1 + \frac{k_s X_0}{D}\right)}$$

For dry oxidation, $N = 2.2 \times 10^{22}$ molecules per cm^3, while for wet oxidation $N = 4.4 \times 10^{22}$ molecules per cm^3. Integrating Eq. (2.18) and using the boundary condition $X_0(t=0) = X_i$, yields the following equation for X_0:

Eq. (2.19) $$X_o^2 + AX_o = B(t + \tau)$$

where τ is an integration constant and where we have denoted:

$$\text{Eq. (2.20)} \qquad \begin{cases} A = \dfrac{2D}{k_s} \\ B = \dfrac{2DC_0}{N} \\ \tau = \dfrac{X_i^{\,2}}{B} + \dfrac{X_i}{B/A} \end{cases}$$

where X_i is the initial thickness of the oxide. For dry oxidation, an initial oxide thickness of 250 Å must be accounted for by letting $X_i = 25$ nm, in order to make Eq. (2.19) universal to both oxidation methods.

Solving for the oxide thickness in Eq. (2.19) as a function of oxidation time t, one obtains the following positive expression for X_0:

$$\text{Eq. (2.21)} \qquad X_0 = \frac{A}{2}\left\{\sqrt{1 + \frac{(t+\tau)}{A^2/4B}} - 1\right\}$$

The growth time t is given directly by Eq. (2.19):

$$\text{Eq. (2.22)} \qquad t = \frac{A^2}{4B}\left[\left(\frac{2X_0}{A} + 1\right)^2 - 1\right] - \tau$$

For the limiting case of "short oxidation time", where $(t+\tau) << A^2/4B$, we can simplify the expression in Eq. (2.21):

$$\text{Eq. (2.23)} \qquad X_0 \approx \frac{A}{2}\left\{1 + \frac{1}{2}\frac{(t+\tau)}{A^2/4B} - 1\right\}$$

which is obtained after considering the Taylor expansion of the square root. We then obtain the so-called *linear oxidation law*:

$$\text{Eq. (2.24)} \qquad X_0 = \frac{B}{A}(t+\tau)$$

where B/A is the linear growth rate constant, and can be calculated using Eq. (2.20) and Table 2.1.

For the other limiting case of "long oxidation time", when $t>>A^2/4B$, one obtains the *parabolic oxidation law*:

Eq. (2.25) $$X_o^2 = Bt$$

where B is the parabolic growth rate constant, and can be calculated using Eq. (2.20) and Table 2.1.

2.2.3. Factors influencing oxidation rate

Numerous factors can influence the oxidation rate by governing each of the mechanisms discussed in the previous model. For example, one of them is the diffusion coefficient in Eq. (2.8). This parameter generally follows an Arrhenius relationship as given by:

Eq. (2.26) $$D = D_0 \exp\left(-\frac{E_A}{k_b T}\right)$$

where k_b is the Boltzmann constant, E_A is the activation energy, and T is the temperature. Values for activation energy and D_0 coefficient can be found in Table 2.1. This relation indicates the strong dependence of oxide growth rate on temperature as the diffusion rate of the oxidant increases exponentially with temperature.

There exist four other factors which are commonly known to affect the oxidation rate of silicon: type of oxidation, orientation of the silicon wafer, pressure and impurity effects. For the type of oxidation, wet oxidation has a higher growth rate due to the higher solubility of the water vapor. The orientation dependence of the oxidation rate can be easily understood because the oxidation process depends on the total number of available Si atoms per unit area for oxidation at the oxide-silicon interface. Only the linear oxidation rate is expected to significantly change as a function of orientation, i.e. for short oxidation durations. For example, the oxidation rate for (111) oriented Si is faster than that for (100) oriented Si initially, in the linear region, as shown in Fig. 2.4(a) and (b). As the oxidation kinetics change from the linear rate to the parabolic rate, i.e. for longer oxidation durations, the difference between the two orientations diminishes. The pressure is proportional to the number of oxidants, and is directly proportional to both linear and parabolic growth rate constants. As can be seen in Eq. (2.19), an increase in pressure results in a slower growth rate.

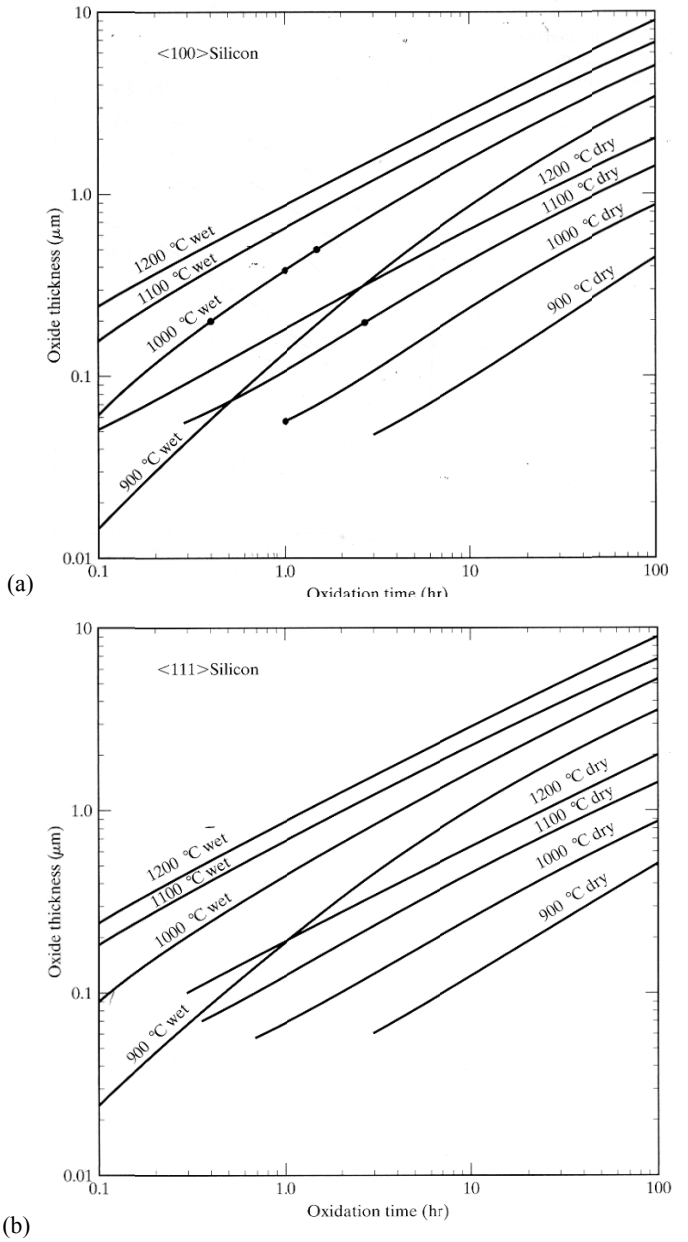

Fig. 2.4. Oxide thickness as a function of oxidation time under various conditions: (a) wet and dry oxidation of (100) silicon at several temperatures, (b) wet and dry oxidation of (111) silicon at various temperatures. [Jaeger, R.C., Introduction to Microelectronics Fabrication: Vol. 5 of Modular Series on Solidstate Devices, 2nd Edition, Fig. 3.6, p. 35.© 2002. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.]

	Wet O_2 (X_i=0 nm)		Dry O_2 (X_i=25 nm)	
	D_0	E_A (eV)	D_0	E_A (eV)
<100> Si Linear	9.7×10^7 μm·hr^{-1}	2.05	3.71×10^6 μm·hr^{-1}	2.00
<100> Si Parabolic	386 μm^2·hr^{-1}	0.78	772 μm^2·hr^{-1}	1.23
<111> Si Linear	1.63×10^8 μm·hr^{-1}	2.05	6.23×10^6 μm·hr^{-1}	2.00
<111> Si Parabolic	386 μm^2·hr^{-1}	0.78	772 μm^2·hr^{-1}	1.23

Table 2.1. D_0 coefficient values and activation energy E_A for wet and dry oxygen for different types of silicon. [Jaeger et al. 1988.]

2.2.4. Oxide thickness characterization

The accurate measurement of the thickness of a dielectric film such as silicon dioxide is very important in the fabrication of optoelectronic devices. Various techniques are available for measuring this oxide thickness, including optical interference, ellipsometry, capacitance, and the use of a color chart.

The optical interference method is a simple and nondestructive technique, which can be used to routinely measure thermal oxide thickness from less than 100 Å to more than 1 μm. The method is based on characterizing the interference pattern created by light reflected from the air/SiO_2 interface and that from the Si/SiO_2 interface, as illustrated in Fig. 2.5.

The equation governing this interference is:

Eq. (2.27) $$X_0 = \frac{\lambda(g - \Delta\varphi)}{2n^*}$$

where X_0 is the thickness of the oxide, λ is the wavelength of the incident radiation, g the order of the interference, and $\Delta\varphi$ is the net phase shift and is equal to $\varphi_s - \varphi_o$ where φ_o is the phase shift at the air/SiO_2 interface and φ_s is the phase shift at the Si/SiO_2 interface. The parameter n^* is given by:

Eq. (2.28) $$n^* = \sqrt{\overline{n}_i^2 - \sin^2\theta}$$

where $\overline{n}_i$ is the refractive index of the oxide film and θ is the angle of incidence of the light relative to the wafer. All these parameters are illustrated in Fig. 2.5.

