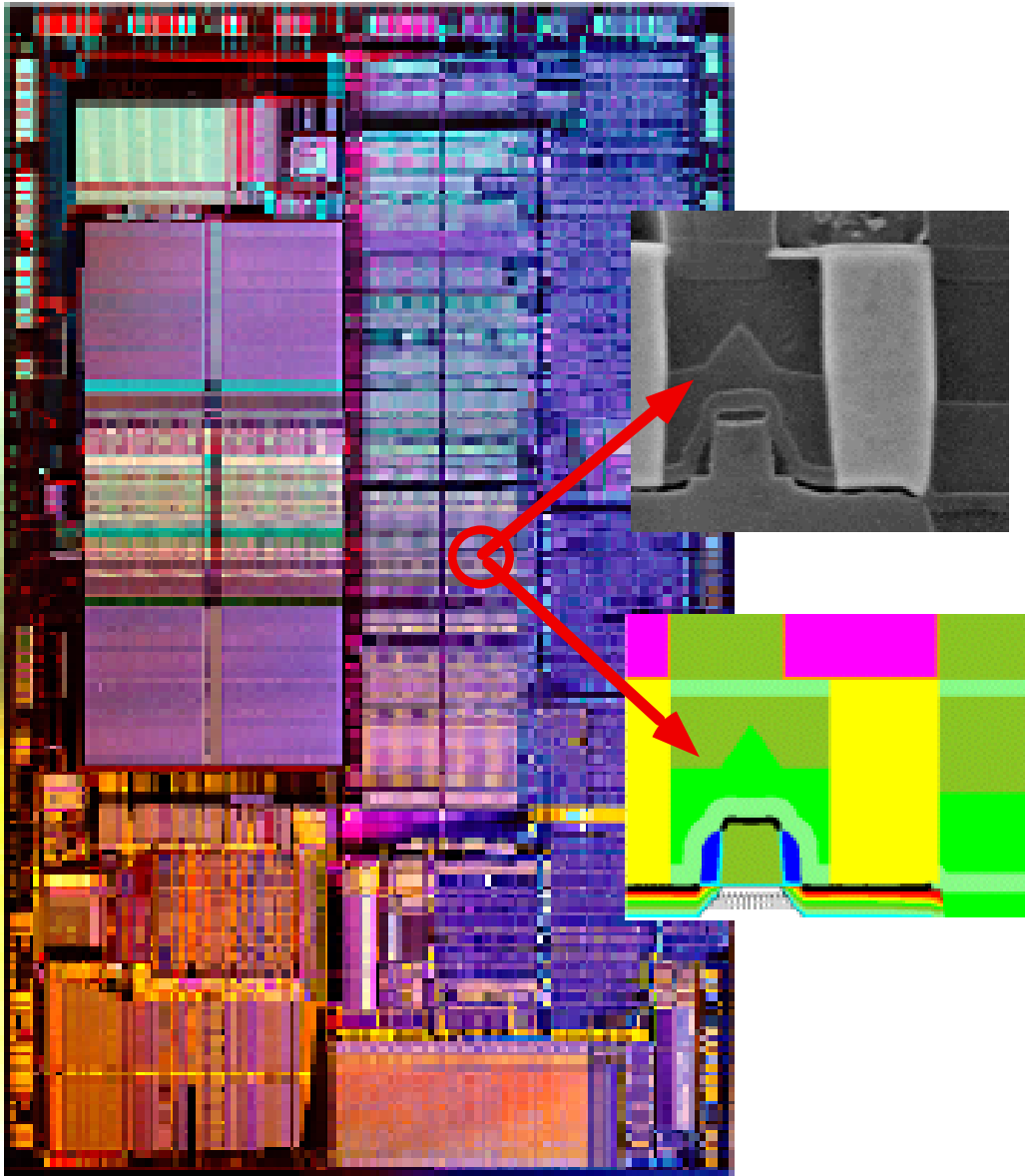


SILICON VLSI TECHNOLOGY Fundamentals, Practice and Models

Solutions Manual for Instructors

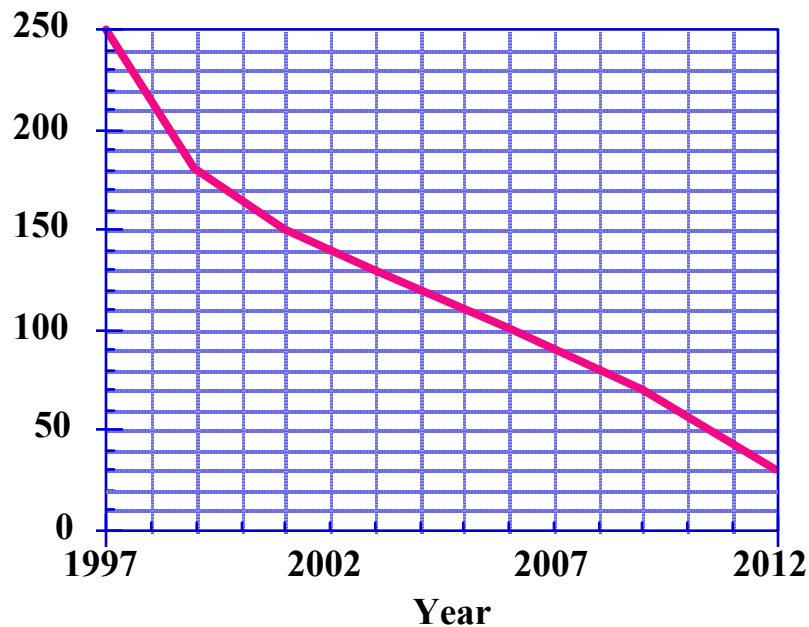


**James D. Plummer
Michael D. Deal
Peter B. Griffin**

Chapter 1 Problems

- 1.1. Plot the NRTS roadmap data from Table 1.1 (feature size vs. time) on an expanded scale version of Fig. 1.2. Do all the points lie exactly on a straight line? If not what reasons can you suggest for any deviations you observe?

Answer:



Interestingly, the actual data seems to consist of two slopes, with a steeper slope for the first 2 years of the roadmap. Apparently the writers of the roadmap are more confident of the industry's ability to make progress in the short term as opposed to the long term.

- 1.2. Assuming dopant atoms are uniformly distributed in a silicon crystal, how far apart are these atoms when the doping concentration is a). 10^{15} cm^{-3} , b). 10^{18} cm^{-3} , c). $5 \times 10^{20} \text{ cm}^{-3}$.

Answer:

The average distance between the dopant atoms would just be one over the cube root of the dopant concentration:

$$x = N_A^{-1/3}$$

$$\text{a) } x = (1 \times 10^{15} \text{ cm}^{-3})^{-1/3} = 1 \times 10^{-5} \text{ cm} = 0.1 \mu\text{m} = 100 \text{ nm}$$

$$\text{b) } x = (1 \times 10^{18} \text{ cm}^{-3})^{-1/3} = 1 \times 10^{-6} \text{ cm} = 0.01 \mu\text{m} = 10 \text{ nm}$$

$$c) x = \left(5 \times 10^{20} \text{ cm}^{-3}\right)^{-1/3} = 1.3 \times 10^{-7} \text{ cm} = 0.0013 \mu\text{m} = 1.3 \text{ nm}$$

- 1.3. Consider a piece of pure silicon 100 μm long with a cross-sectional area of 1 μm^2 . How much current would flow through this “resistor” at room temperature in response to an applied voltage of 1 volt?**

Answer:

If the silicon is pure, then the carrier concentration will be simply n_i . At room temperature, $n_i \approx 1.45 \times 10^{10} \text{ cm}^{-3}$. Under an applied field, the current will be due to drift and hence,

$$\begin{aligned} I &= I_n + I_p = qAn_i(\mu_n + \mu_p)E \\ &= \left(1.6 \times 10^{-19} \text{ coul}\right) \left(10^{-8} \text{ cm}^2\right) \left(1.45 \times 10^{10} \text{ carriers cm}^{-3}\right) \left(2000 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}\right) \left(\frac{1 \text{ volt}}{10^{-2} \text{ cm}}\right) \\ &= 4.64 \times 10^{-12} \text{ amps or } 4.64 \text{ pA} \end{aligned}$$

