DIFFUSION - Chapter 7

- Doping profiles determine many short-channel characteristics in MOS devices.
- Resistance impacts drive current.
- Scaling implies all lateral and vertical dimensions scale by the same factor.
- Generally doping levels need to increase and $x_J$ values need to decrease.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology Node (half pitch)</td>
<td>250 nm</td>
<td>180 nm</td>
<td>130 nm</td>
<td>90 nm</td>
<td>65 nm</td>
<td>45 nm</td>
<td>32 nm</td>
<td>22 nm</td>
<td>18 nm</td>
</tr>
<tr>
<td>MPU Printed Gate Length</td>
<td>100 nm</td>
<td>70 nm</td>
<td>53 nm</td>
<td>35 nm</td>
<td>25 nm</td>
<td>18 nm</td>
<td>13 nm</td>
<td>10 nm</td>
<td></td>
</tr>
<tr>
<td>Contact Resistivity $r_C$ ($\Omega$ cm$^2$)</td>
<td>2x10$^{-7}$</td>
<td>1.6x10$^{-7}$</td>
<td>1x10$^{-7}$</td>
<td>6x10$^{-8}$</td>
<td>1.7x10$^{-8}$</td>
<td>8.7x10$^{-9}$</td>
<td>5x10$^{-9}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/D Extension Abruptness (nm/decade)</td>
<td>4.1</td>
<td>2.8</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/D Extension Sheet Resistance (PMOS) ($\Omega$/sq)</td>
<td>660</td>
<td>890</td>
<td>1875</td>
<td>510</td>
<td>550</td>
<td>580</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S/D Extension $x_J$ (nm)</td>
<td>50-100</td>
<td>42-70</td>
<td>25-43</td>
<td>20</td>
<td>14</td>
<td>7.2</td>
<td>10</td>
<td>7.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Min Supply Voltage (volts)</td>
<td>1.8-2.5</td>
<td>1.5-1.8</td>
<td>1.2-1.5</td>
<td>0.9-1.2</td>
<td>0.8-1.1</td>
<td>0.7-1.0</td>
<td>0.6-0.9</td>
<td>0.5-0.8</td>
<td>0.5-0.7</td>
</tr>
</tbody>
</table>
• The resistivity of a cube is given by

\[ J = n q v = n q \mu \varepsilon = \frac{1}{\rho} \varepsilon \quad \therefore \quad \rho = \frac{\varepsilon}{J} \quad \text{Ωcm} \quad (1) \]

• The sheet resistance of a shallow junction is

\[ R = \frac{\rho}{x_j} \quad \text{Ω / Square} \equiv \rho_s \quad (2) \]

• For a non-uniformly doped layer,

\[ \rho_s = \frac{\rho}{x_j} = \frac{1}{q \int_0^{x_j} [n(x) - N_B] \mu[n(x)] \, dx} \quad (3) \]

• Eqn. (3) has been numerically integrated by Irvin for simple analytical profiles (example later).

• Sheet resistance can be experimentally measured by a four point probe technique.

• Doping profiles can be measured by SIMS (chemical) or spreading R (electrical).
• Diffusion is the redistribution of atoms from regions of high concentration of mobile species to regions of low concentration. It occurs at all temperatures, but the diffusivity has an exponential dependence on $T$.

• **Predeposition**: doping often proceeds by an initial predep step to introduce the required dose of dopant into the substrate.

• **Drive-In**: a subsequent drive-in anneal then redistributes the dopant giving the required $x_J$ and surface concentration.

