

## DIFFUSION

Diffusion is a technique for introducing controlled amounts of chemical impurities into the silicon lattice. It describes the process by which atoms move in the crystal lattice and includes self-diffusion as well as the diffusion of impurity atoms that are introduced in order to alter the electrical properties of the material. The basic driving force is the concentration gradient  $dC/dx$ . At elevated temperatures (typically 800°C - 1200°C for silicon) the dopant atoms move away from a high concentration region to a low concentration region. Lattice features such as crystal structure and defect concentration also play an important role.

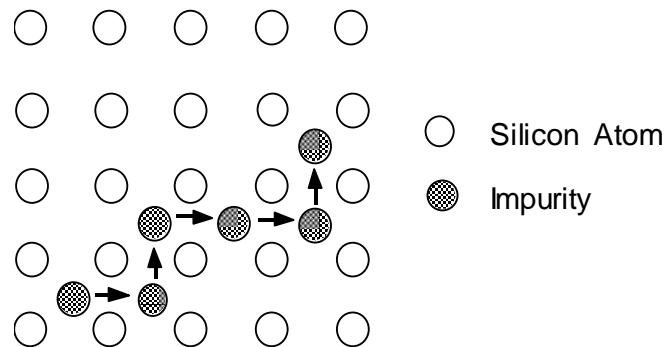
### 1.1 Diffusion Theory:

There are two major approaches for tackling diffusion problems:

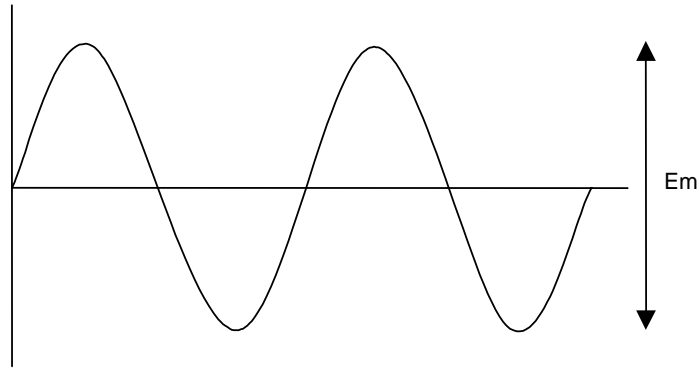
**1.1.1 Continuum theory** of Fick's diffusion equation. This approach is suitable when impurity concentrations are low and the diffusivity is constant (i.e. not concentration dependent).

**1.1.2 Atomistic theory** This involves interactions between point defects, vacancies and interstitial atoms, and impurity atoms. This theory is still under active development.

**1.2 Interstitial Diffusion** Here impurity atoms move through the crystal by jumping from one interstitial site to the next. There are five interstitial voids in a unit cell of the diamond lattice and these are nearly all available to accept an impurity atom as it wanders through the lattice.



For diffusion of an interstitial atom, energy is required to enable the atom to squeeze through the constrictions in the lattice. This may be considered to be periodic in nature.



$E_m$  = interaction energy (eV)

$T$  = lattice temperature ( $^{\circ}\text{K}$ )

$\nu_0$  = frequency of lattice vibrations ( $10^{13} - 10^{14}\text{Sec}^{-1}$ ); this is the frequency with which atoms strike the potential barrier shown above.

$\nu$  = frequency with which barrier is overcome

Assuming a Boltzmann energy distribution, the probability that an atom has an energy in excess of  $E_m$  is given by

$$\exp \left[ - \frac{E_m}{kT} \right]$$

Since an atom can jump in four different directions.

$$\nu = 4\nu_0 \exp \left[ - \frac{E_m}{kT} \right]$$

Ex 1 Calculate the rate at which an interstitial impurity atom will jump from one void to another

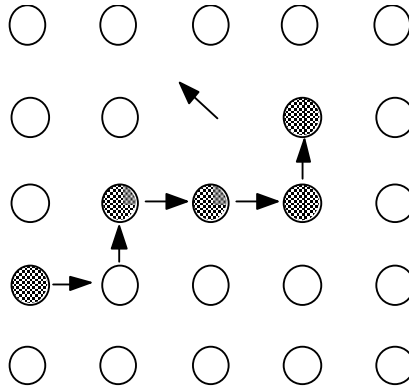
(a) at room temperature

(b) at the diffusion temperature of  $1,000^{\circ}\text{C}$

Assume a value of 1.0eV for  $E_m$  and a lattice vibration frequency of  $10^{14} \text{Sec}^{-1}$ .