- 1.4. Estimate the resistivity of pure silicon in $\Omega \text{ ohm cm}$ at a) room temperature, b) 77K, and c) 1000 °C. You may neglect the temperature dependence of the carrier mobility in making this estimate.**

Answer:

The resistivity of pure silicon is given by Eqn. 1.1 as

$$\rho = \frac{1}{q(\mu_n n + \mu_p p)} = \frac{1}{qn_i(\mu_n + \mu_p)}$$

Thus the temperature dependence arises because of the change in n_i with T. Using Eqn. 1.4 in the text, we can calculate values for n_i at each of the temperatures of interest. Thus

$$n_i = 3.1 \times 10^{16} T^{3/2} \exp\left(-\frac{0.603 \text{ eV}}{kT}\right)$$

which gives values of $\approx 1.45 \times 10^{10} \text{ cm}^{-3}$ at room T, $7.34 \times 10^{21} \text{ cm}^{-3}$ at 77K and $5.8 \times 10^{18} \text{ cm}^{-3}$ at 1000 °C. Taking room temperature values for the mobilities, $\mu_n = 1500 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$ and $\mu_p = 500 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$, we have,

$$\begin{aligned}
\rho &= 2.15 \times 10^5 \Omega \text{cm at room T} \\
&= 4.26 \times 10^{35} \Omega \text{cm at 77K} \\
&= 5.39 \times 10^{-4} \Omega \text{cm at 1000 K}
\end{aligned}$$

Note that the actual resistivity at 77K would be much lower than this value because trace amounts of donors or acceptors in the silicon would produce carrier concentrations much higher than the n_i value calculated above.

1.5. a). Show that the minimum conductivity of a semiconductor sample occurs

when $n = n_i \sqrt{\frac{\mu_p}{\mu_n}}$.

b). What is the expression for the minimum conductivity?

c). Is this value greatly different than the value calculated in problem 1.2 for the intrinsic conductivity?

Answer:

a).

$$\sigma = \frac{1}{\rho} = q(\mu_n n + \mu_p p)$$

To find the minimum we set the derivative equal to zero.

$$\therefore \frac{\partial \sigma}{\partial n} = \frac{\partial}{\partial n} \{q(\mu_n n + \mu_p p)\} = \frac{\partial}{\partial n} \left\{ q \left(\mu_n n + \mu_p \frac{n_i^2}{n} \right) \right\} = q \left(\mu_n + \mu_p \frac{n_i^2}{n^2} \right) = 0$$

$$\therefore n^2 = n_i^2 \frac{\mu_p}{\mu_n} \quad \text{or} \quad n = n_i \sqrt{\frac{\mu_p}{\mu_n}}$$

b). Using the value for n derived above, we have:

$$\sigma_{\min} = q \left(\mu_n n_i \sqrt{\frac{\mu_p}{\mu_n}} + \mu_p \frac{n_i^2}{n_i \sqrt{\frac{\mu_p}{\mu_n}}} \right) = q \left(\mu_n n_i \sqrt{\frac{\mu_p}{\mu_n}} + \mu_p n_i \sqrt{\frac{\mu_n}{\mu_p}} \right) = 2q n_i \sqrt{\mu_n \mu_p}$$

c). The intrinsic conductivity is given by

$$\sigma_i = q n_i (\mu_n + \mu_p)$$

Taking values of $n_i = 1.45 \times 10^{10} \text{ cm}^{-3}$, $\mu_n = 1500 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$ and $\mu_p = 500 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$, we have:

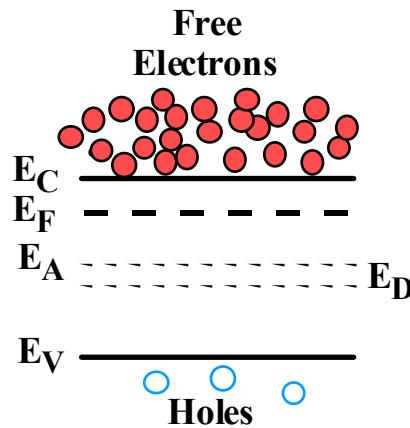
$$\sigma_i = 4.64 \times 10^{-6} \Omega \text{cm} \quad \text{and} \quad \sigma_{\min} = 4.02 \times 10^{-6} \Omega \text{cm}$$

Thus there are not large differences between the two.

- 1.6. When a Au atom sits on a lattice site in a silicon crystal, it can act as either a donor or an acceptor. E_D and E_A levels both exist for the Au and both are close to the middle of the silicon bandgap. If a small concentration of Au is placed in an N type silicon crystal, will the Au behave as a donor or an acceptor? Explain.**

Answer:

In N type material, the Fermi level will be in the upper half of the bandgap as shown in the band diagram below. Allowed energy levels below E_F will in general be occupied by electrons. Thus the E_D and E_A levels will have electrons filling them. This means the donor level will not have donated its electron whereas the acceptor level will have accepted an electron. Thus the Au atoms will act as acceptors in N type material.



- 1.7. Show that E_F is approximately in the middle of the bandgap for intrinsic silicon.**

Answer:

Starting with Eqn. 1.9 and 1.10 in the text, we have

$$n \cong N_C \exp\left(-\frac{E_C - E_F}{kT}\right) \quad \text{and} \quad p \cong N_V \exp\left(-\frac{E_F - E_V}{kT}\right)$$

In intrinsic material, $n = p = n_i$, so we have

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_V \exp\left(-\frac{E_F - E_V}{kT}\right)$$

$$\therefore \ln\left(\frac{N_C}{N_V}\right) = -\frac{E_F - E_V}{kT} + \frac{E_C - E_F}{kT}$$

Using the definition of N_C and N_V from Eqn. 1.13, we have

$$-\frac{3}{4} kT \ln \frac{m_e^*}{m_h^*} = -2E_F + E_C + E_V$$

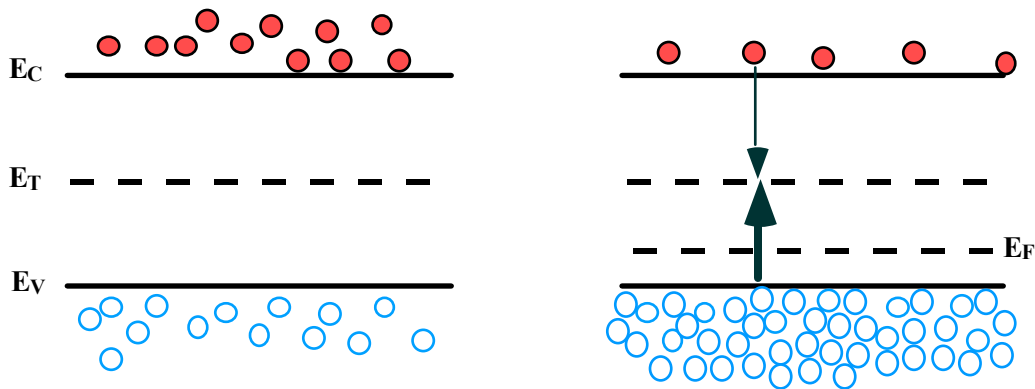
$$\therefore E_F = \frac{E_C + E_V}{2} - \frac{3}{4} kT \ln \frac{m_e^*}{m_h^*}$$

The second term on the right is negligible compared to the first term on the right, so we have finally that

$$\therefore E_F \cong \frac{E_C + E_V}{2} = E_i$$

- 1.8. Construct a diagram similar to Figure 1-27b for P-type material. Explain physically, using this diagram, why the capture of the minority carrier electrons in P-type material is the rate limiting step in recombination.**

Answer:



In P type material (on the right), there are many more holes than electrons. Thus the probability of capturing a hole by the trap is high compared to electrons. So the normal state of occupancy of the trap level will be with a hole present. The recombination rate will then be limited by the capture of the "scarce" species which is the minority carrier or electron in this case.

Stated another way, since E_F is below the trap level in P type material, E_T will normally not be occupied by an electron. Thus the limiting process is the electron capture.