---

### Ion Implantation and Annealing vs. Solid/Gas Phase Diffusion

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room temperature mask</td>
<td>Implant damage enhances diffusion</td>
</tr>
<tr>
<td>Precise dose control</td>
<td>Dislocations caused by damage may cause</td>
</tr>
<tr>
<td></td>
<td>junction leakage</td>
</tr>
<tr>
<td>$10^{11} - 10^{16}$ atoms cm$^{-2}$ doses</td>
<td>Implant channeling may affect profile</td>
</tr>
<tr>
<td>Accurate depth control</td>
<td>Low surface concentration hard to achieve</td>
</tr>
<tr>
<td></td>
<td>without a long drive-in</td>
</tr>
<tr>
<td></td>
<td>Low dose predeps very difficult</td>
</tr>
</tbody>
</table>
### Dopants are soluble in bulk silicon up to a maximum value before they precipitate into another phase.

### Dopants may have an “electrical” solubility that is different than the solid solubility defined above.

### One example - As$_4$V - electrically inactive complex.
Macroscopic dopant redistribution is described by Fick’s first law, which describes how the flux (or flow) of dopant depends on the doping gradient.

\[ F = -D \frac{dC}{dx} \]  \hspace{1cm} (4)

- F is the flux (cm² sec⁻¹). D is isotropic in the silicon lattice.

Fick’s second law describes how the change in concentration in a volume element is determined by the fluxes in/out of the volume.

\[ \frac{\partial C}{\partial t} = \frac{\partial F}{\partial x} = \frac{\partial}{\partial x} \left( D \frac{\partial C}{\partial x} \right) \]  \hspace{1cm} (5)

- If D is a constant this gives

\[ \frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \]  \hspace{1cm} (6)

- This is rarely true in practice but this is the only form of Fick's second law which can be solved analytically.
Analytic Solutions Of Fick’ s Laws

1. **Limited Source**: Consider a fixed dose $Q$, introduced as a delta function at the origin.

- The solution that satisfies Fick’ s second law is

$$C(x,t) = \frac{Q}{2\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$$  \hspace{2cm} (7)

- Important consequences:
  1. Dose $Q$ remains constant
  2. Peak concentration decreases as $1/\sqrt{t}$
  3. Diffusion distance increases as $2\sqrt{Dt}$
2. Constant Source Near A Surface:

- This is similar to the previous case except the diffusion only goes in one direction.

\[
C(x,t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) = C(0,t) \exp\left(-\frac{x^2}{4Dt}\right) \quad (8)
\]

3. Infinite Source:

- The infinite source is made up of small slices each diffusing as a Gaussian.

\[
C(x,t) = \frac{C}{2\sqrt{\pi Dt}} \sum_{i=1}^{n} \Delta x_i \exp\left(-\frac{(x - x_i)^2}{4Dt}\right) \quad (9)
\]

- The solution which satisfies Fick’s second law is

\[
C(x,t) = \frac{C}{2} \left[1 - \operatorname{erf}\left(\frac{x}{2\sqrt{Dt}}\right)\right] = C_s \left[\operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)\right] \quad (10)
\]
• Important consequences of error function solution:
  • Symmetry about mid-point allows solution for constant surface concentration to be derived.
  • Error function solution is made up of a sum of Gaussian delta function solutions.
  • Dose beyond \( x = 0 \) continues to increase with annealing time.
  
(See Appendix A.9 in text for properties of \( \text{erfc} \).)

4. Constant Surface Concentration:

• Just the right hand side of the above figure.

\[
C(x, t) = C_s \left[ \text{erfc} \frac{x}{2\sqrt{Dt}} \right]
\]  

(11)

• Note that the dose is given by

\[
Q = \int_0^\infty C_s \left[ 1 - \text{erf} \left( \frac{x}{2\sqrt{Dt}} \right) \right] = \frac{2C_s}{\sqrt{\pi}} \sqrt{Dt}
\]

(12)
Intrinsic Dopant Diffusion Coefficients

- Intrinsic dopant diffusion coefficients are found to be of the form:

\[ D = D^0 \exp\left(-\frac{E_A}{kT}\right) \]  

(13)

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>B</th>
<th>In</th>
<th>As</th>
<th>Sb</th>
<th>P</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D^0)</td>
<td>560</td>
<td>1.0</td>
<td>1.2</td>
<td>9.17</td>
<td>4.58</td>
<td>4.70</td>
<td>cm² sec⁻¹</td>
</tr>
<tr>
<td>(E_A)</td>
<td>4.76</td>
<td>3.5</td>
<td>3.5</td>
<td>3.99</td>
<td>3.88</td>
<td>3.68</td>
<td>eV</td>
</tr>
</tbody>
</table>

- Note that \(n_i\) is very large at process temperatures, so "intrinsic" actually applies under many conditions.