### 1.3 Substitutional diffusion

Before this type of diffusion can occur a host atom must acquire sufficient energy to leave the lattice site and become an interstitial atom, thereby creating a vacancy. A neighbouring impurity atom can then migrate to the vacant site.



SUBSTITUTIONAL DIFFUSION

The probability of a neighbouring site being vacant is given by

$$\exp \left[ - \frac{E_s}{kT} \right]$$

where  $E_s$  is energy of formation of a vacancy (Schottky defect). Each lattice site has four tetrahedrally situated nearest neighbours so each jump can be made in four different ways. The substitutionally diffusing atom will also experience a periodic potential as certain covalent bonds are broken and others are made. If the height of this barrier is  $E_n$ , then the probability of one of these atoms having an energy in excess of this is

$$\exp \left[ - \frac{E_n}{kT} \right]$$

and so

$$\begin{aligned} v &= 4v_0 \exp \left[ - \frac{E_s}{kT} \right] \exp \left[ - \frac{E_n}{kT} \right] \\ &= 4v_0 \exp \left[ - \frac{(E_s + E_n)}{kT} \right] \end{aligned}$$

Ex 2 Calculate, as in Ex 1, the rate at which a substitutional impurity jumps from one lattice site to another. Assume  $E_S = 2.3\text{eV}$  and  $E_I = 1.5\text{eV}$ .

### **Schottky Defect**

This occurs when an atom leaves its site and moves to the surface. The density of Schottky defects is calculated from:

$$n = N \exp \left( -\frac{E_S}{kT} \right)$$

where  $N$ : atom density of Si

$E_S$ : energy to create vacancy

Ex.3 Calculate the density of Schottky defects at room temperature and at  $1000^\circ\text{C}$

## 1.4 Diffusion Equation

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}$$

where D is the diffusion coefficient or diffusivity. D has dimensions of  $L^2T^{-1}$  and the units commonly used are  $\mu m^2h^{-1}$ . Diffusion coefficients for the important impurities in silicon are strong functions of temperature i.e.

$$D = D_0 \exp \left[ - \frac{E_A}{kT} \right]$$

where  $D_0$  is the asymptotic value of D at infinite temperature, k is Boltzmann's constant and  $E_A$  is the activation energy for the process (see Fig.2 of the process curves). In silicon, n and p-type impurities are substitutional and enough energy (heat) must be supplied to break the bonds and create vacancies (3-4 times  $E_A$ ). It is possible, by selecting suitable boundary conditions and by using complex derivations to obtain solutions to the diffusion equation (see Ghandhi). In IC fabrication there are two situations of interest, namely diffusion from a **Constant** source and diffusion from a **Limited** source.

## 1.5 Diffusion from constant source

In this case the impurity source is present during the entire diffusion thereby maintaining a constant surface concentration  $C_s$ .

BOUNDARY CONDITIONS:  $C_{(\infty,t)} = 0$ ;  $C_{(x,t=0)} = 0$   
surface concentration  $C_s = C_{(0,t)}$

Solution:

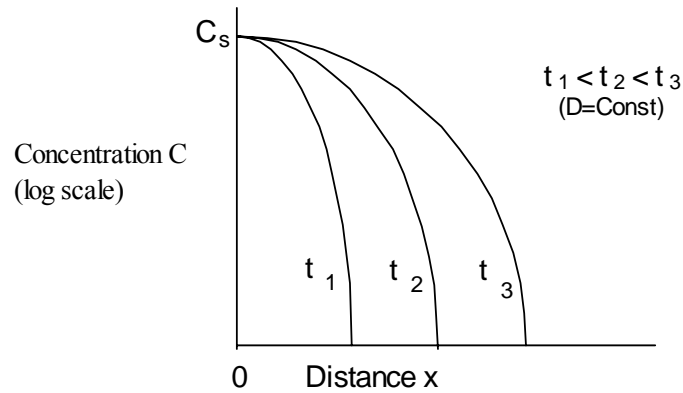
$$C_{(x,t)} = C_s \operatorname{erfc} \left[ \frac{x}{2\sqrt{D_1 t_1}} \right] \dots \dots (1)$$