- 1.9. A silicon diode has doping concentrations on the N and P sides of $N_D = 1 \times 10^{19} \text{ cm}^{-3}$ and of $N_A = 1 \times 10^{15} \text{ cm}^{-3}$. Calculate the process temperature at which the two sides of the diode become intrinsic. (Intrinsic is defined as $n_i = N_D$ or N_A .)**

Answer:

Each side of the diode will become intrinsic at the temperature at which $n_i = N_D$ or N_A . We can estimate these temperatures by looking at the graph in Fig. 1-16. From the graph, the N side of the diode will become intrinsic around 1400K or 1100°C. The P side will become intrinsic at a lower temperature since N_A is smaller than N_D . From the figure, the P side becomes intrinsic at about 500K or 200°C. More accurately, we solve Eqn. 1.4 to find the exact temperature. Thus

$$n_i = 3.1 \times 10^{16} T^{3/2} \exp\left(-\frac{0.603 \text{ eV}}{kT}\right)$$

Setting $n_i = 10^{19} \text{ cm}^{-3}$ and 10^{15} cm^{-3} respectively, we and solving by iteration, we find that the N side becomes intrinsic at about 1380K or about 1110°C and the P side becomes intrinsic at about 545K or about 270°C.

- 1.10. A state-of-the-art NMOS transistor might have a drain junction area of $0.5 \times 0.5 \text{ } \mu\text{m}$. Calculate the junction capacitance associated with this junction at an applied reverse bias of 2 volts. Assume the drain region is very heavily doped and the substrate doping is $1 \times 10^{16} \text{ cm}^{-3}$.**

Answer:

The capacitance of the junction is given by Eqn. 1.25.

$$\frac{C}{A} = \frac{\epsilon_S}{x_d} = \left[\sqrt{\frac{q\epsilon_S}{2} \left(\frac{N_A N_D}{N_A + N_D} \right) \frac{1}{(\phi_i \pm V)}} \right]$$

The junction built-in voltage is given by Eqn. 1.24. N_D is not specified except that it is very large, so we take it to be 10^{20} cm^{-3} (roughly solid solubility). The exact choice for N_D doesn't make much difference in the answer.

$$\phi_i = \frac{kT}{q} \ln\left(\frac{N_D N_A}{n_i^2}\right) = (.0259 \text{ volts}) \ln\left(\frac{(10^{20} \text{ cm}^{-3})(10^{16} \text{ cm}^{-3})}{(1.45 \times 10^{10} \text{ cm}^{-3})^2}\right) = 0.934 \text{ volts}$$

Since $N_D \gg N_A$ in this structure, the capacitance expression simplifies to

$$\frac{C}{A} \cong \frac{\epsilon_S}{W} = \left[\sqrt{\frac{q\epsilon_S}{2} (N_A)} \frac{1}{(\phi_i \pm V)} \right]$$

$$= \left[\frac{(1.6 \times 10^{-19} \text{ coul})(11.7)(10^{16} \text{ cm}^{-3})(8.86 \times 10^{-14} \text{ Fcm}^{-1})}{(2)(2.934 \text{ volts})} \right] = 1.68 \times 10^{-8} \text{ Fcm}^{-2}$$

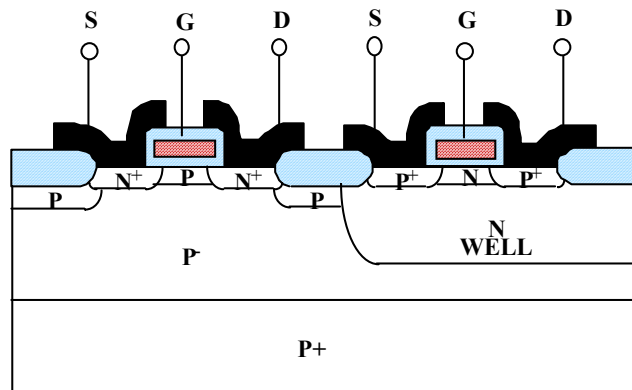
Given the area of the junction ($0.25 \times 10^{-8} \text{ cm}^2$), the junction capacitance is thus 4.2×10^{-17} Farads.

Chapter 2 Problems

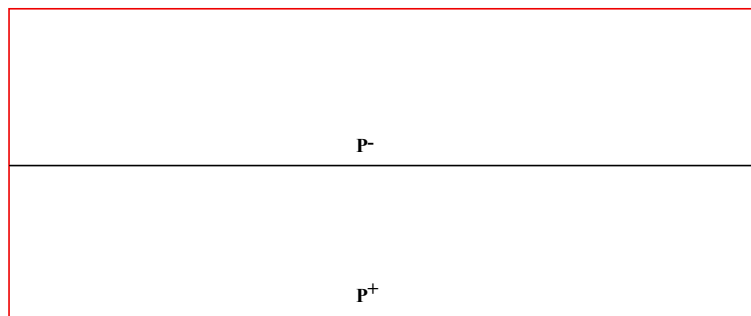
2.1 Sketch a process flow that would result in the structure shown in Figure 1-34 by drawing a series of drawings similar to those in this chapter. You only need to describe the flow up through the stage at which active device formation starts since from that point on, the process is similar to that described in this chapter.

Answer:

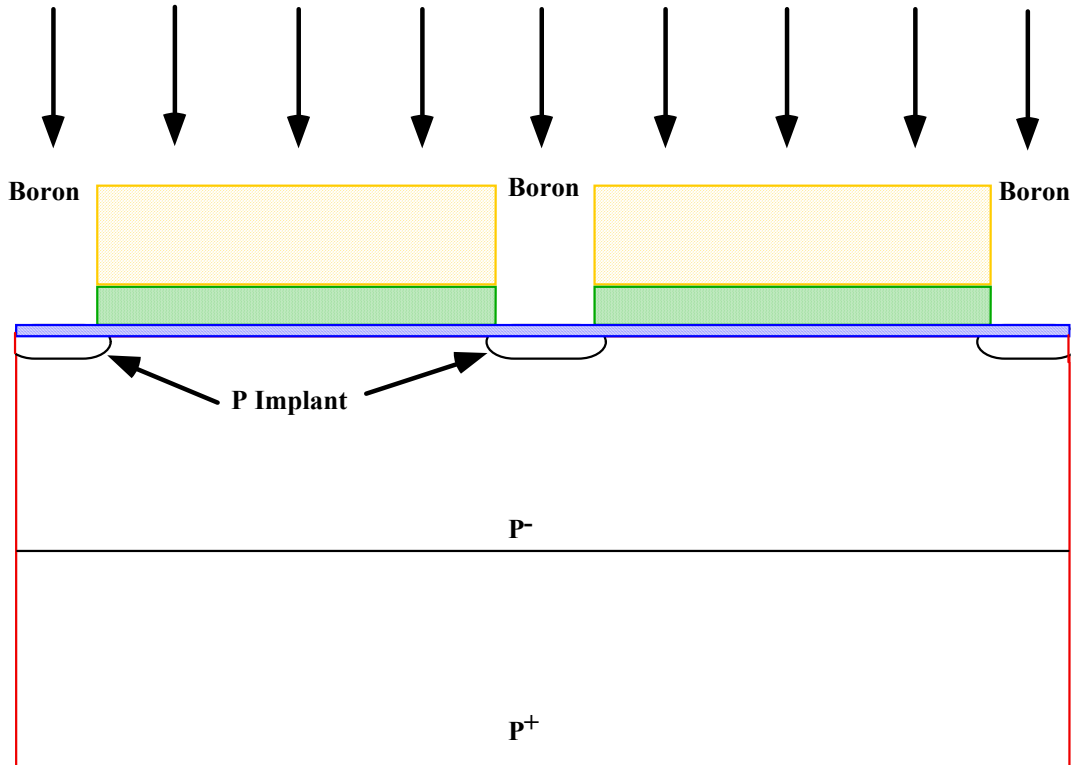
The CMOS technology we need to realize is shown below, from Figure 1-34 in the text.