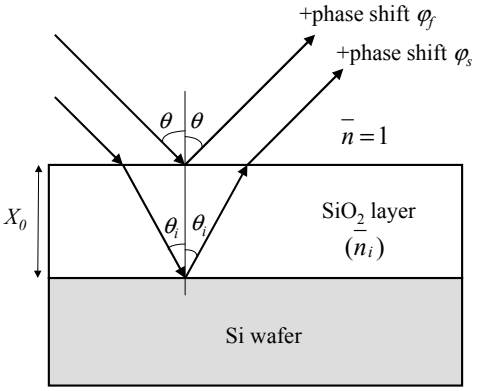

Fig. 2.5. Optical interference method for the measurement of oxide film thickness. Two rays of light with the same wavelength are shown incident on the wafer. One of them is reflected from the oxide-air interface. The other enters the oxide layer which has a different refractive index than air and is reflected at the oxide-silicon interface. A difference in optical path occurs between these two rays of light and a phase shift difference results. If the phase shift difference is an integer multiple of 2π, these two reflected rays of light interfere constructively, whereas if the phase shift difference is a half integer multiple of 2π, these rays interfere destructively.

The second method for the measurement of the oxide film thickness is ellipsometry. Ellipsometry is the most popular technique used to assess the properties of silicon dioxide films. Ellipsometry provides a non-destructive technique for accurately determining the oxide thickness, as well as the refractive index at the measuring wavelength. An illustration of an ellipsometry system is shown in Fig. 2.6. It is the most widely used tool to measure the refractive index of a wide variety of materials on any substrate, in particular SiO_2, Si_3N_4, photoresist, and aluminum oxide (Al_2O_3) on silicon substrates. Such systems can measure film thickness in the range of 20 Å to 60,000 Å with an accuracy of ±2%. An ellipsometer operates by shining polarized monochromatic light onto the wafer surface at an angle. The light is then reflected from both the oxide and the silicon surface. A phase modulation unit, numerical data acquisition and processing system work together to measure the difference in polarization. The result is then used to calculate the oxide thickness.

Fig. 2.6. A typical ellipsometer system, including a light source and its power supply, a sample stage, a detector and the analyzing circuits.

The third oxide film thickness measurement technique is the capacitance method, which requires the fabrication of a metal-oxide-semiconductor (MOS) capacitor. The oxide thickness is given by the following equation:

$$\text{Eq. (2.29)} \qquad X_0 = \frac{A_g \varepsilon_{ox} \varepsilon_0}{C_{ox}}$$

where C_{ox} is the experimentally measured oxide capacitance, A_g is the area of the capacitor, ε_{ox} is the dielectric constant of the oxide film, and ε_o the permittivity in vacuum.

Finally, the fourth and simplest method used to measure an oxide film thickness is by comparing the film color with a calibrated chart as shown in Table 2.2 for SiO_2. Each oxide thickness has a specific color when it is viewed under white light perpendicular to its surface. The colors are cyclically repeated for different orders of reflection.

Film thickness (μm)	Color	Film thickness (μm)	Color
0.05	Tan	0.68	Bluish
0.07	Brown	0.72	Blue green to green
0.10	Dark violet to red violet	0.77	"Yellowish"
0.12	Royal blue	0.80	Orange
0.15	Light blue to metallic blue	0.82	Salmon
0.17	Metallic to very light yellow green	0.85	Dull, light red violet
0.20	Light gold to yellow, metallic	0.86	Violet
0.22	Gold with slight yellow orange	0.87	Blue violet
0.25	Orange to melon	0.89	Blue
0.27	Red violet	0.92	Blue green
0.30	Blue to violet blue	0.95	Dull yellow green
0.31	Blue	0.97	Yellow to "yellowish"
0.32	Blue to blue green	0.99	Orange
0.34	Light green	1.00	Carnation pink
0.35	Green to yellow green	1.02	Violet red
0.36	Yellow green	1.05	Red violet
0.37	Green yellow	1.06	Violet
0.39	Yellow	1.07	Blue violet
0.41	Light orange	1.10	Green
0.42	Carnation pink	1.11	Yellow green
0.44	Violet red	1.12	Green
0.46	Red violet	1.18	Violet
0.47	Violet	1.19	Red violet
0.48	Blue violet	1.21	Violet red
0.49	Blue	1.24	Carnation pink to salmon
0.50	Blue green	1.25	Orange
0.52	Green	1.28	"Yellowish"
0.54	Yellow green	1.32	Sky blue to green blue
0.56	Green yellow	1.40	Orange
0.57	Yellow to "yellowish"	1.45	Violet
0.58	Light orange or yellow to pink borderline	1.46	Blue violet
0.60	Carnation pink	1.50	Blue
0.63	Violet red	1.54	Dull yellow green

Table 2.2. SiO_2 oxide film color chart.

Example

Q: Using the Deal-Grove oxidation model, calculate the time needed to grow a 150 nm thick oxide on top of (100) silicon by wet oxidation at a temperature of 1000 ºC.

A: T=1000 ºC=1273 K

From Table 2.1 we obtain the value of the pre-exponential factor D_0=3.71 × 10^6 μm·hr^{-1} (linear oxidation) and E_A = 2.00 eV. Using Eq. (2.26) we determine the diffusion coefficient D and then the ratio:

$$B/A = 3.71\times10^{6}\exp[-2.00/(8.617\times10^{-5})(1273)]$$
$$= 0.04478\,\mu m\cdot hr^{-1}$$

From Table 2.1 we obtain D_0 = 772 $\mu m^2 \cdot hr^{-1}$ (parabolic oxidation) and E_A = 1.23 eV. Using Eq. (2.26) we calculate D and then the value for B:

$$B = 772\exp[-1.23/(8.617\times10^{-5})(1273)]$$
$$= 0.01042\,\mu m^2\cdot hr^{-1}$$

We can find the necessary oxidation time by using Eq. (2.19), which when rearranged becomes:

$$t = \frac{X_0^{\,2}}{B} + \frac{X_0}{B/A} - \tau$$

Since we are using wet oxidation, X_i = 0 so τ = 0,

$$\Rightarrow \quad t = \frac{(0.15)^2}{0.04478} + \frac{0.15}{0.10420} = 5.5\ .$$

Therefore, 5.5 hours is needed to grow a 150 nm thick oxide layer on (100) silicon using wet oxidation at 1000 ºC.

2.3. Diffusion of dopants

Doping is a method to control the electrical properties in semiconductors. Doping is achieved by replacing the constituting atoms of the semiconductor with atoms which contain fewer or more electrons. Through doping, the crystal composition is thus slightly altered so that it contains either a higher

concentration of electrons or holes, which makes the semiconductor *n*-type or *p*-type, respectively.

The doping of semiconductors can be performed during the bulk crystal or epitaxial film growth by introducing the dopant along with the precursor chemicals. This way, the entire crystal or film is uniformly doped with the same concentration of dopants. Another method consists of carrying out the doping after the film deposition by performing the diffusion or the implantation of dopants. These have the advantage that the doping can be localized to certain regions only, by using an adequate mask to prevent the doping in undesired areas. In this section, we will focus on the diffusion of dopants and we will illustrate our discussion with the doping of silicon.

The diffusion of dopants in compound semiconductor epitaxial films generally follows a similar model. However, the effects of diffusion doping in compound semiconductor heterostructures are subtler and have been discussed in detail in specialized texts [Razeghi 1989].

2.3.1. Diffusion process

Diffusion is the process whereby a particle moves from regions of higher concentrations to regions of lower concentrations. The process could be visualized by thinking of a drop of black ink dropped into a glass of clean water. Initially, the ink stays in a localized area, appearing as a dark region in the clean water. Gradually, some of the ink moves away from the region of high concentration, and instead of there being a dark region and a clean region, there is a graduation of colors. As time passes, the ink spreads out until it is possible to see through it. Finally, after a very long time, a steady state is reached and the ink is uniformly distributed in the water. The movement of the ink from the region of high concentration (ink drop) to the region of low concentration (the rest of the glass of water) is an illustration of the process of diffusion.

In the doping of silicon by diffusion, the silicon wafer is placed in an atmosphere containing the impurity or dopant to incorporate. Because the silicon does not initially contain the dopant in its lattice, we are in the presence of two regions with different concentrations of impurities. At high temperatures (900–1200 °C), the impurity atoms can move into the crystal and diffusion can therefore occur, as schematically illustrated in Fig. 2.7.

The wafers are loaded vertically into a quartz boat and put into a furnace similar to the furnace used for oxidation. There are three types of sources to be used for the dopant atoms: solid, liquid, and gas, as shown in Fig. 2.8.

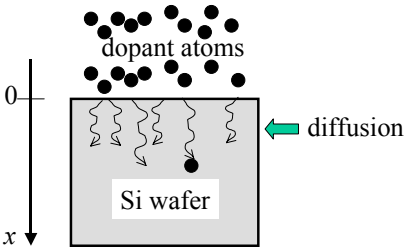

Fig. 2.7. Diffusion of dopants in a silicon wafer. The wafer is placed in an atmosphere containing the dopant. The gradient of dopant concentration between the atmosphere and the silicon crystal leads to their diffusion into the silicon.