where  $\operatorname{erfc}$  = complementary error function

x = distance into the silicon

D = diffusion coefficient

t = time



### CONSTANT-SOURCE DIFFUSION PROFILE

Note: surface concentration remains constant whereas diffusion depth increases with time. This type of diffusion is used when a shallow layer with high surface concentration is required, e.g. an emitter diffusion. It is also used as the first step in the two-stage limited-source diffusion where it is referred to as **Predeposition**.

A useful quantity is the total number of impurity atoms  $Q$  deposited in a  $1 \text{ cm}^2$  section of the silicon. This may be found by integrating eqn (1) i.e.

$$\int_0^{\infty} C_{(x,t)} dx = \int_0^{\infty} C_s \operatorname{erfc} \left[ \frac{x}{2\sqrt{D_1 t_1}} \right] dx$$

Note: we wish to integrate between 0 and  $x_j$ . However, as the number of impurity atoms  $\rightarrow 0$  as  $x \rightarrow x_j$  we may integrate between 0 and  $\infty$ .

$$Q(t) = C_s \int_0^{\infty} \operatorname{erfc} \frac{x}{2\sqrt{D_1 t_1}} dx$$

$$Q(t) = \frac{2}{\sqrt{\pi}} \sqrt{D_1 t_1} \times 10^{-4} \cdot C_s \quad \text{-----(2)}$$

### 1.6 Diffusion from a limited source

Typically diffused layers are formed in a two-step process. First, impurities are introduced into the silicon to a depth of a few tenths of a micron (predeposition). Later, the impurities are diffused deeper to provide a suitable concentration distribution without any

further impurities being added. This second step is called **Drive-in** and is often carried out in an oxidising ambient.

Boundary conditions:

$$\left. \frac{\partial C}{\partial x} \right|_{(0,t)} = 0, C(\infty, t) = 0$$

These provide for a constant Q being maintained in the silicon during drive-in. The initial condition is

$$C_{(x,0)} = C_s \operatorname{erfc} \left[ \frac{x}{2\sqrt{D_1 t_1}} \right]$$

The solution of the diffusion equation subject to these boundary conditions is a Gaussian profile given by:

$$C_{(x,t)} = \frac{Q}{\sqrt{\pi D_2 t_2} \times 10^{-4}} \exp - \left( \frac{x^2}{4D_2 t_2} \right) \quad \text{-----}(3)$$

Note: This assumes that

$$\sqrt{D_1 t_1} \ll \sqrt{D_2 t_2}$$

and hence the extent of penetration of the predeposited profile is small. As a result, the predeposited profile may be represented mathematically by a delta function. A comparison of the exact solution with the distribution obtained using the delta-function approximation indicates good agreement for

$$\sqrt{D_2 t_2} > 4\sqrt{D_1 t_1}$$

Errors become noticeable when

$$\sqrt{D_2 t_2} \approx \sqrt{D_1 t_1}$$

Since a constant Q is maintained during drive-in, the surface concentration must drop as the impurity spreads deeper into the semiconductor.