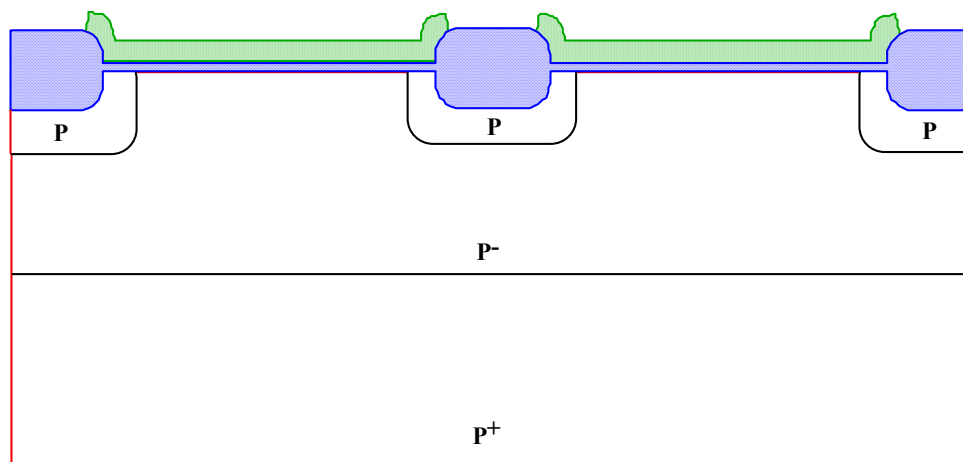
We can follow many of the process steps used in the CMOS process flow in Chapter 2. The major differences are that an epi layer is needed, only one well (P well) used, and the device structures are considerably simplified from those in the text because there are no LDD regions etc.



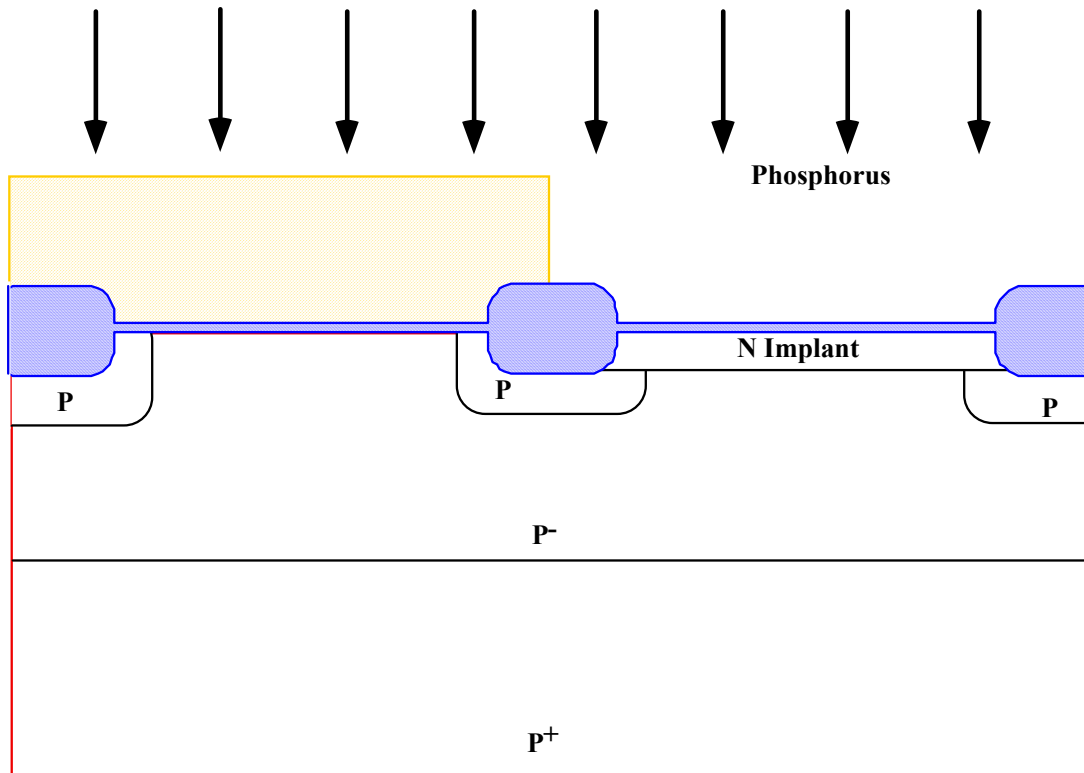
The first step is to grow the blanket epi layer shown in the final cross-section. A heavily doped P^+ substrate is chosen and a lightly doped boron epitaxial layer is grown uniformly on its surface.



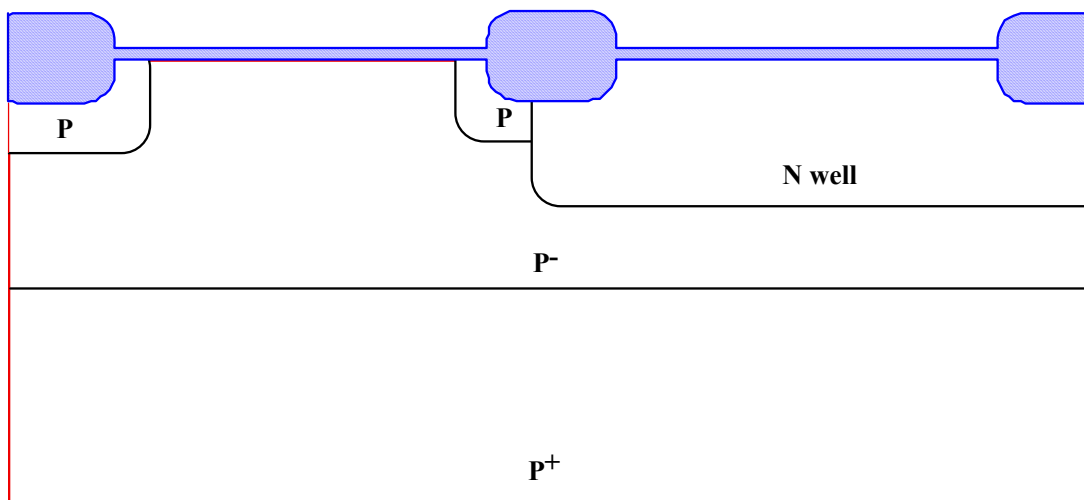
Mask #1 patterns the photoresist. The Si_3N_4 layer is removed where it is not protected by the photoresist by dry etching. Since the technology uses field implants below the field oxide, a boron implant is used to dope these P regions.



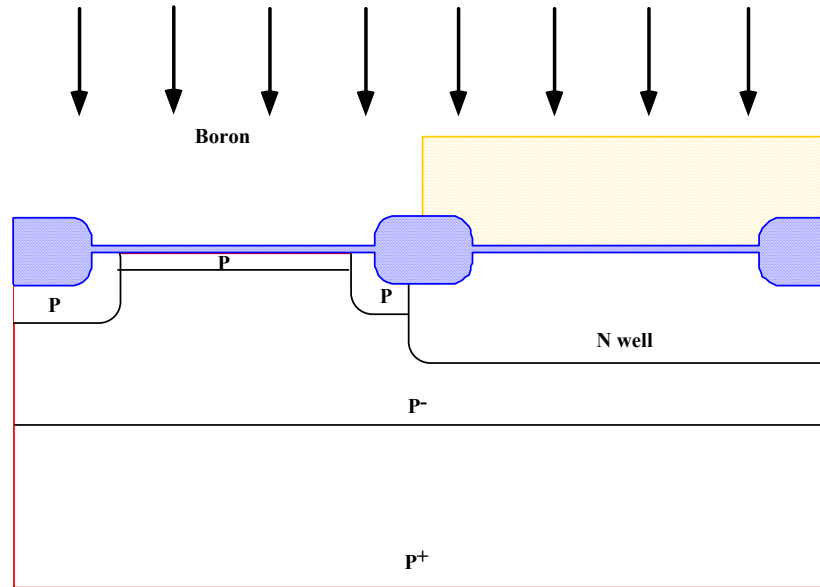
During the LOCOS oxidation, the boron implanted regions diffuse ahead of the growing oxide producing the P doped regions under the field oxide. The Si_3N_4 is stripped after the LOCOS process.