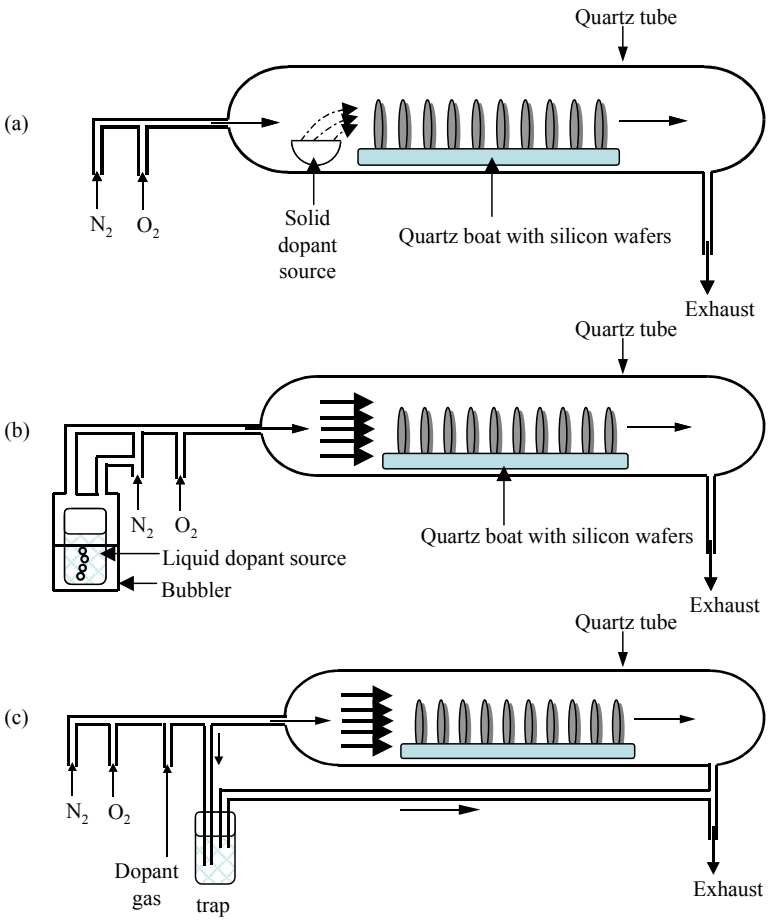

Fig. 2.8. Diffusion furnaces. (a) solid source diffusion with the source in a platinum source boat, (b) liquid source diffusion with the carrier gas passing through the bath, and (c) gas source diffusion with gaseous impurity sources.

There exist several types of diffusion mechanisms. An impurity can diffuse into an interstitial site in the lattice and can move from there to another interstitial site, as shown in Fig. 2.9(a). We then talk about interstitial diffusion Sometimes a silicon atom can be knocked into an interstitial site, leaving a vacancy in the lattice where a diffusing dopant atom can fit, as shown in Fig. 2.9(b). A third possible mechanism consists of a dopant directly diffusing into a lattice vacancy (Fig. 2.9 (c)). We then talk about substitutional diffusion. It is only in the cases that an impurity occupies a vacated lattice site that *n*-type or *p*-type doping can occur. The presence of such vacancies in the lattice can be due to defects or to heat which increases atomic vibrations thus giving enough energy to the silicon atoms to move out of their equilibrium positions into interstitial sites.

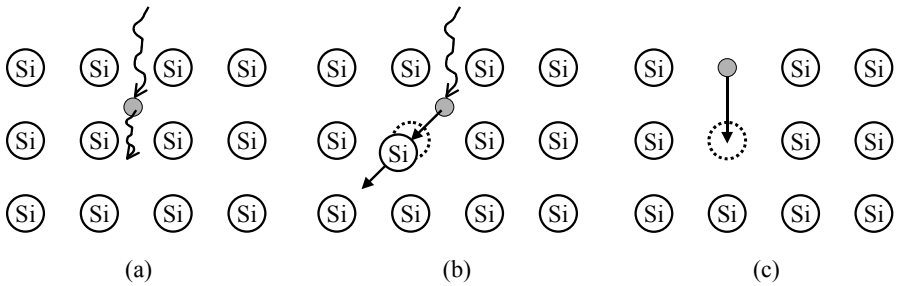

Fig. 2.9. Three possible diffusion mechanisms in a silicon wafer: (a) an impurity moves from one interstitial site to another, (b) a silicon atom is knocked into an interstitial site, thus leaving a vacancy which can be occupied by a diffusing impurity, (c) an impurity diffuses directly into a vacancy.

There are many different types of impurities that can be used for diffusion, the most common being boron, phosphorus, arsenic, and antimony. Table 2.3 lists the reactions for the materials for the three different types of diffusion sources.

The rate at which the diffusion of impurities takes place depends on how fast they are moving through the lattice. This phenomenon is quantitatively characterized by the diffusion coefficient of the impurity in silicon. Table 2.4 lists diffusion coefficient values for common impurities in silicon. We can then model the diffusion process by combining Fick's first and second law of diffusion:

Eq. (2.30) $\partial N / \partial t = D\partial^2 N / \partial x^2$

Impurity	Type	Reaction
Boron	Solid	$2(CH_3O_3)B + 9O_2 \rightarrow B_2O_3 + 6CO_2 + 9H_2O$
	Liquid	$4BBr_3 + 3O_2 \rightarrow 2B_2O_3 + 6Br_2$
	Gas	$2B_2O_3 + 3Si \rightarrow 4B + 3SiO_2$
Phosphorus	Solid	$2P_2O_5 + 5Si \rightarrow 4P + 5SiO_2$
	Liquid	$4POCl_3 + 3O_2 \rightarrow 2P_2O_5 + 6Cl_2$
	Gas	$2PH_3 + 4O_2 \rightarrow P_2O_5 + 3H_2O$
Arsenic	Solid	$2As_2O_3 + 3Si \rightarrow 3SiO_2 + 4As$
Antimony	Solid	$2Sb_2O_3 + 3Si \rightarrow 3SiO_2 + 4Sb$

Table 2.3. Diffusion reactions for common impurity types.

Element	D_0 ($cm^2 \cdot s^{-1}$)	E_A (eV)
B	10.50	3.69
Al	8.00	3.47
Ga	3.60	3.51
In	16.5	3.9
P	10.5	3.69
As	0.32	3.56
Sb	5.6	3.95

Table 2.4. Diffusion coefficient and activation energy values for common impurities in silicon.

The technology of diffusion in semiconductor processing consists of introducing a controlled amount of chosen impurities into selected regions of the semiconductor crystal. To prevent the diffusion of dopants in undesired areas, it is common to use a dielectric mask such as SiO_2 to selectively block the diffusion as show in Fig. 2.10. Fig. 2.11 shows a plot of the minimum mask thickness needed for a given diffusion time for boron and phosphorus diffusion.

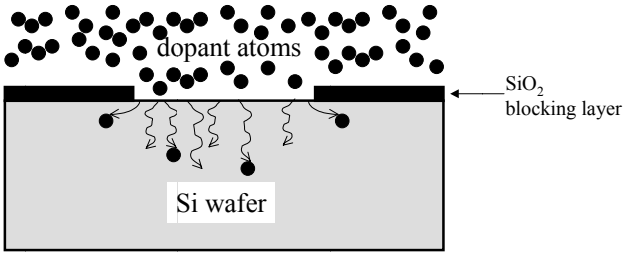

Fig. 2.10. Schematic illustration of the selective diffusion in a silicon wafer. The SiO_2 layer acts as a blocking layer for the diffusion of dopant atoms. Some dopant atoms can diffuse laterally under the blocking layer to some extent.

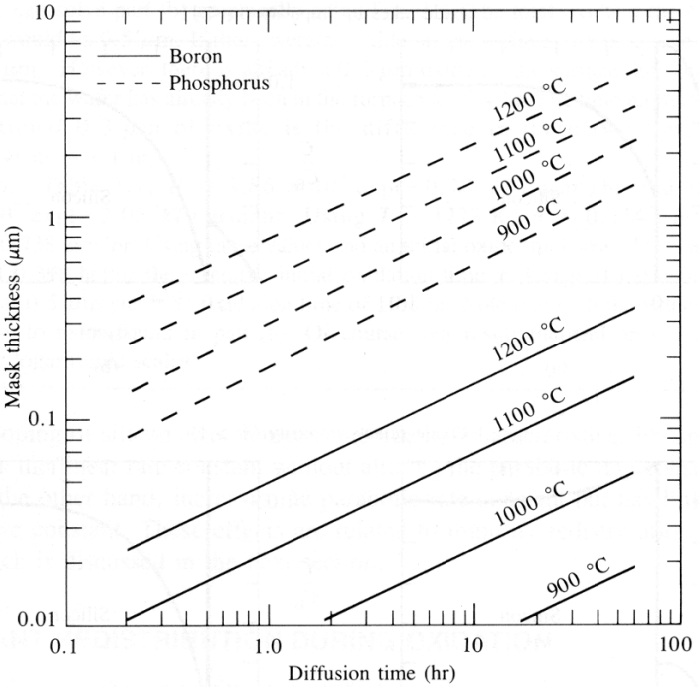

Fig. 2.11. Minimum SiO_2 mask thickness needed for successful diffusion of boron and phosphorus in silicon for a given temperature and time. [Jaeger, R.C., Introduction to Microelectronics Fabrication: Vol. 5 of Modular Series on Solidstate Devices, 2nd Edition,© 2002, p.40, fig 3.10. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.]