The new surface concentration is given by

$$C_{(0,t)} = C_s(t) = \frac{Q}{\sqrt{\pi D_2 t_2} \times 10^{-4}} \text{-----(4)}$$

### 1.7 Successive Diffusions

In a typical IC process there are multiple drive-in steps during oxidations and other diffusions. For example, when diffusing the emitter in a bipolar transistor the base is subjected to further drive-in. Thus the impurities in the base are subjected to two sets of time and temperature values. The effective Dt product is given by:

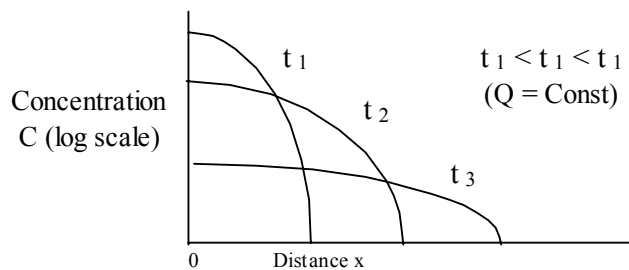
$$(Dt)_{\text{eff}} = \Sigma (D_1 t_1 + D_2 t_2 + \dots D_n t_n)$$

where  $t_1, t_2$  are the different diffusion times.

$D_1, D_2$  are the appropriate diffusion constants in effect at these times.

The impurity profile is given by:

$$C_{(x,t)} = \frac{Q}{\sqrt{\pi D_t t_t} \times 10^{-4}} \exp\left(-\frac{x^2}{4D_t t_t}\right)$$

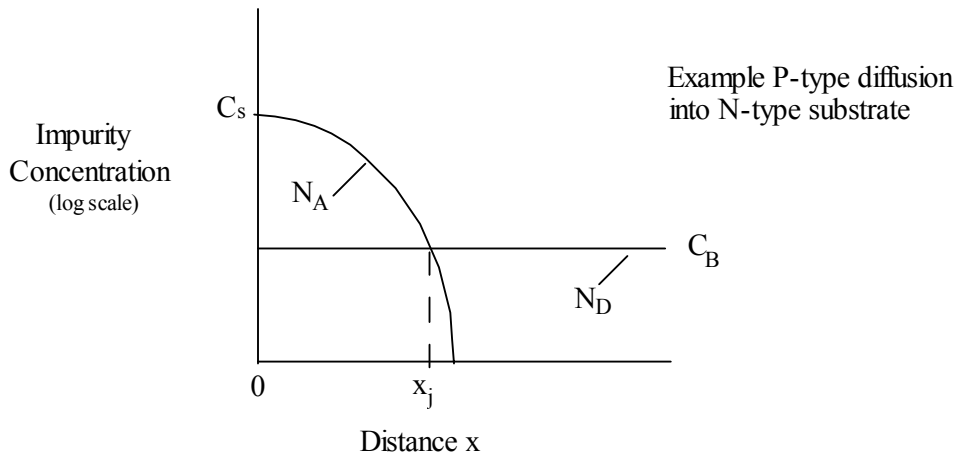


LIMITED-SOURCE DIFFUSION PROFILE

### 1.8 Junction formation

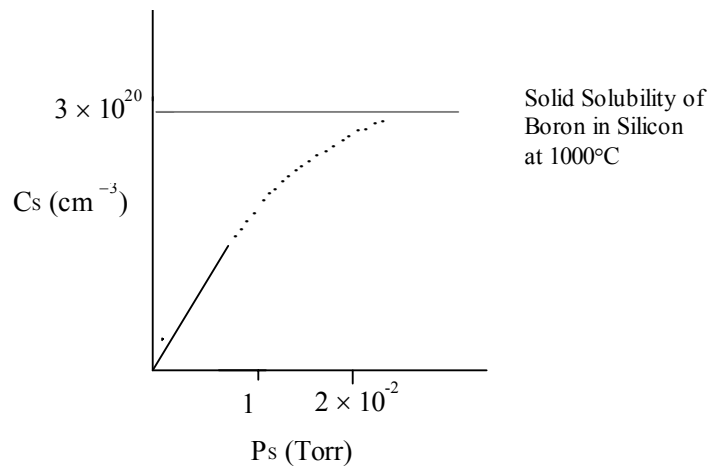
When an impurity of opposite type to the substrate is diffused, a junction is formed. This occurs at  $x = x_j$  when  $C_{(x_j,t)} = C_B$  (where  $C_B$  is the substrate (background) concentration). The net concentration is then given by the difference in acceptor and donor impurities, the conductivity type being determined by whichever impurity is in the majority.