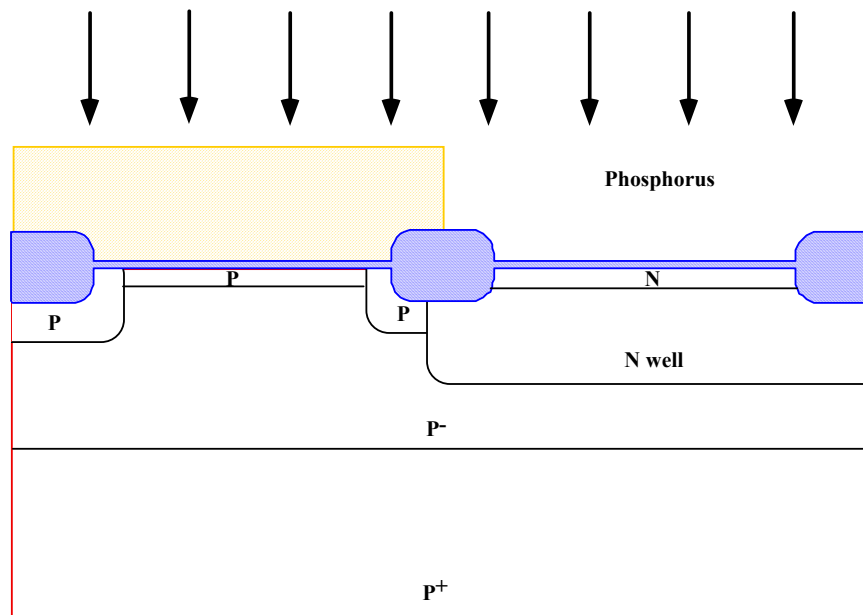
Mask #2 is used to form the N well. Photoresist is used to mask the regions where NMOS devices will be built. A phosphorus implant provides the doping for the N wells for the PMOS devices



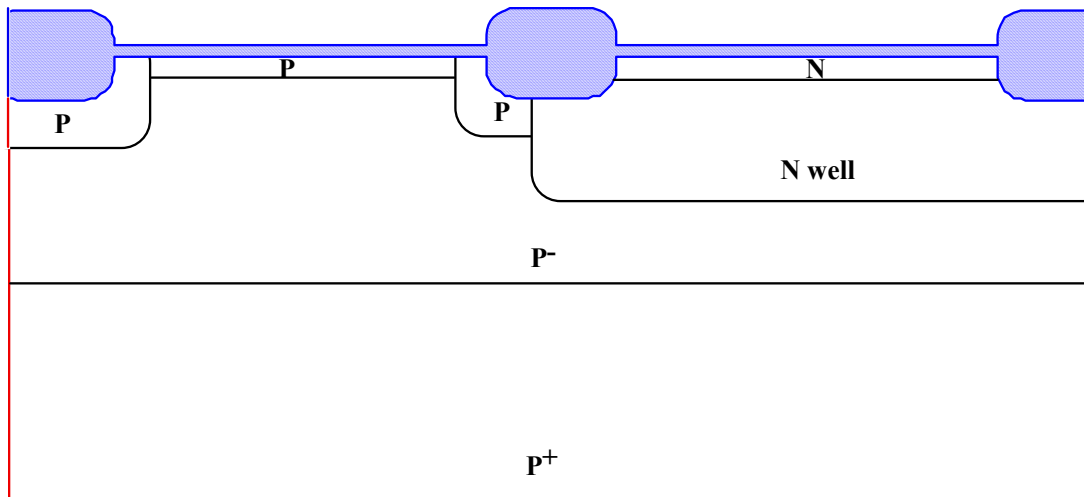
A high temperature drive-in completes the formation of the N well.



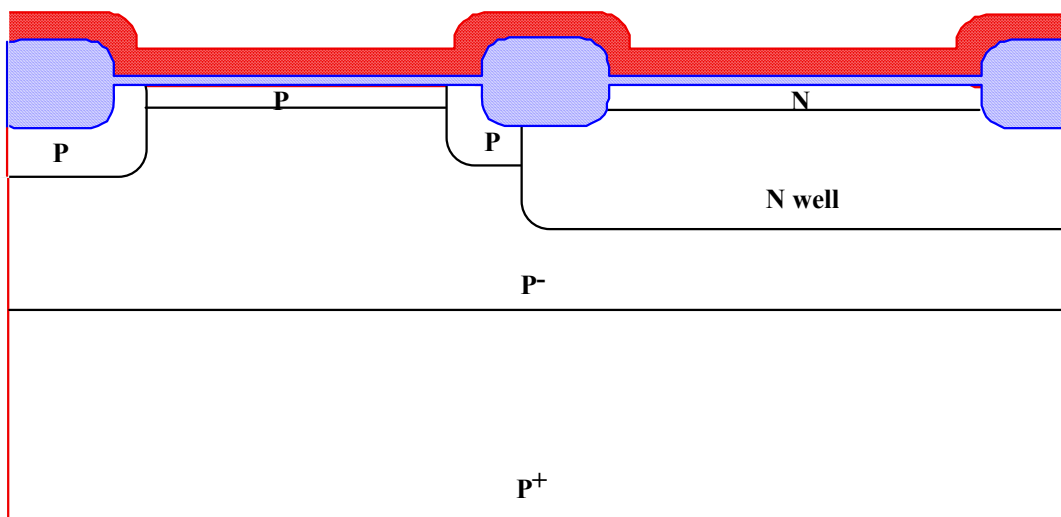
After spinning photoresist on the wafer, mask #3 is used to define the NMOS transistors. A boron implant adjusts the N channel V_{TH} .



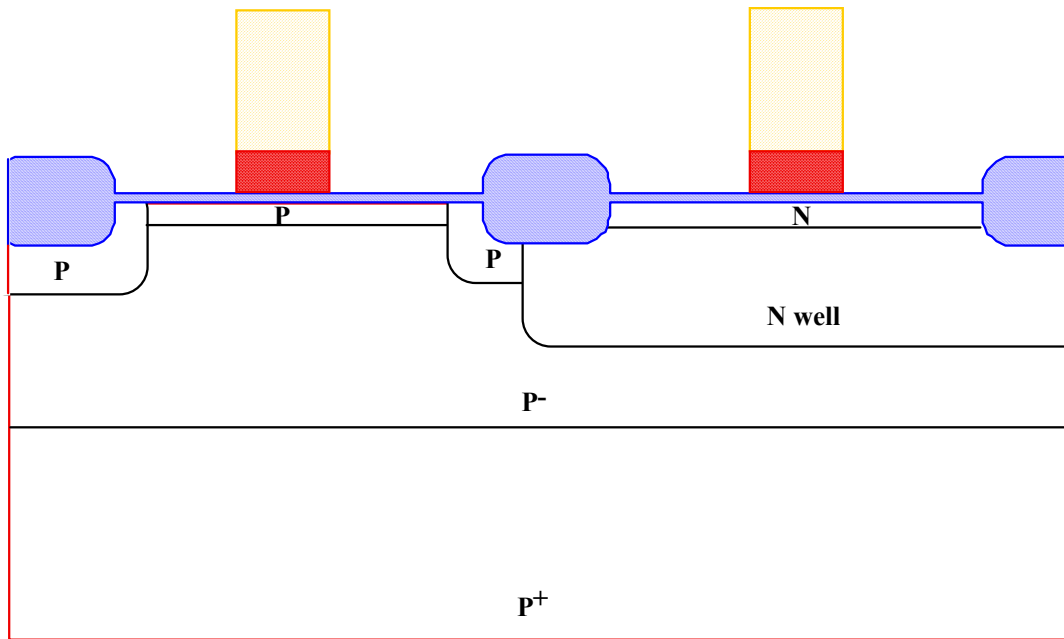
After spinning photoresist on the wafer, mask #4 is used to define the PMOS transistors. A phosphorus or arsenic implant adjusts the P channel V_{TH} . (Depending on the N well doping, a boron implant might actually be needed at this point instead of an N type implant, to obtain the correct threshold voltage.)