- Note the "slow" and "fast" diffusers. Solubility is also an issue in choosing a particular dopant.
Effect Of Successive Diffusions

• If a dopant is diffused at temperature $T_1$ for time $t_1$ and then is diffused at temperature $T_2$ for time $t_2$, the total effective $Dt$ is given by the sum of all the individual $Dt$ products.

$$Dt_{\text{eff}} = \sum Dt = D_1 t_1 + D_2 t_2 + .....$$  \hspace{1cm} (14)

• The Gaussian solution only holds if the $Dt$ used to introduce the dopant is small compared with the final $Dt$ for the drive-in; i.e. if an initial delta function approximation is reasonable.

• Example: In a bipolar transistor, if the emitter profile is formed by a predep and the base profile by an implant + drive-in, then the junctions occur where:

$$C_S \text{erfc} \left( \frac{x}{2\sqrt{Dt}} \right) = \frac{Q}{\sqrt{\pi Dt}} \exp \left( - \frac{x^2}{4Dt} \right)$$

$$(\text{Emitter Dt}) \quad (\text{Base+Emitter Dt})$$

$$\frac{Q}{\sqrt{\pi Dt}} \exp \left( - \frac{x^2}{4Dt} \right) = C_B$$
Design Of Diffused Layers

- Eqn. (3) has been numerically integrated for specific cases (erfc and Gaussian).

- Example of Irvin’s curves, in this case for P type Gaussian profiles. (Other examples are in Appendix A.8 in the text.)

- We can now consider how to design a boron diffusion process (say for the well or tub of a CMOS process - Figs. 2.10 - 2.12), such that:

\[ \rho_s = 900 \Omega/\text{square} \]

\[ x_j = 3 \mu\text{m} \]

\[ N_{BC} = 1 \times 10^{15} \text{ cm}^{-3} \text{ (substrate concentration)} \]
• The average conductivity of the layer is
\[ \bar{\sigma} = \frac{1}{\rho_S x_j} = \frac{1}{(900 \Omega/\text{sq})(3 \times 10^{-4} \text{ cm})} = 3.7 (\Omega \cdot \text{cm})^{-1} \]

• From Irvin's curve we obtain
\[ C_s \approx 4 \times 10^{17} / \text{cm}^3 \]

• We can surmise that the profile is Gaussian after drive-in.
\[ \therefore C_{BC} = \frac{Q}{\sqrt{\pi Dt}} \exp \left( -\frac{x_j^2}{4Dt} \right) = C_s \exp \left( -\frac{x_j^2}{4Dt} \right) \]

so that
\[ Dt = \frac{x_j^2}{4 \ln \frac{C_s}{C_{bc}}} = \frac{(3 \times 10^{-4})^2}{4 \ln \left( \frac{4 \times 10^{17}}{10^{15}} \right)} = 3.7 \times 10^{-9} \text{ cm}^2 \]

• If the drive-in is done at 1100 °C, then the boron diffusivity is
\[ D = 1.5 \times 10^{-13} \text{ cm}^2 \text{ sec}^{-1} \]

• The drive-in time is therefore
\[ t_{\text{drive-in}} = \frac{3.7 \times 10^{-9} \text{ cm}^2}{1.5 \times 10^{-13} \text{ cm}^2 / \text{sec}} = 6.8 \text{ hours} \]
• Given both the surface concentration and \( \text{Dt} \), the initial dose can be calculated for this Gaussian profile.

\[
Q = C_s \sqrt{\pi \text{Dt}} = (4 \times 10^{17}) (\sqrt{\pi}) (\sqrt{3.7 \times 10^{-9}}) = 4.3 \times 10^{13} \text{ cm}^{-2}
\]

• This dose could easily be implanted in a narrow layer close to the surface, justifying the implicit assumption in the Gaussian profile that the initial distribution approximates a delta function.