There are two major techniques for conducting diffusion, depending on the state of the dopant on the surface of the wafer: (1) constant-source diffusion, also called predeposition or thermal predeposition, in which the concentration of the desired impurity at the surface of the semiconductor is kept constant; and (2) limited-source diffusion, or drive-in, in which a fixed

total quantity of impurity is diffused and redistributed into the semiconductor to obtain the final profile.

2.3.2. Constant-source diffusion: predeposition

During predeposition, the silicon wafer is heated to a specific temperature, and an excess of the desired dopant is maintained above the wafer. The dopants diffuse into the crystal until their concentration near the surface is in equilibrium with the concentration in the surrounding ambient above it. At a given temperature, the maximum concentration that can be diffused into a solid is called the solid solubility. Having more dopants available outside the solid than can enter the solid guarantees that solid solubility will be maintained during the predeposition. For example, the solid solubility of phosphorus in silicon at 1000 °C is 9×10^{20} atoms/cm^3, while for boron in silicon at the same temperature it is only 2×10^{20} atoms/cm^3. These values only depend on the temperature, for a given dopant in a given semiconductor. As a result, the substrate temperature also determines the concentration of the dopant at the surface of the crystal wafer during diffusion.

Under predeposition conditions, let us denote N_0 the dopant concentration in the wafer near the surface. N_0 would be equal to the solid solubility of the dopant at the predeposition temperature if the excess dopant in the ambient above the wafer is sufficient. The concentration of dopant in the crystal at a depth x below the surface and after a diffusion time t can be known and is equal to:

Eq. (2.31) $$N(x,t) = N_0 erfc\left[\frac{x}{2\sqrt{Dt}}\right]$$

where D is the diffusion coefficient of the dopant at the predeposition temperature and *erfc* refers to the complementary error function. The complementary error function is found by complementing the integral of the normalized Gaussian function, and is shown in Fig. 2.12:

Eq. (2.32) $$erfc(\bar{x}) = 1 - erf(\bar{x}) = \frac{2}{\sqrt{\pi}} \int_{\bar{x}}^{\infty} e^{-t^2} dt$$

The shape of the dopant concentration function is shown in Fig. 2.13 for several values of the product Dt. We see that, as the diffusion coefficient increases or, equivalently, as the diffusion time increases, the dopant reaches deeper into the crystal. The surface concentration remains the same at N_0. The concentration N_B represents the background carrier concentration and

refers to the concentration of majority carriers in the semiconductor before diffusion. The value of x for which $N(x,t)$ is equal to N_B is conventionally termed the junction depth.

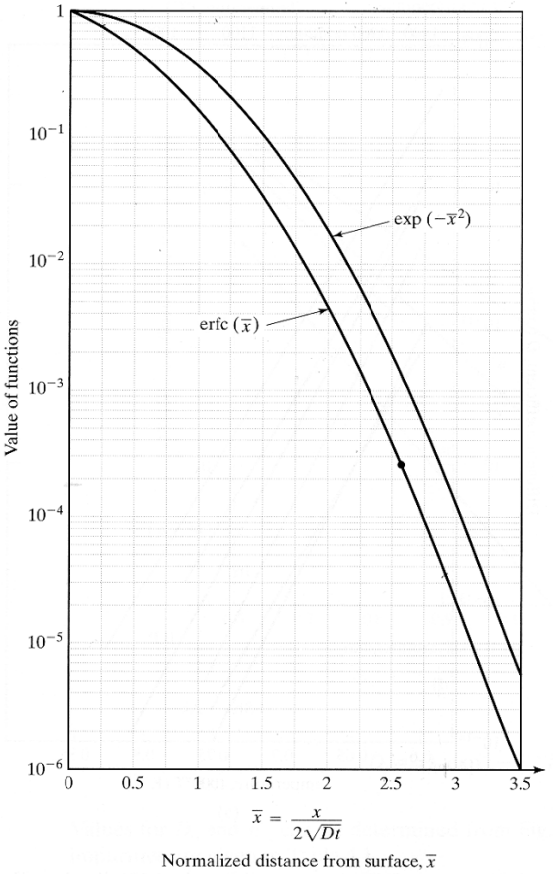

Fig. 2.12. Complementary error function, used in the calculation of the dopant concentration. [Jaeger, R.C., Introduction to Microelectronics Fabrication: Vol. 5 of Modular Series on Solidstate Devices, 2nd Edition, © 2002, p.54, fig. 4.4. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.]

The total amount of impurities Q introduced per unit area, also called the dose, after a diffusion duration t in a predeposition process is found by integrating the function in Eq. (2.31) for values of $x > 0$, which leads to the following expression:

Eq. (2.33) $$Q(t) = N_0\sqrt{\frac{4Dt}{\pi}}$$

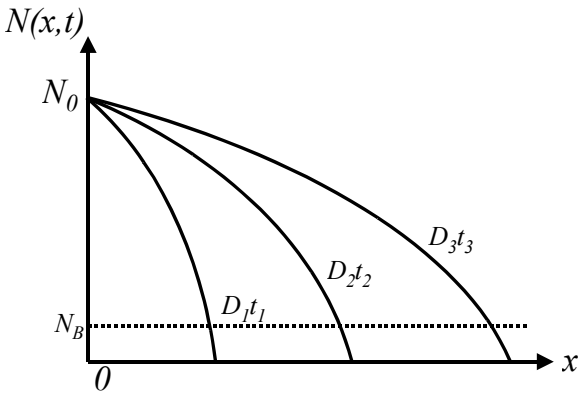

Fig. 2.13. Semi-logarithmic graph of the complementary error function, representing the dopant concentration in the crystal during predeposition, where the surface concentration is kept constant, for several values of Dt: $D_3t_3>D_2t_2>D_1t_1$. As the diffusion coefficient and/or the diffusion time is increased, the dopant reaches deeper into the crystal.

2.3.3. Limited-source diffusion: drive-in

Unlike predeposition, the drive-in diffusion process is carried out with a fixed total amount of impurity. This method allows us to better control the resulting doping profile and depth, which are important parameters in the fabrication of semiconductor devices.

During drive-in, the parameters which can be controlled include the duration of diffusion, the temperature, and the ambient gases. The dopant concentration profile of the drive-in has the shape of a Gaussian function, as shown in Fig. 2.14. In this type of diffusion, the dose remains constant causing the surface concentration to decrease. This relationship explains the shape of the curve, which can be expressed by solving Eq. (2.34) and using the boundary condition that the impurity concentration at the surface is equal to the dose:

Eq. (2.34) $$N(x,t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left[\frac{-x^2}{4Dt}\right]$$

which is expressed in units of atoms per unit volume. D is the diffusion coefficient of the impurity at the drive-in temperature and t is the drive-in time. The drive-in can be performed after a predeposition step, in a high temperature diffusion furnace once the excess dopant remaining on the surface of the wafer from the predeposition step has been removed. In this case, Q is the total dose introduced into the crystal during a predeposition step.

The limited-source diffusion process is ideally suited when a relatively low value of surface concentration is needed in conjunction with a high diffusion depth. Typically, a short period of constant-source diffusion is followed by a period of limited-source diffusion. Predeposition is used to establish the dose into a shallow layer of the surface creating a diffusion front. Then the drive-in step moves this diffusion front to the desired depth.

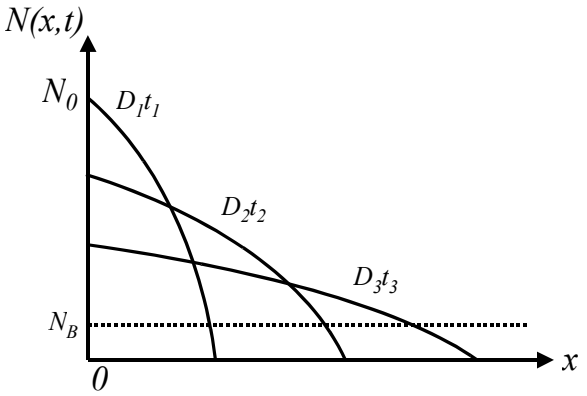

Fig. 2.14. Semi-logarithmic graph of the dopant concentration in the crystal during drive-in for several values of Dt: $D_3t_3>D_2t_2>D_1t_1$. As the diffusion coefficient and/or the diffusion time is increased, the dopant reaches deeper into the crystal. At the same time, the concentration at the surface is reduced because the drive-in is a limited-source diffusion process.

2.3.4. Junction formation

When diffusing *p*-type impurity dopants in an originally *n*-type doped semiconductor, a *p-n* junction can be formed, as shown in Fig. 2.15. In fact, the purpose of most diffusion processes is to form a *p-n* junction by changing a region of an *n*-type semiconductor into a *p*-type or vice versa.