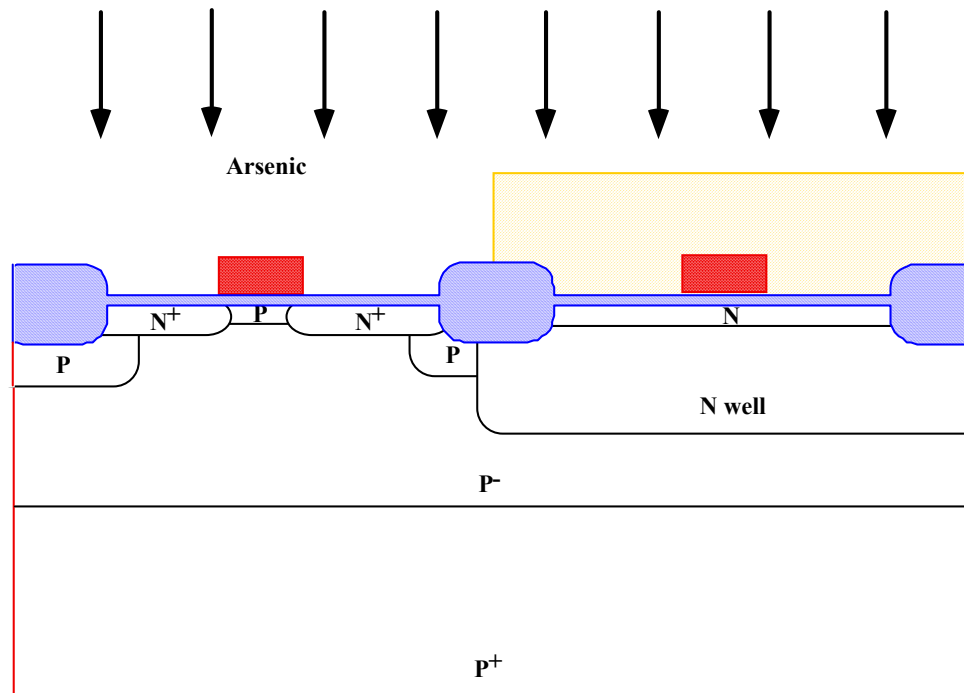
After etching back the thin oxide to bare silicon, the gate oxide is grown for the MOS transistors.



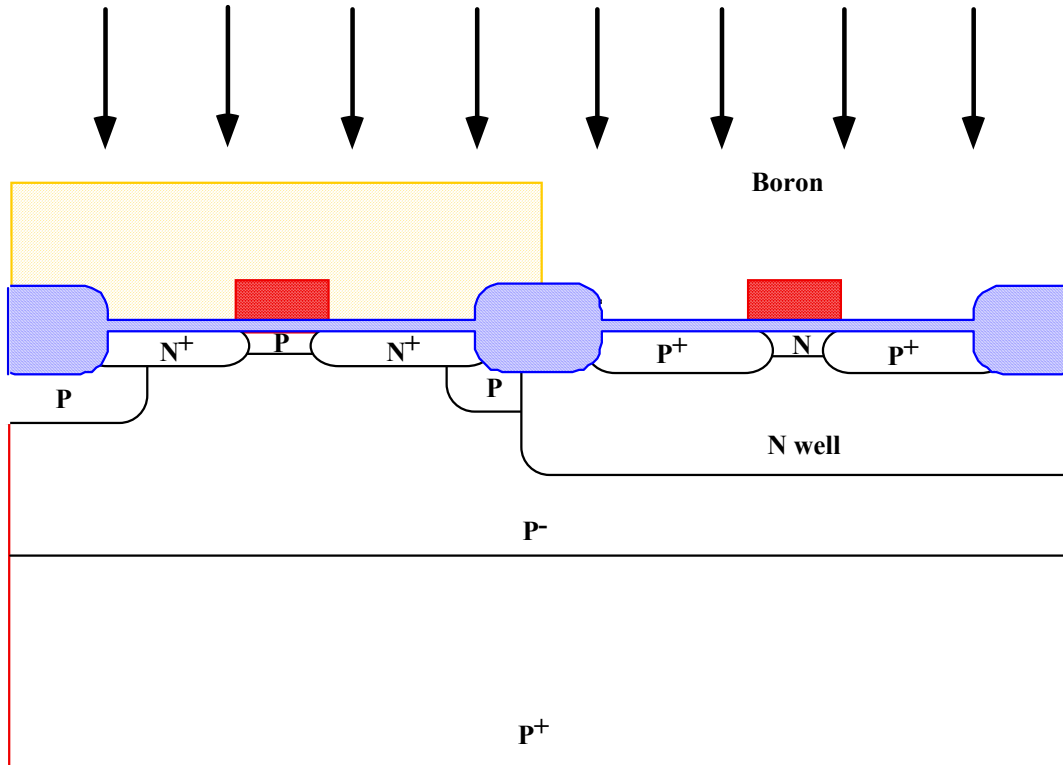
A layer of polysilicon is deposited. Ion implantation of phosphorus follows the deposition to heavily dope the poly.



Photoresist is applied and mask #5 is used to define the regions where MOS gates are located. The polysilicon layer is then etched using plasma etching.



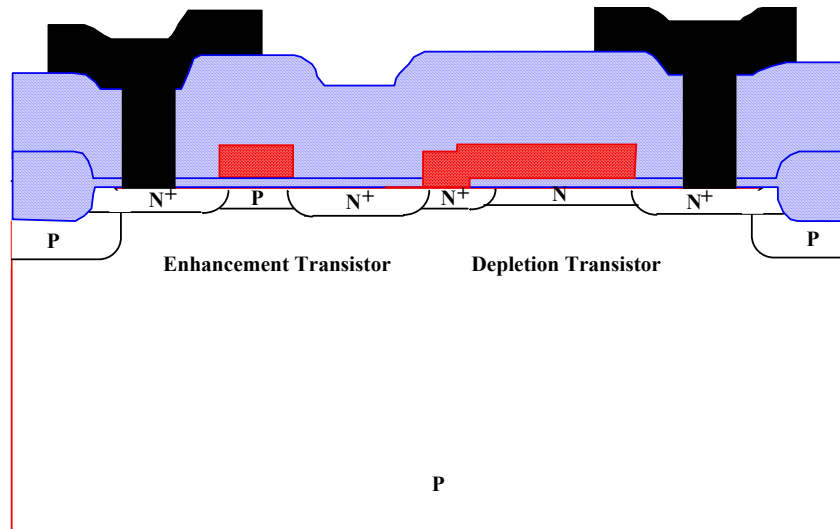
Photoresist is applied and mask #6 is used to protect the PMOS transistors. An arsenic implant then forms the NMOS source and drain regions.



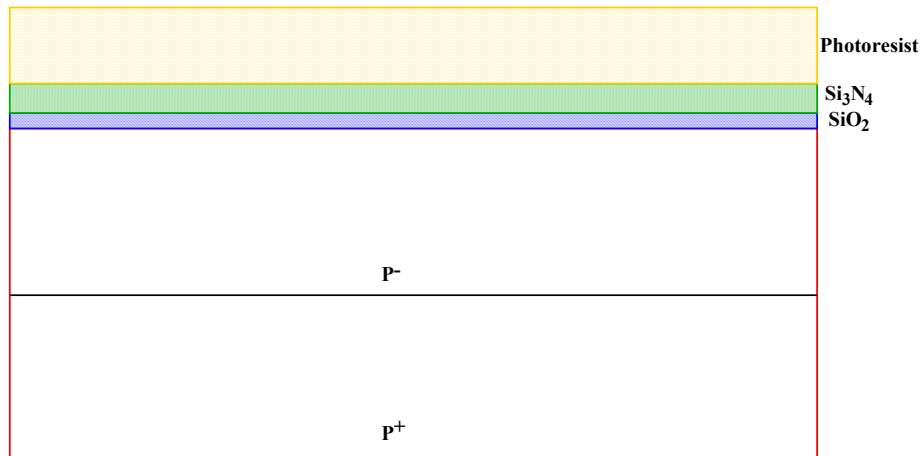
After applying photoresist, mask #7 is used to protect the NMOS transistors. A boron implant then forms the PMOS source and drain regions.

At this point we have completed the formation of the active devices, except for a final high temperature anneal to activate the dopants and drive in the junctions to their final depth. The rest of the process flow would be similar to the CMOS flow in the text.