• If a gas/solid phase predeposition step at 950 \( ^\circ \text{C} \) were used, then

\[
\begin{align*}
\text{B solid solubility at 950 }^\circ\text{C} & = 2.5 \times 10^{20} \text{ cm}^{-3} \\
\text{B diffusivity} & = 4.2 \times 10^{-15} \text{ cm}^{2} \text{ sec}^{-1}
\end{align*}
\]

• The dose for an erfc profile is

\[
Q = \frac{2C_s}{\sqrt{\pi}} \sqrt{\text{Dt}}
\]

• So that the time required for the predeposition is

\[
\begin{align*}
t_{\text{pre-dep}} & = \left( \frac{4.3 \times 10^{13}}{2.5 \times 10^{20}} \right)^{\frac{1}{2}} \left( \frac{\sqrt{\pi}}{2} \right)^{\frac{1}{2}} \frac{1}{4.2 \times 10^{-15}} = 5.5 \text{ sec}
\end{align*}
\]

Check: \( \text{Dt}_{\text{pre-dep}} \left( 2.3 \times 10^{-14} \right) \ll \text{Dt}_{\text{drive-in}} \left( 1.5 \times 10^{-13} \right) \)
Modifications Of Fick's Laws

A. Electric Field Effects

- When the doping is higher than $n_i$, electric field effects become important.
- Electric field induced by higher mobility of electrons and holes compared with dopant ions.
- Electric field enhances the diffusion of dopants causing the field (see derivation in text).

\[ J = -hD \frac{\partial C}{\partial x} \quad (16) \]

where
\[ h = 1 + \frac{C}{\sqrt{C^2 + 4n_i^2}} \quad (17) \]

- SUPREM simulation at 1000°C. Note the boron profile ($h \leq 2$ for the As but electric field effects dominate the B diffusion).
- Field effects can dominate the doping distribution near the source/drain of a MOS device.
B. Concentration Dependent Diffusivity

- At high doping concentrations, the diffusivity appears to increase. Fick's equation must then be solved numerically since $D \neq \text{constant}$.

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left( D_{\text{eff}} \frac{\partial C}{\partial x} \right)$$  \hspace{1cm} (5)

- Isoconcentration experiments indicate the dependence of $D$ on concentration; e.g. $\text{B}^{10}$ in a $\text{B}^{11}$ background.

- Often, $D$ is well described by

$$D_{\text{eff}} = D^0 + D^- \left( \frac{n}{n_i} \right) + D^+ \left( \frac{n}{n_i} \right)^2$$  \hspace{1cm} (18)

- The $n$ and $n^2$ ($p$ and $p^2$ for $P$ type dopants) terms are due to charged defect diffusion mechanisms.

$$D = D_0 \exp \left( -\frac{D.E}{kT} \right)$$  \hspace{1cm} (19)
• SUPREM simulation including field and concentration dependent D effects.

Note:
- field effects around junctions (As determines the field).
- Steep As profile - concentration dependent effects.
- Boron in the base region does not show concentration dependent effects. Boron inside As is slowed down by concentration dependent effects.
- Phosphorus diffusing up inside As profile is affected by concentration effects.
C. Segregation

- Dopants segregate at interfaces. Recall from Chapter 3 in the text,

- This gives an interface flux of

\[ F = h \left( C_A - \frac{C_B}{k_O} \right) \]  \hspace{1cm} (21)

Oxidation of a uniformly doped boron substrate depletes the boron into the growing SiO$_2$ (SUPREM simulation).