### 1.9 Solid Solubility:

Henry's Law states that in equilibrium, the concentration of a species within a solid is proportional to the partial pressure of that species in the surrounding gas. There is however, a limit, i.e. a maximum concentration of an impurity that may be incorporated in a solid at any given temperature. This concentration is called the **Solid Solubility** of the impurity. Thus, Henry's Law holds only up to the limit of solubility.



Ex: Surface concentration of boron in silicon as a function of the partial pressure of  $B_2O_3$  in the ambient at 1100°C.

In industrial practice, predeposition is usually performed with a partial pressure of the impurity gas high enough so that the surface concentration in the silicon corresponds to the solid solubility of the impurity. Under such conditions the surface concentration is independent of the partial pressure of the impurity and will be both reproducible and uniform even if the partial pressure is not.

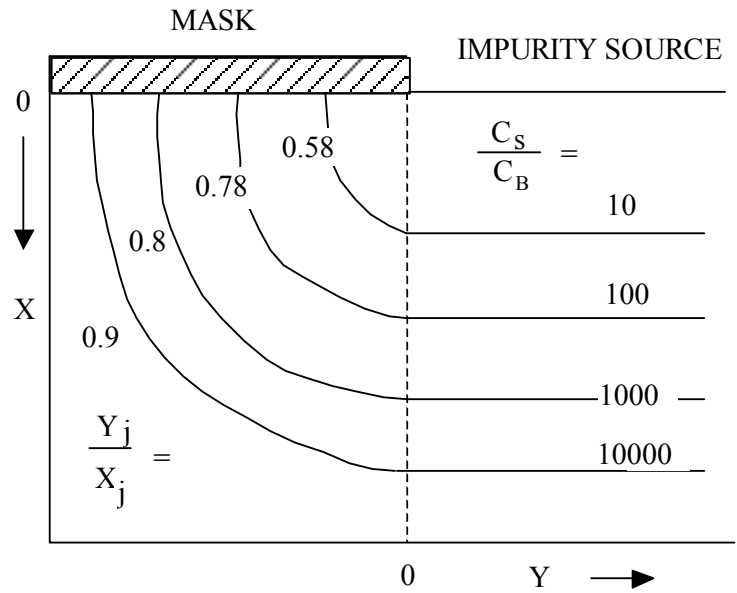
Ex 4 A P<sup>+</sup>N junction is made by diffusing boron into an n-type substrate with background concentration  $C_B = 10^{16} \text{cm}^{-3}$ . A constant source concentration is maintained during the diffusion. Calculate the time required to form the junction at a depth of 1  $\mu\text{m}$  if the diffusion temperature is 1050°C. How long would it take to form the junction if the temperature is reduced to 950°C?

Ex. 5 A boron predeposition lasting 30 min is performed at 950°C on a wafer having a background concentration  $C_B = 5 \times 10^{15} \text{cm}^{-3}$ . Calculate the junction depth  $x_j$ . How many impurity atoms/cm<sup>2</sup> have been deposited? The wafer is then subjected to a drive-in lasting 2 hr at 1150°C. Calculate the junction depth. What is the new surface concentration?

Ex. 6 A p-type wafer of background concentration  $10^{16}\text{cm}^{-3}$  is diffused with phosphorous at a temperature of  $1,000^{\circ}\text{C}$ . Predeposition lasts 15 min and drive-in lasts 1 hr. Plot the resultant impurity concentration profile and comment on the validity of any assumptions made.

Ex. 7 Determine the diffusivity of phosphorus at 1200°C if the diffusivity at 1050°C and 1150°C is  $0.04 \mu\text{m}^2/\text{hr}$  and  $0.64 \mu\text{m}^2/\text{hr}$  respectively.

**1.10 Lateral Diffusion:** The one-dimensional diffusion equation satisfactorily describes the diffusion process except at the edge of the mask window. Here the impurities will diffuse downward and sideways (laterally). The sketch below shows the contours for a Gaussian diffusion. From this it can be seen that when the background concentration is a factor of 10 below the surface concentration, sideways diffusion is roughly 60% of the vertical diffusion. This increases to about 90% when the background concentration is a factor of  $10^{-4}$  below the surface concentration. The reduced lateral diffusion is due to a limited supply of the diffusion species at the edge of the mask window.