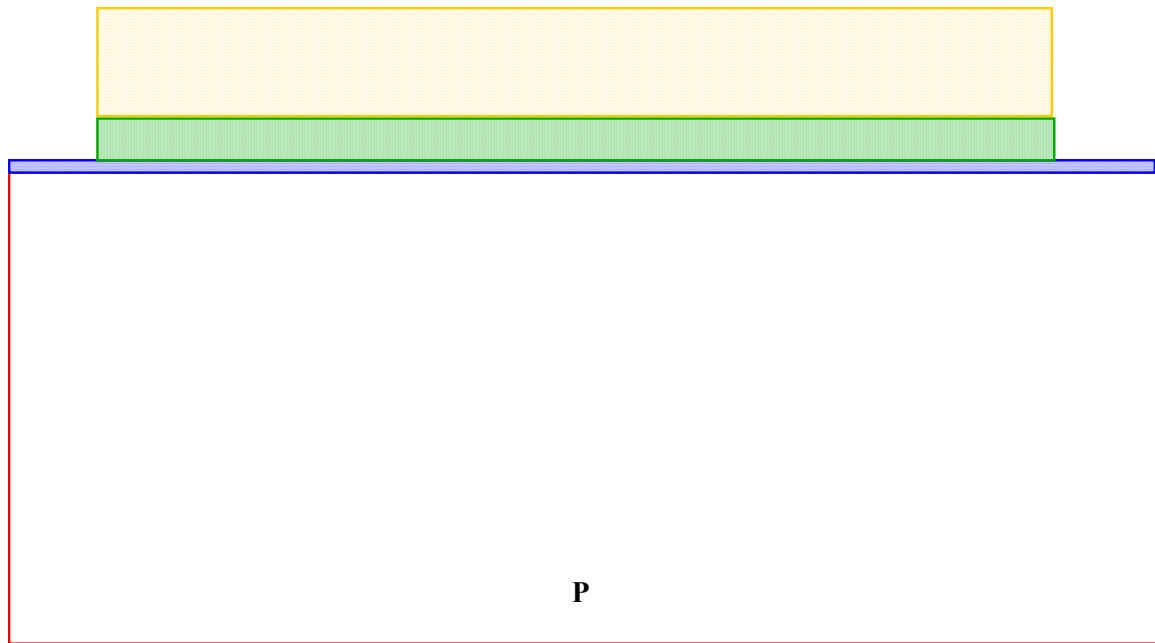
- 2.2. During the 1970s, the dominant logic technology was NMOS as described briefly in Chapter 1. A cross-sectional view of this technology is shown below (see also Figure 1-33). The depletion mode device is identical to the enhancement mode device except that a separate channel implant is done to create a negative threshold voltage. Design a plausible process flow to fabricate such a structure, following the ideas of the CMOS process flow in this chapter. You do not have to include any quantitative process parameters (times, temperatures, doses etc.) Your answer should be given in terms of a series of sketches of the structure after each major process step, like the figures in Chapter 2. Briefly explain your reasoning for each step and the order you choose to do things.

**Answer:**

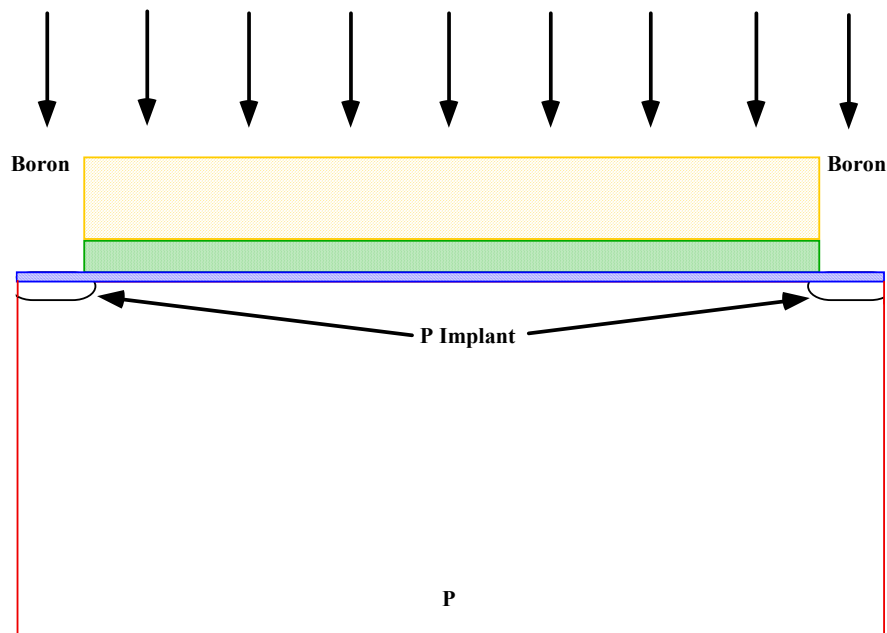
We can follow many of the process steps used in the CMOS process flow in Chapter 2. The major differences are that only NMOS devices are required (2 different threshold voltages however), and there is a buried contact connecting the poly gate of the depletion device to the source region of that device.



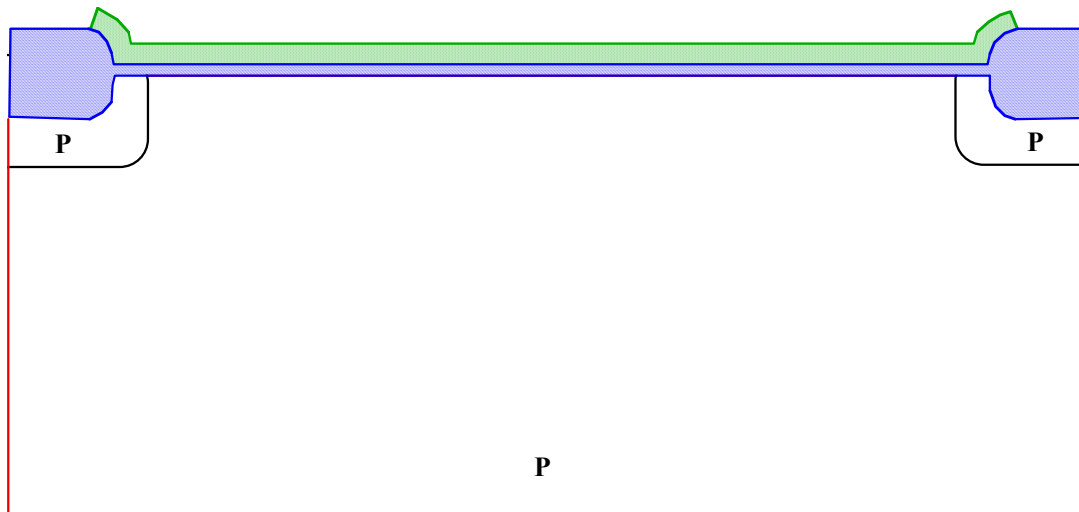
Following initial cleaning, an SiO₂ layer is thermally grown on the silicon substrate. A Si₃N₄ layer is then deposited by LPCVD. Photoresist is spun on the wafer to prepare for the first masking operation.