N-type dopants tend to pile-up while boron depletes (SUPREM simulation).
D. Interfacial Dopant Pile-up

- Dopants may also segregate to an interface layer, perhaps only a monolayer thick. Interfacial dopant dose loss or pile-up may consume up to 50% of the dose in a shallow layer.

- In the experiment (right) 40% of the dose was lost in a 30 sec anneal.

Kasnavi et. al.
Atomic Scale Diffusion

- Fick's first law macroscopically describes dopant diffusion at low concentrations.
- "Fixes" to this law to account for experimental observations (concentration dependent diffusion and field effects), are useful, but at this point the complexity of the "fixes" begins to outweigh their usefulness.
- Many effects (OED, TED etc) that are very important experimentally, cannot be explained by the macroscopic models discussed so far.

We turn to an atomistic view of diffusion for a deeper understanding.

Vacancy Assisted Mechanism:  
$$A + V \leftrightarrow AV$$

Kick-out and Interstitial(cy) Assisted Mechanisms  
(Identical from a mathematical viewpoint.)  
$$A + I \leftrightarrow AI$$
A. Modeling I And V Components Of Diffusion

- Oxidation provides an I injection source.
- Nitridation provides a V injection source.
- Stacking faults serve as "detectors" as do dopant which diffuse.
- Experiments like these have "proven" that both point defects are important in silicon. Therefore,

\[
D_A = D_A^* \left( f_I \frac{C_I}{C_I^*} + f_V \frac{C_V}{C_V^*} \right) \tag{22}
\]

- Thus dopant diffusion can be enhanced or retarded by changes in the point defect concentrations.

<table>
<thead>
<tr>
<th></th>
<th>$f_I$</th>
<th>$f_V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silicon</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Boron</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>1.0</td>
<td>0</td>
</tr>
<tr>
<td>Arsenic</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Antimony</td>
<td>0.02</td>
<td>0.98</td>
</tr>
</tbody>
</table>
• Measurements of enhanced or retarded diffusion under oxidizing or nitriding conditions allow an estimate of the I or V component of diffusion to be made.

• Oxidation injects interstitials, raises $\frac{C_I}{C_{I}^*}$ and reduces $\frac{C_V}{C_{V}^*}$ through I-V recombination in the bulk silicon. Nitridation does exactly the opposite.

• TSUPREM IV simulations of oxidation enhanced diffusion of boron (OED) and oxidation retarded diffusion of antimony (ORD) during the growth of a thermal oxide on the surface of silicon. The two shallow profiles are antimony, the two deeper profiles are boron.

• Note that the $\frac{C_I}{C_{I}^*}$ and $\frac{C_V}{C_{V}^*}$ profiles are relatively flat indicating the “stiff source” characteristic of the oxidation process.
B. Modeling Atomic Scale Reactions

- Consider the simple chemical reaction
- This contains a surprising amount of physics.
- For example OED is explained because oxidation injects I driving the equation to the right, creating more AI pairs and enhancing the dopant D.

\[ \text{A} + \text{I} \rightleftharpoons \text{AI} \quad (23) \]

Phosphorus diffuses with I, and releases them in the bulk. This enhances the tail region D.

"Emitter push" is also explained by this mechanism.
• If we assume “chemical equilibrium” between dopants and defects in Eqn. (23), then from the law of mass action,

$$C_{AI} = kC_A C_I$$  \hspace{1cm} (24)

• Applying Fick’s law to the mobile species

$$F_{AI} = -d_{AI} \frac{\partial C_{AI}}{\partial x}$$  \hspace{1cm} (25)

• Applying the chain rule from calculus,

$$F_{AI} = -d_{AI} \left( kC_I \frac{\partial C_A}{\partial x} + kC_A \frac{\partial C_I}{\partial x} \right)$$  \hspace{1cm} (26)

• Thus, gradients in defects as well as gradients in dopant concentrations can drive diffusion fluxes.