Let us consider the example of an *n*-type doped silicon wafer which exhibits a background concentration N_B, and a *p*-type diffusing impurity with surface concentration N_0. Where the diffusing impurity profile concentration intersects the background concentration N_B, a metallurgical junction depth, x_j, is formed as shown in Fig. 2.15.

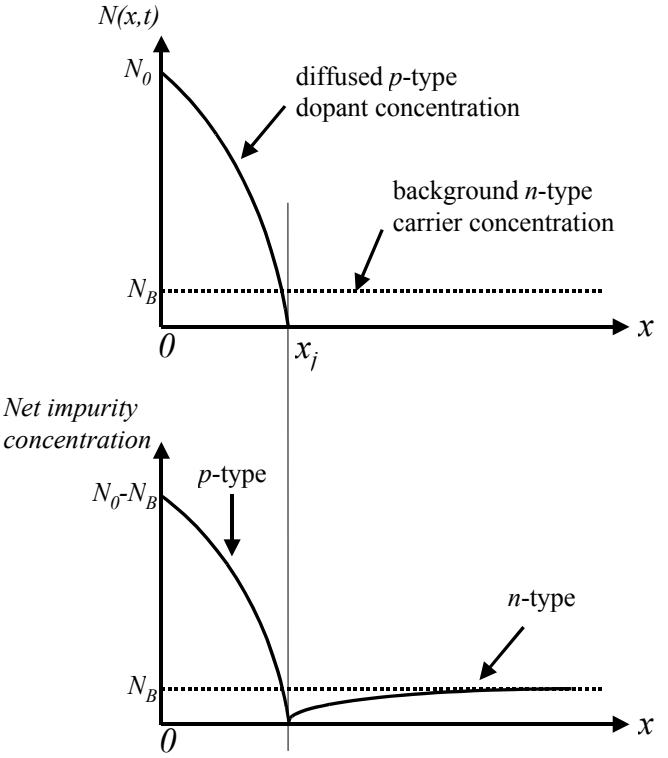

Fig. 2.15. Semi-logarithmic graphs illustrating the formation of a p-n junction through diffusion. A p-type dopant is diffused into an n-type semiconductor which has a background concentration of N_B. The p-type dopant concentration profile after diffusion is shown in the top graph. The p-n junction will occur where the p-type dopant concentration is equal to the n-type background concentration as shown on the bottom graph.

At the metallurgical junction depth, the background concentration is equal to the surface concentration, so the net impurity concentration is zero. In the predeposition process with a complementary diffusion profile, the junction depth is found by solving Eq. (2.31) and using the boundary condition that $N(x_j,t) = N_B$:

Eq. (2.35) $$x_j = \left(2\sqrt{Dt}\right) erfc^{-1}\left(\frac{N_B}{N_0}\right)$$

where $erfc^{-1}$ refers to the reciprocal function of the complementary error function. In the drive-in process with a Gaussian diffusion profile, the junction depth is found by solving Eq. (2.34) and using the boundary condition that $N(x_j,t) = N_B$:

Eq. (2.36) $$x_j = 2\sqrt{Dt \ln\left(\frac{N_0}{N_B}\right)}$$

By successively diffusing two impurities of different types into an originally doped wafer, more complex structures can be achieved, such as for example an *n*-*p*-*n* transistor structure as illustrated in Fig. 2.16. The starting wafer would be an *n*-type (with a background concentration N_C in this example), the first diffusion process would introduce *p*-type dopants (N_B in this example) and the second diffusion would introduce *n*-type impurities (N_E) such that $N_E >> N_B >> N_C$.

Example

Q: Calculate the dose for a boron diffusion process at 1000 ºC for 30 minutes using an *n*-type silicon substrate with a concentration of 10^{19} cm^{-3}.

A: T=1000 ºC = 1273 K

T = 30 min = 1800 s

From Table 2.4, boron has diffusion coefficient value D_0 = 10.5 cm^2·s^{-1} and activation energy E_A = 3.69 eV. Using Eq. (2.26), the diffusion coefficient becomes

$$D = 10.5\exp[-3.69/(8.617\times10^{-5}eV)(1273K)]$$
$$= 2.5822\times10^{-14}cm^2\cdot s^{-1}$$

From Eq. (2.33),

$$Q = (10^{19})\sqrt{(4\times2.5822\times10^{-14}\times1800)/\pi}$$
$$= 7.6928\times10^{13}\ cm^{-2}$$

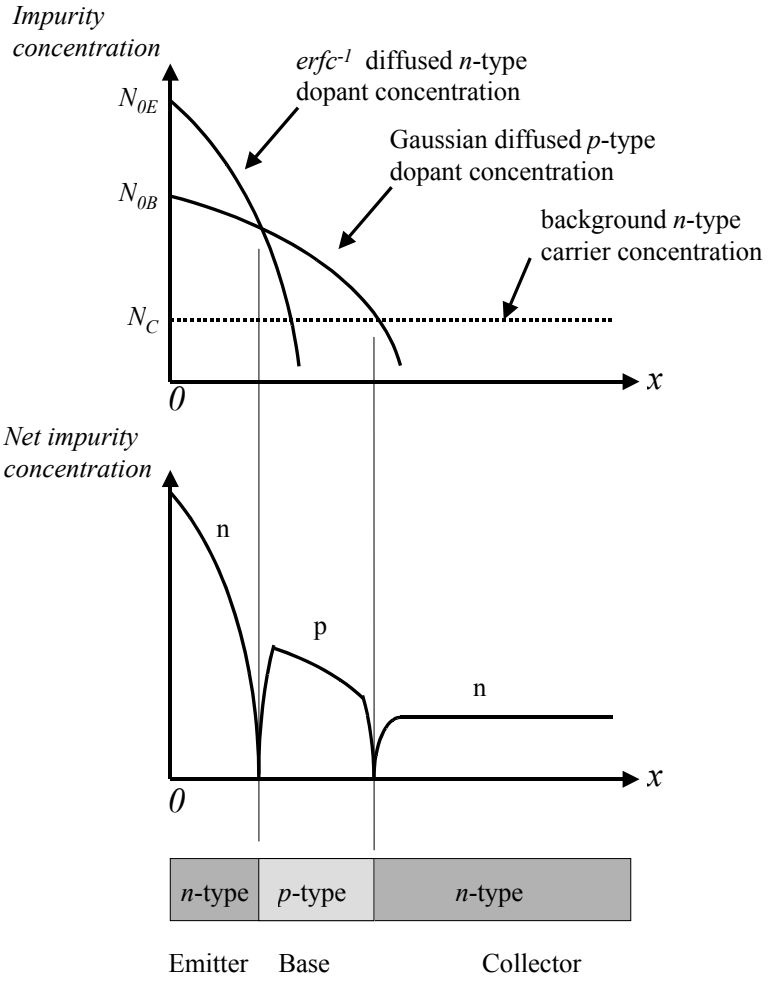

Fig. 2.16. Semi-logarithmic graphs illustrating the formation of an n-p-n transistor through diffusion. p-type dopants are first diffused into an n-type semiconductor to form the first junction. n-type dopants are subsequently diffused to form the second junction.

2.4. Ion implantation of dopants

Another technique to introduce dopants into a semiconductor wafer is through ion implantation. This technique is actually a direct alternative to the thermal predeposition described previously, and can be followed by a drive-in diffusion step.

The ion implantation process selects ions of a desired dopant, accelerates them using an electric field to form a beam of ions, and scans

them across a wafer to obtain a uniform predeposition dopant profile inside the crystal. The energy imparted to the dopant ion determines the ion implantation depth. Using this technique, a controlled dose of dopant impurities can be introduced deep inside the semiconductor. This is in contrast to diffusion, where the dose of dopant is introduced only at the wafer surface. In addition, like diffusion, it is possible to conduct the ion implantation in only certain well-defined areas of the wafer by using an appropriate mask. This method yields reproducible and controlled dopant concentration for semiconductor devices.

We can choose to perform selective implantation, in which regions are selectively implanted with accelerated ions by using a patterned layer of material such as silicon dioxide or photoresist, as shown in Fig. 2.17.

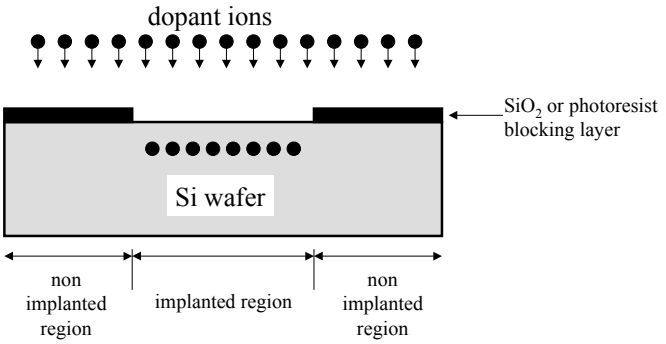

Fig. 2.17. Method of masking during ion implantation. The SiO_2 or photoresist layer acts as a blocking layer for the implantation of dopant atoms. In this process, no dopant atom can be found under the blocking layer if it is thick enough.