Lateral v Vertical Diffusion under an oxide mask - Gaussian

Ex. 8 A wafer with  $C_B = 5 \times 10^{16} \text{ cm}^{-3}$  is diffused with a  $C_S = 5 \times 10^{19} \text{ cm}^{-3}$  resulting in a junction depth of  $1 \text{ } \mu\text{m}$ . Determine the lateral impurity penetration beneath the oxide masking layer.

### 1.11 Choice of Impurity:

The table below lists the possible choice of impurity with solid solubilities and temperatures at which these are attainable.

	n-type				p-type		
	P	As	Sb	B	Al	Ga	In
SS cm <sup>-3</sup>	1.3x10 <sup>21</sup>	2x10 <sup>21</sup>	6x10 <sup>19</sup>	2.5x10 <sup>20</sup>	10 <sup>19</sup> -10 <sup>20</sup>	4x10 <sup>19</sup>	10 <sup>19</sup>
T°C	1150	1150	1300	1200	1150	1250	1300

**n-type:** None of the n-type impurities listed exhibits undesirable characteristics. All are highly soluble in silicon and may be used with oxide masking.

**Phosphorus:** High limit of solid solubility (SS), low diffusivity (D) in SiO<sub>2</sub>. D is ~ 10 times higher than for As or Sb hence it is often used for reasons of economy.

**Arsenic:** High limit of SS, low D in Si which makes it suitable for such applications as the buried layer in bipolar circuits. Low D in SiO<sub>2</sub>. Good lattice match. Normally used as a spin-on dopant.

**Antimony:** Similar to A<sub>S</sub> but with a lower SS.

**p-type :** Of the p-type impurities listed, only one, boron is suitable for the following reasons.

## 1.12 Source of Diffusion

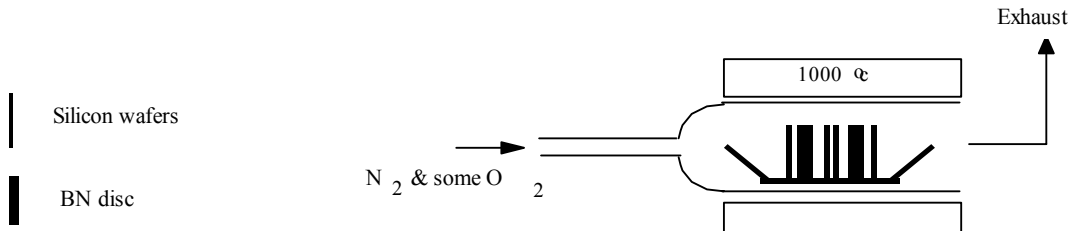
Diffusions may be carried out using either solid, liquid or gas sources.

### 1.12.1 Solid source:

Nowadays diffusion from a solid source is carried out by means of solid discs. In the case of boron these take the form of boron nitride (BN) discs and in the case of phosphorus they take the form of silicon pyrophosphate SiP<sub>2</sub>O<sub>7</sub> discs. The boron nitride discs must first be heated to ~1000°C in O<sub>2</sub> to form a surface layer of boron oxide. Once this surface layer becomes depleted it becomes necessary to re-oxidise them. Recently boron<sup>+</sup> discs have become available and these require no oxidising treatment.