• The overall flux equation solved by simulators like SUPREM (see text) is

$$F_{BI}^{\text{tot}} = D_{BI}^* \cdot \left( \frac{1 + \beta \frac{p}{n_i}}{1 + \beta} \right) \cdot C_{I^o}^* \cdot C_{B^-} \frac{\partial}{\partial x} \ln \left( \frac{C_{B^-}}{C_{I^o}^* \frac{p}{n_i}} \right)$$  \hspace{1cm} (27)

(written for boron diffusing with neutral and positive interstitials as an example).
Thus there are several distinct effects that drive the dopant diffusion:

- inert, low concentration diffusion, driven by the dopant gradient \( D_{BI}^* \)
- the interstitial supersaturation \( C_{I_o}/C_{I_o}^* \)
- high concentration effects on the dopant diffusivity \( \left(1 + \beta \frac{p}{n_i}\right)/(1 + \beta) \)
- the electric field effect \( \frac{\partial}{\partial x} \ln \frac{p}{n_i} \)

\[ F_{BI}^{tot} = D_{BI}^* \left(1 + \beta \frac{p}{n_i}\right) \cdot \frac{C_{I_o}}{C_{I_o}^*} \cdot C_B^- \cdot \frac{\partial}{\partial x} \ln \left( \frac{C_B^-}{C_{I_o}^*} \cdot \frac{p}{n_i} \right) \]  

(27)

Compare the above expression with Fick’s Law, which is where we started.

\[ F = -D \frac{\partial C}{\partial x} \]  

(4)
• TSUPREM IV simulations of boron diffused from a polysilicon source for 8 hours at 850°C, with experimental data (diamonds) from [7.29]. The simulations use identical coefficients, but with and without full coupling between dopants and defects. The $C_I/C_I^*$ curve (right axis) is for the fully coupled case. Without full coupling, $C_I/C_I^* = 1$.

• No coupling produces a “boxier” profile because of concentration dependent diffusion.

• TSUPREM IV simulations of the interstitial supersaturations generated by a phosphorus versus an arsenic diffusion to the same depth. The fast diffusing phosphorus profile has a larger effect on $C_I/C_I^*$ than the slow diffusing arsenic profile.
• 2D SUPREM simulation of small MOS transistor.
• Ion implantation in the S/D regions generates excess I. These diffuse into the channel region pushing boron (channel dopant) up towards the surface.
• Effect is more pronounced in smaller devices.

(See text for more details on these examples.)
• Assuming $N_D$ or $N_A \leq 2 \times 10^{20} \text{ cm}^{-3}$ and $\mu \leq 52 \text{ cm}^2\text{volt}^{-1} \text{ sec}^{-1}$ an ideal box-shaped profile limits the $x_J r_S$ product to values outside the red area.

• The ITRS goals for S/D extension regions are not physically achievable towards the end of the roadmap without metastable doping concentrations $> 2 \times 10^{20} \text{ cm}^{-3}$. 
Flash annealing - ramp to intermediate $T \approx 800 ^\circ C$ then msec flash to high $T \approx 1300 ^\circ C$.

Recent flash annealing results with boron are much better than RTA.

We’ll see why this works in the next set of notes on ion implantation and damage annealing.

Overall, the shallow junction problem seems manageable through process innovation.
Summary of Key Ideas

- Selective doping is a key process in fabricating semiconductor devices.
- Doping atoms generally must sit on substitutional sites to be electrically active.
- Both doping concentration and profile shape are critical in device electrical characteristics.

- Ion implantation is the dominant process used to introduce dopant atoms. This creates damage and thermal annealing is required to repair this damage.
- During this anneal dopants diffuse much faster than normal (Chapter 8 - TED).

- Atomistic diffusion processes occur by pairing between dopant atoms and point defects.
- In general diffusivities are proportional to the local point defect concentration.
- Point defect concentrations depend exponentially on temperature, and on Fermi level, ion implant damage, and surface processes like oxidation.
- As a result dopant diffusivities depend on time and spatial position during a high temperature step.

- Powerful simulation tools exist today which model these processes and which can predict complex doping profiles.