2.4.1. Ion generation

The first requirement of an ion implantation system is to generate ions of the desired species, accelerate them and direct them onto the wafer. A schematic of a typical ion implantation system is shown in Fig. 2.18. The dopant often comes in a gaseous form, and their ions are generated by heating them with a hot filament. These ions are then accelerated through an electric field. A magnetic field then curves the beams of ions and separates the ions, according to their atomic masses and charges, through a preset angle and output aperture. The selected ions are then further accelerated using an electric field. The beam is collimated and focused before striking the target wafer and penetrating the crystal lattice. An $x-y$ rastering mechanism ensures that a large area of the sample is scanned by the ion implantation beam.

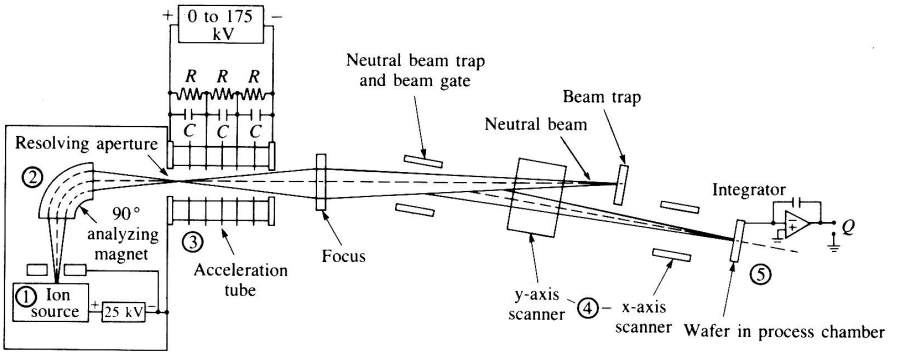

Fig. 2.18. Schematic of a typical ion implantation equipment, including an ion source (1), a primary acceleration and an analyzing magnet (2), an acceleration tube (3), a collimating and rastering magnets (4), and the sample to be implanted. [Jaeger, R.C., Introduction to Microelectronics Fabrication: Vol. 5 of Modular Series on Solidstate Devices, 2nd Edition, Fig 5.1, p. 90., © 2002. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.]

2.4.2. Parameters of ion implantation

There are four major parameters to be controlled during ion implantation: the *energy* of the ions that reach the wafer, the dose Q of the dopant (the total number of ions that reach the wafer per unit of area), and the *depth* and *width* of the resulting implanted dopant profile.

The energy of the ions is directly controlled by the voltage used to accelerate them. It is easily understood that more energetic ions will penetrate deeper into the crystal, and potentially cause more physical damage than less energetic ones.

Because the selected ions all carry the same electrical charge, by measuring the electrical current carried by the ion beam (amount of electrical charge flowing per unit time), we can directly determine the dose. Mathematically, the latter is related to the ion beam current I through:

$$\text{Eq. (2.37)} \qquad Q = \frac{I}{qA}t$$

where q is the elementary charge, A is the implanted area and t is the duration of the ion implantation. For example, a 100 μA beam current of single ionized ions swept across a 200 cm^2 area for 60 seconds yields a dose equal to:

$$\text{Eq. (2.38)} \qquad Q = \frac{\left(100\times10^{-6}\right)\times(60)}{\left(1.6\times10^{-19}\right)\times(200)} = 1.875\times10^{14} \text{ dopants per cm}^2$$

By controlling the beam current and the implantation time, values of Q between 5×10^9 and 5×10^{15} cm^{-2} can typically be achieved. This range of available doses is wider than that obtainable with thermal predeposition. It is therefore possible to reach doping profiles unobtainable by any other means. If the dopants were distributed uniformly over a depth of 50 nm, the dopant concentration could be controlled between values of 10^{15} and 10^{21} dopants per cm^3.

The depth and width of the resulting implanted doping profile can be represented by the projected range and straggle, as will be discussed in the next sub-section.

2.4.3. Ion range distribution

The dopant concentration profile after implantation follows a Gaussian distribution as illustrated in Fig. 2.19. As seen in the figure, the peak concentration N_p is found at a certain depth called the projected range R_p. The projected range measures the average penetration depth of the ions.

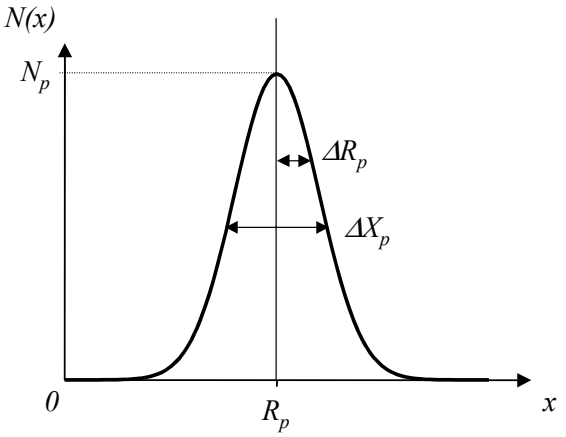

Fig. 2.19. Gaussian distribution for the concentration profile of implanted ions. The distribution is determined by its projected range, denoted R_p, corresponding to the peak concentration, and its straggle denoted by ΔR_p.

The depth at which the ions are implanted is mainly determined by the energy and the atomic number of the ions, as well as the atomic number of the substrate material. This can be easily understood because, as an impinging ion penetrates the semiconductor, it undergoes collisions with atoms and electrical repulsion with the surrounding electrons. The distance traveled between collisions and the amount of energy lost per collision are determined by a random process. Hence, even though all the ions are of the same type and have the same incident energy, they do not necessarily yield

the same implantation depth. Instead, there is a distribution of depths represented by a standard deviation, called the straggle ΔR_p. Using Fig. 2.20, the impurity concentration at a given depth x can be found if the acceleration energy E_A is known:

Eq. (2.39) $$N(x) = N_p \exp\left[-\frac{(x - R_p)^2}{2\Delta R_p^2}\right]$$

For a Gaussian distribution shown in Fig. 2.19, the full width at half-maximum, denoted ΔX_p, is given by:

Eq. (2.40) $$\Delta X_p = \left(2\sqrt{2\ln 2}\right)\Delta R_p = 2.35\Delta R_p$$

The implanted dose can be determined by integrating Eq. (2.39) over all the possible depths inside the crystal:

Eq. (2.41) $$Q = \int_0^\infty N(x)dx = \sqrt{2\pi} N_p \Delta R_p$$

Example

Q: Find the full width at half-maximum for the ion-implantation using boron with an acceleration energy of 100 keV.

A: From Fig. 2.20, we obtain the normal straggle $\Delta R_p = 0.07$ μm. Using Eq. (2.40),

$$\Delta X_p = (2\sqrt{2\ln 2})(0.07) = 0.1648\,\mu m$$

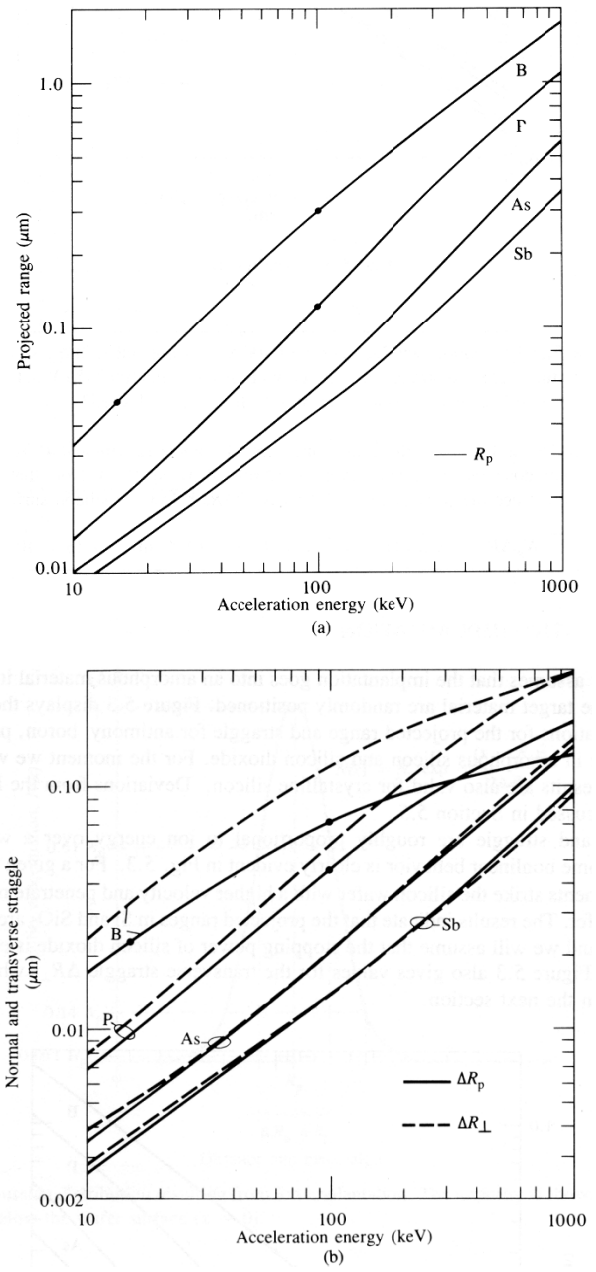

Fig. 2.20. (a) Projected range and (b) normal and transverse straggle for the ion-implantation of boron, phosphorus, arsenic, and antimony impurities for a given acceleration energy. [Jaeger, R.C., Introduction to Microelectronics Fabrication: Vol. 5 of Modular Series on Solidstate Devices, 2nd Edition, © 2002, p. 93-94, fig 5.3. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.]