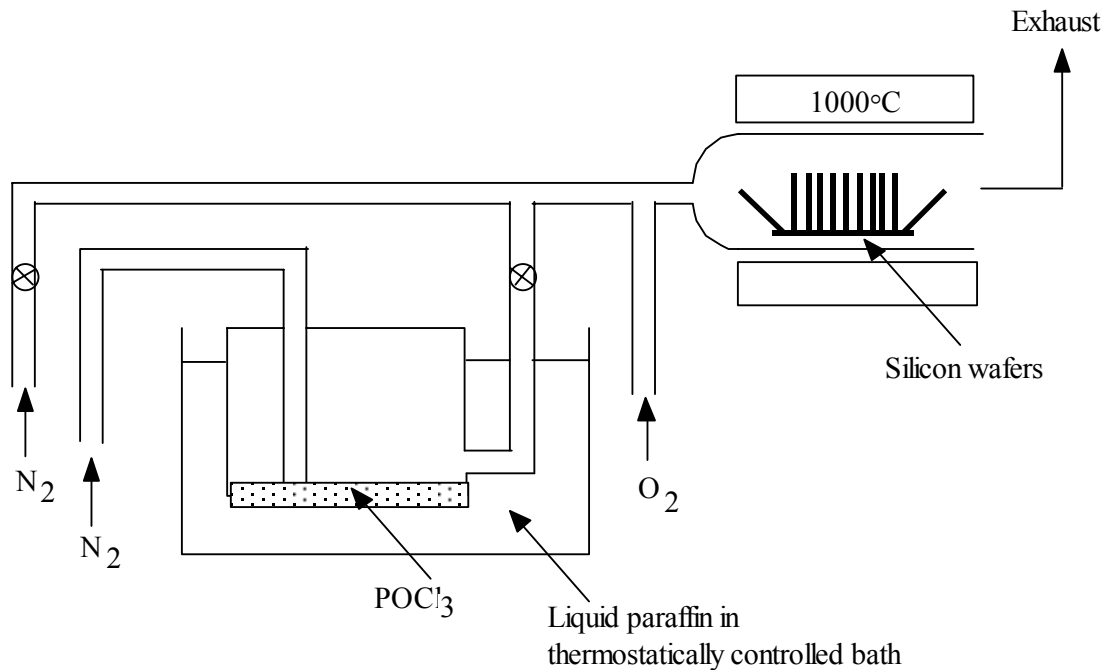
To carry out a diffusion the discs are placed parallel with the silicon wafers about 2 mm apart and at right angles to the carrier gas flow. Using this set-up good yield and reproducibility are obtained. Each wafer has its own diffusion source so gas-flow patterns are less critical leading to better uniformity across the wafer and good reproducibility. The need for connection of liquid or gaseous sources is obviated.



### 1.12.2 Liquid sources:

**i) Phosphorus** Here the source is phosphorus oxychloride,  $\text{POCl}_3$  which is maintained near room temperature and the carrier gas is bubbled through it to transport the dopant to the furnace tube.  $\text{O}_2$  is present in the carrier gas to protect the silicon surface and to facilitate the decomposition of  $\text{POCl}_3$  into  $\text{P}_2\text{O}_5$  which becomes the local source.

**ii) Boron** Here the process is similar and boron tribromide,  $\text{BBr}_3$  is used as the source. It differs in that the  $\text{N}_2$  is not bubbled through, but simply passed over the  $\text{BBr}_3$ .  $\text{O}_2$  is included to protect the Si and assist in forming  $\text{B}_2\text{O}_3$ . Diffusion without  $\text{O}_2$  can result in insoluble block deposits.



### Spin-on dopant

Alternatively a doped glass, which comes in liquid form, may be spun onto the wafer using a standard photoresist spinner. After a brief curing period at low temperature the wafer is loaded into the diffusion furnace. Two disadvantages of this method are:

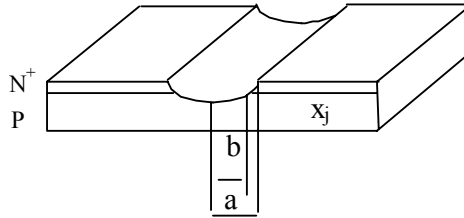
- 1) the technique is not attractive for mass production purposes (dopant must be spun-on wafers individually).
- 2) for limited-source diffusions the wafer must be removed from the furnace and the doped glass dissolved prior to drive-in (this also applies to solid sources). The advantages are similar to those mentioned for solid sources.

**1.12.3 Gaseous source:** Using this system metered quantities of the required gas are simply passed down the furnace tube. The gases commonly used are arsine  $\text{AsH}_3$ , phosphine  $\text{PH}_3$  and diborane  $\text{B}_2\text{H}_6$ . These are highly toxic gases which must be handled with extreme care.