2.5. Characterization of diffused and implanted layers

Two parameters are of interest in assessing the properties of diffused or implanted layers and junctions: their electrical resistivity, junction depth, and impurity concentration. A number of techniques are available for evaluating each of these parameters.

2.5.1. Sheet resistivity

In diffused or implanted layers, we are not interested in the values of the bulk resistivity, because the carrier concentration is not uniform in space as shown in the profiles in Fig. 2.13 and Fig. 2.14. Rather, we are interested in the sheet resistivity, a quantity which can be directly measured. In order to visualize the physical meaning of this parameter, let us consider a parallelepiped semiconductor bar with a length L, a width W and a thickness H as shown in Fig. 2.21.

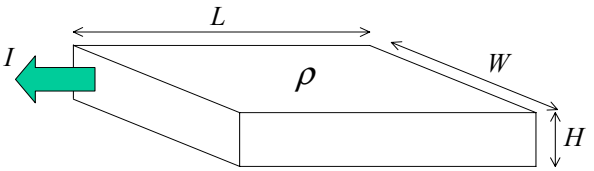

Fig. 2.21. Geometry used in determining the resistance of a block of material having uniform resistivity. When the current is flowing in the direction shown, the resistance in this direction is proportional to the length L and inversely proportional to the cross-section area WH.

From elementary physics considerations we know that the resistance of this block for a current flowing in the direction of the shown arrow is given by:

$$\text{Eq. (2.42)} \qquad R = \frac{L}{WH}\rho = \frac{L}{W}R_s$$

where ρ is the resistivity of the material, assumed to be uniform in this case. The sheet resistivity is a quantity which does not take into account the thickness of the layer, and is defined as the resistivity divided by the thickness:

$$\text{Eq. (2.43)} \qquad R_s = \frac{\rho}{H}$$

and expressed in units of "ohm per square" or Ω/□. In practice, because the thickness of the conducting layer is not always known during experiments, the sheet resistivity is the quantity that is actually measured, and the bulk resistivity is calculated subsequently.

One of the measurement methods for the sheet resistivity is the linear four-point probe method as shown in Fig. 2.22. It consists of placing four equally spaced probes on the surface of the wafer in a linear manner. The probe spacing s is typically on the order of either 1000 or 1250 μm. By sending a fixed current I through the two outer probes and measuring the voltage V across the two inner probes, we can determine the resistivity, in units of Ω·m, given by the following expression:

Eq. (2.44) $$\rho = 2\pi s \frac{V}{I} = \left[\frac{\pi t}{\ln 2}\right]\frac{V}{I}$$

Sheet resistivity can then be determined from Eq. (2.44) as:

Eq. (2.45) $$R_x = \frac{\rho}{t} = \left[\frac{\pi}{\ln 2}\right]\frac{V}{I}$$

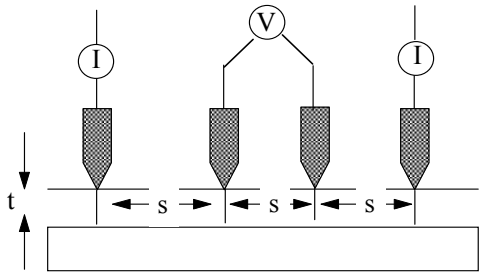

Fig. 2.22. An in-line four-point probe. Four equally spaced probes are placed on the surface of the wafer in a linear manner. A fixed current is sent through two of the outer probes and the voltage measured between the two inner probes gives a value for the sheet resistivity of the material.

Another method to measure the sheet resistivity of doped layers is van der Pauw method which can be used for any arbitrarily shaped sample of material by placing four contacts on its periphery as shown in Fig. 2.23(a). Square shaped test areas with contact regions at the four corners are usually preferred and are prepared by lithographic techniques as shown in Fig. 2.23(b).

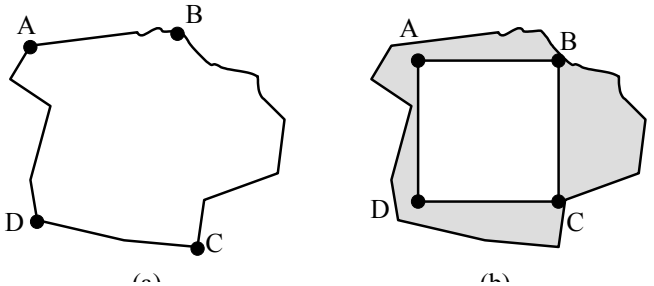

Fig. 2.23. A simple van der Pauw test structure. (a) Four contacts are placed at the periphery of any arbitrarily shaped sample and can be used to determine the resistivity of the material. (b) It is generally preferred to use a square shaped sample, which can be obtained by etching away undesired areas off the original sample.

In this configuration, a current is injected through one pair of the contacts and the voltage is measured across the remaining pair of contacts. Repeating these measurements for another pair, we are then able to define two resistances such as for example:

$$\text{Eq. (2.46)} \qquad R_{AB,CD} = \frac{V_{CD}}{I_{AB}} \text{ and } R_{BC,DA} = \frac{V_{DA}}{I_{BC}}$$

These resistances are related by the following equation relation:

$$\text{Eq. (2.47)} \qquad \exp\left[\frac{-\pi R_{AB,CD}}{R_S}\right] + \exp\left[\frac{-\pi R_{BC,DA}}{R_S}\right] = 1$$

where R_s is the sheet resistivity of the semiconductor layer. This expression allows us to implicitly determine the sheet resistivity of the sample, and thus the bulk resistivity if the layer thickness is known. For a symmetrical measurement geometry, the two resistances in Eq. (2.46) are equal and Eq. (2.47) yields a simple expression:

$$\text{Eq. (2.48)} \qquad R_s = \frac{\pi}{\ln 2}\frac{V_{CD}}{I_{AB}}$$

2.5.2. Junction depth

The junction depth x_j is defined as the distance from the top surface within the diffused or implanted layer at which the dopant concentration equals the

background concentration. There exist two methods to measure the junction depth: the groove and stain method and the angle-lap method.

In the groove and stain method, a cylindrical groove is mechanically ground into the surface of the wafer as shown in the cross-section schematic in Fig. 2.24. A chemical stain creates a color contrast between the differently doped layers, thus revealing the junction.

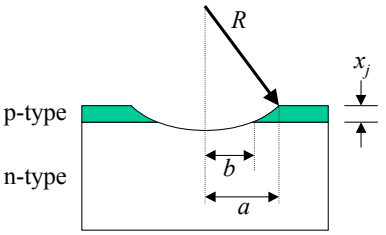

Fig. 2.24. Cross-section illustration of the junction depth measurement by the groove and stain technique. A cylindrical groove is mechanically ground into the surface of the wafer. A chemical stain creates a color contrast between the differently doped layers, thus revealing the junction.

Through purely geometrical arguments, the junction depth can be found to be equal to:

Eq. (2.49) $$x_j = \sqrt{\left(R^2 - b^2\right)} - \sqrt{\left(R^2 - a^2\right)}$$

For $R >> a$ and b, this expression can be simplified into:

Eq. (2.50) $$x_j \approx \frac{\left(a^2 - b^2\right)}{2R}$$

In the angle-lap method, a piece of the wafer is mounted on a special fixture which permits the edge of the wafer to be lapped at an angle between 1 and 5° as shown in Fig. 2.25. The sample is then stained with, for example, a 100 ml hydrofluoric acid/nitric acid solution. Once stained, the sample is observed under a collimated monochromatic light at normal incidence.

An interference pattern can be observed through a cover glass, and the junction depth may be calculated by counting the interference fringes and then applying the equation:

Eq. (2.51) $$x_j \approx d \tan\theta = \frac{N\lambda}{2}$$

where θ is the angle of the etched lap, λ is the wavelength of the monochromatic light and N is the number of fringes.

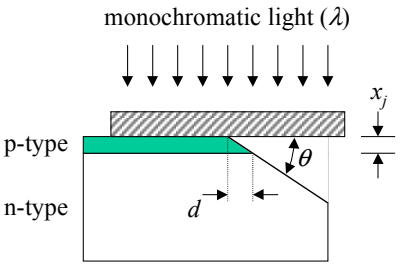

Fig. 2.25. Junction depth measured by the angle-lap method. A piece of the wafer is mounted on a special fixture which permits the edge of the wafer to be lapped at an angle. The sample is then stained and then observed under a collimated monochromatic light at normal incidence. An interference pattern can be observed through a cover glass and the junction depth can be calculated by counting the number of interference fringes.

2.5.3. Impurity concentration

There are many ways to measure the impurity concentration of a sample, but one of the most common techniques is Secondary Ion Mass Spectroscopy (SIMS). A schematic representation of the experimental setup is shown in Fig. 2.26. SIMS is a destructive characterization technique that operates with a highly energetic beam of ions hitting the sample, causing the sputtering or ejection of atoms from the sample material. Some of these ejected atoms are charged ions, or secondary ions. A mass spectrometer then separates and collects the secondary ions. The number of collected secondary ions then allows a detector to determine the material composition. An example plot of the impurity concentration information obtained from SIMS is shown in Fig. 2.27.

SIMS is an excellent technique for identifying all types of elements, unlike other measurement resources. One disadvantage to the SIMS measurement is the sensitivity of the technique. The sensitivity can be affected by the built-up charge from the sputtering process, type of ion beam. Another disadvantage to SIMS is the limitation of the beam area that hits the sample.

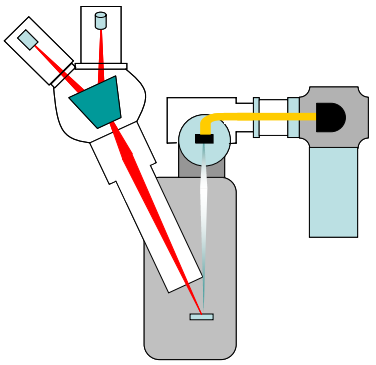

Fig. 2.26. Secondary Ion Mass Spectrometer. Low energy ion-beam hits the sample surface and atoms sputter off of it. The atoms are then ionized and sent through a mass spectrometer, where secondary ions are collected. The mass spectrometer identifies the atomic species and then the detector uses these secondary ions to determine the profile as a function of depth.

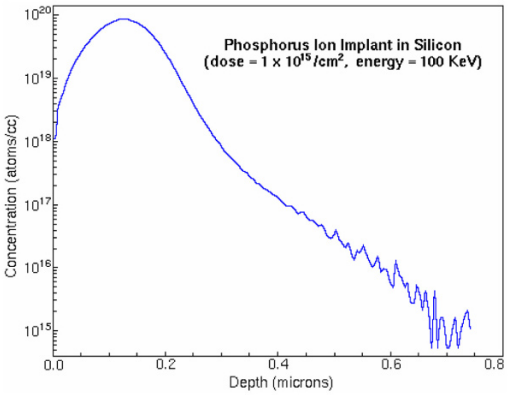

Fig. 2.27. Secondary Ion Mass Spectroscopy plot measuring the impurity concentration as a function of depth for ion implantation of phosphorus into silicon. [From http://www.me.ust.hk/~mejswu/MECH343/343SIMS.pdf.]

2.6. Summary

In this Chapter, we have reviewed a few of the steps involved in the fabrication of semiconductor devices, including oxidation, diffusion and ion implantation. Although the discussion was primarily based on silicon, the concepts introduced are applicable for the entire semiconductor industry.
We described the oxidation experimental process, mathematically modeled the formation of a silicon oxide film, discussed the factors influencing the oxidation and reviewed the methods used to characterize the oxide film. The diffusion and ion implantation of impurity dopants in silicon to achieve controlled doping in selected areas of a wafer was described, along with the

resulting dopant concentration profiles inside the semiconductor. The predeposition and drive-in conditions of diffusion were discussed. Methods used to assess the electrical properties of the diffused or implanted layers were reviewed.

References

Jaeger, R.C., *Introduction to Microelectronic Fabrication*, 2nd Edition, Prentice-Hall, Upper Saddle River, NJ, 2002.

Razeghi, M., *The MOCVD Challenge Volume 1: A Survey of GaInAsP-InP for Photonic and Electronic Applications*, Adam Hilger, Bristol, UK, pp. 188-193, 1989.

Razeghi, M., *The MOCVD Challenge Volume 2: A Survey of GaInAsP-GaAs for photonic and electronic device applications*, Institute of Physics, Bristol, UK, 1995.

Further reading

Campbell, S.A., *The Science and Engineering of Microelectronic Fabrication*, Oxford University Press, New York, 1996.

Diaz, J.E., *Fabrication of High Power Aluminum-Free 0.8 μm to 1.0 μm InGaP/InGaAsP/GaAs Lasers for Optical Pumping*, Ph.D. dissertation, Northwestern University, 1997.

Fogiel, M., *Microelectronics-Principle, Design Techniques, and Fabrication Processes*, Research and Education Association, New York, 1968.

Ghandi, S., *VLSI Fabrication Principles*, John Wiley & Sons, New York, 1983.

Soclof, S., *Design and Application of Analog Integrated Circuits*, Prentice-Hall, Englewood Cliffs, NJ, pp. 8-23, 1991.

Streetman, B.G., *Solid State Electronic Devices*, Prentice-Hall, Englewood Cliffs, NJ, pp. 65-70, 1980.

Trumbore, F.A., "Solid solubilities of impurity elements in germanium and silicon," *Bell System Technical Journal* 39, pp. 205-233, 1960.

Problems

1. A (100) Si wafer undergoes the following sequence of oxidation steps: one-hour dry oxidation at 1100 °C, two-hour wet oxidation at 1000 °C, and one-hour dry oxidation at 1100 °C. Calculate the thickness after each oxidation step.

2. Compare the thickness of silicon dioxide film grown on (100) Si for wet and dry oxidation at 1100 °C. Compare the two different orientations (100) and (111) Si using the same conditions for wet oxidation.

3. Calculate the time required to grow 2 μm silicon dioxide on (111) silicon wafer for wet oxidation at 1050 °C. How long would it take to grow an additional 5 μm?

4. A silicon oxide layer is grown for 65 minutes at 1100 °C on (111) silicon by passing oxygen through a 95 °C water bath. How thick an oxide layer is grown on the silicon surface?

5. A (100) oriented silicon wafer is already covered with a 0.5 μm thick oxide film. How long would it take to grow an additional 0.1 μm oxide using wet oxidation at 1373 K? Compare the result with the linear oxidation law using a rate of $B/A = 3\ \mu m \cdot hr^{-1}$.

6. An *npn* transistor is formed by boron diffusion on an *n*-type silicon wafer with impurity concentration of $10^{20}\ cm^{-3}$ and doping concentration $10^{16}\ cm^{-3}$. Constant source diffusion is performed for 30 minutes followed by limited-source diffusion for 2 hours, both at 1000 °C. Find the junction depth after each step.

7. What is the required thickness for a SiO_2 mask used for selective phosphorus diffusion. The diffusion was performed at 1000 °C for 3 hours.

8. An impurity is diffused into silicon in the constant-source diffusion case with a surface concentration $N_0 = 10^{19}\ cm^{-3}$. The diffusion coefficient is known to be equal to $2 \times 10^{-12}\ cm^2 \cdot s^{-1}$ at 1100 °C with an activation energy of 4 eV. A diffusion length of 1 μm is aimed at. (a) After diffusion at 1000 °C, what is the total dose diffused in the layer? (b) How long must the diffusion last?

9. An impurity is diffused into silicon at 1100 °C during 20 minutes. A diffusion length of 1 μm is measured and the dose diffused in the sample is $2 \times 10^{15}\ atoms \cdot cm^{-2}$. Determine the diffusion coefficient. Determine the surface concentration, assuming constant-source diffusion.

10. Find the dose in silicon for phosphorus that is implanted with an energy of 100 keV with a 0.1 μm SiO_2 layer and peak concentration of

10^{17} cm^{-3}. Find the time required for this implantation onto a 2" wafer using 2 μA of current.

11. Implantation using phosphorus is done such that the implantation peak is located at the Si-SiO_2 interface with an energy of 40 keV. The dose is 5.24×10^{14} cm^{-2}, and background concentration is 3×10^{16} cm^{-3}. Find the minimum oxide thickness required for a masking layer.

12. Compare the time required for implantation of phosphorus and boron with an energy of 75 keV, with a desired peak concentration of 10^{18} cm^{-3}, into a 140 mm silicon wafer with 1 μA.

13. A constant voltage of 5 V is applied to each of the five contact pads on a given sample. The location of the pads are at $x_1 = 1$, $x_2 = 5$, $x_3 = 10$, $x_4 = 16$, and $x_5 = 23$ μm. The width is the same for each contact and is given as 1 μm. The thickness is given as 200 μm. Find the contact resistance using transmission line measurement if the current is measured as $I_{12} = 20$ mA, $I_{23} = 17$ mA, $I_{34} = 10$ mA, and $I_{45} = 3$ mA.

14. Calculate the resistivity using the van der Pauw method with measured current $I_{AB} = 1$ mA and $V_{CD} = 2$ V. The length $L = 1$ mm and width $W = 2$ mm.

15. Assume that an As implant leads to a uniform electron concentration of 10^{19} cm^{-3} down to a depth of 0.1 μm and a mobility of 100 cm^2·V^{-1}·s^{-1}. Determine the resistivity and the sheet resistivity of the implanted layer. If a square van der Pauw pattern with 1 cm side length is used with a 10 V applied at two adjacent contacts, what current would be measured through the other two contacts?

http://www.springer.com/978-1-4419-1055-4

Technology of Quantum Devices

Razeghi, M.

2010, XXIII, 546 p. 100 illus., Hardcover
ISBN: 978-1-4419-1055-4