## 1.13 Evaluation of Diffused Layers

**1.13.1 Junction depth:** This is easily determined if the diffused layer forms a PN junction with the underlying substrate. The junction may be delineated by cutting a groove in the silicon and etching the surface with a solution that stains the P-type region darker than the n-type. Beveling magnifies the depth dimensions by purely geometric means. A microscope

is then used to estimate  $x_J$ . Alternatively an interferometer may be used for more accurate results.



$$x_J = \sqrt{R_0^2 - b^2} - \sqrt{R_0^2 - a^2}$$

where  $R_0$  is the radius of the tool used to form the groove.

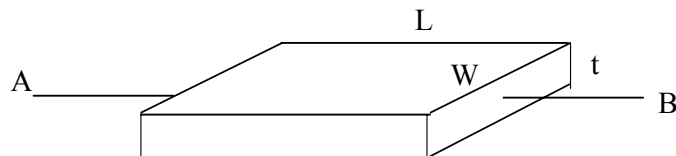
If  $R_0 \gg a$  and  $b$  then

$$x_J \approx \frac{a^2 - b^2}{2R_0}$$

### 1.13.2 Sheet Resistance:

For a diffused layer it is more appropriate to specify a sheet resistance rather than a specific resistivity as the layer is not homogeneously doped.

Consider the section of silicon shown



$$R_{AB} = \frac{\rho L}{tW} \Omega$$

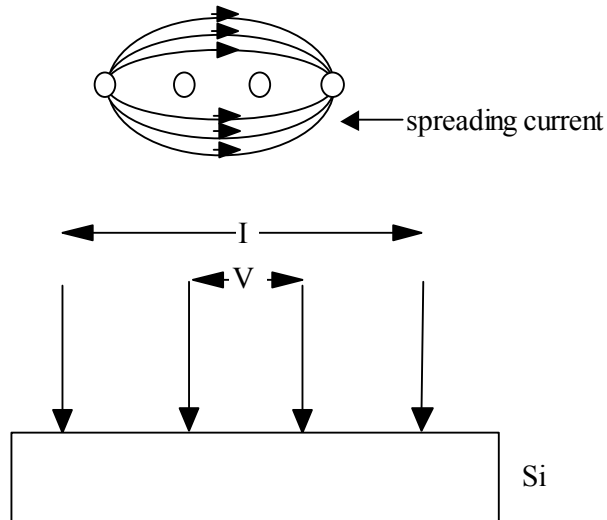
where  $\rho$  is the mean resistivity of the layer in  $\Omega \text{ cm}$

The ratio  $\rho/t$  is called the Sheet Resistance  $R_s$  and the units are  $\Omega$ , or more commonly  $\Omega/\text{sq}$  as this specifies the resistance of one square of the material. The total resistance can be found by multiplying the sheet resistance by the number of squares, i.e.  $L/W$

$$\text{So } R_{AB} = R_s \frac{L}{W}$$

If  $L = W$  then  $R = R_s$

Sheet resistance is conventionally measured using the four-point probe technique. Four equi-spaced probes are placed on the surface of the silicon. Current is passed through the two outermost probes and the potential drop is measured across the two inner probes. This arrangement eliminates any problems caused by probe-to-semiconductor resistance. Assuming the voltage is measured by a high-impedance voltmeter the current drawn will be negligible and hence the voltage drop across the voltage probes will also be negligible. The voltage drop across the outer probe contacts is unimportant since it is the current that is being measured.



The electrostatic problem, which takes into account the spreading of the current from the two outside probes, has been solved to give the solution:

$$R_s = \frac{\rho}{t} = \left[ \frac{\pi}{\ln 2} \right] \cdot \frac{V}{I} = 4.53 \frac{V}{I}$$

where  $t$  is the thickness of the sample. If a diffused layer is being measured then  $t$  is replaced by  $x_j$  the junction depth. Note: this formula only applies to layers whose lateral dimensions are large and whose width  $x_j$  is small in comparison to the probe spacing. Otherwise a correction factor must be introduced.

Revised March 2004.