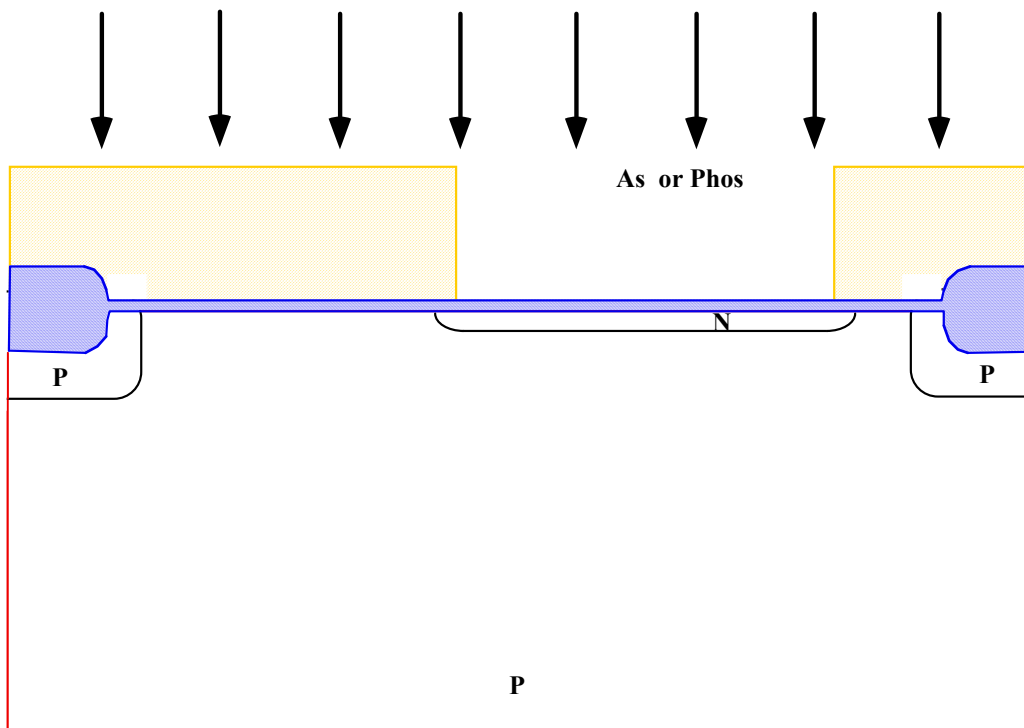
Mask #1 patterns the photoresist. The Si_3N_4 layer is removed where it is not protected by the photoresist by dry etching.



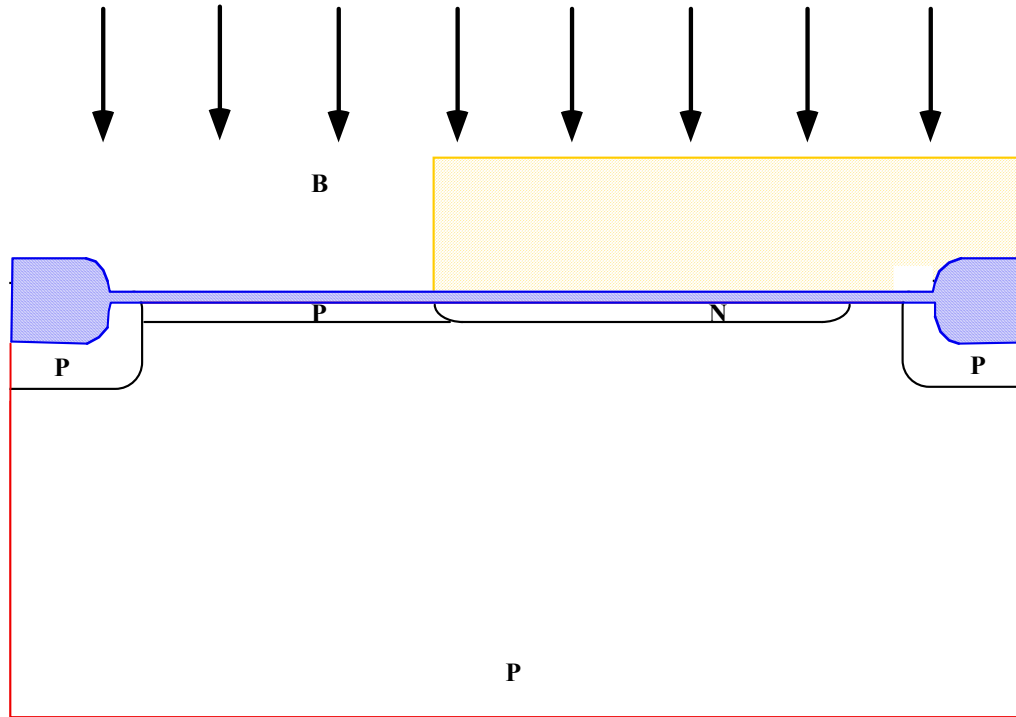
A boron implant prior to LOCOS oxidation increases the substrate doping locally under the field oxide to minimize field inversion problems.



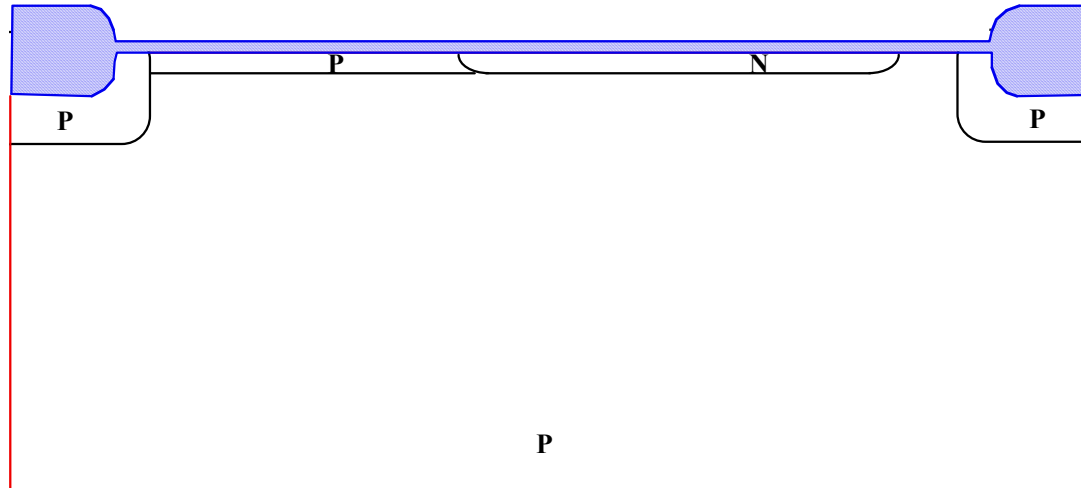
During the LOCOS oxidation, the boron implanted regions diffuse ahead of the growing oxide producing the P doped regions under the field oxide. The Si_3N_4 is stripped after the LOCOS process.



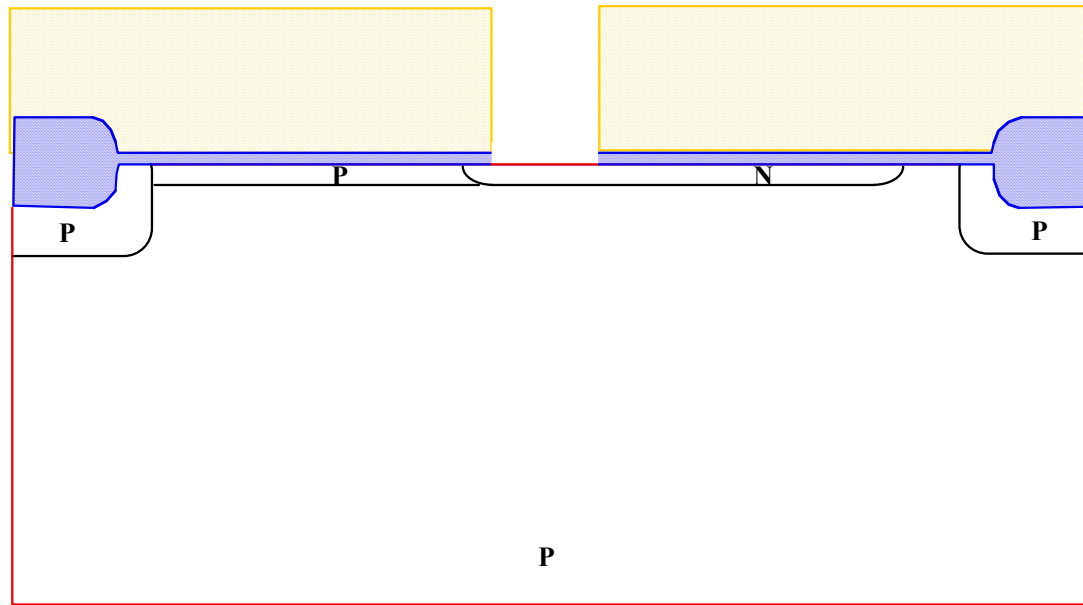
Mask #2 is used for the the threshold shifting implant for the depletion transistors. An N type dopant is implanted.



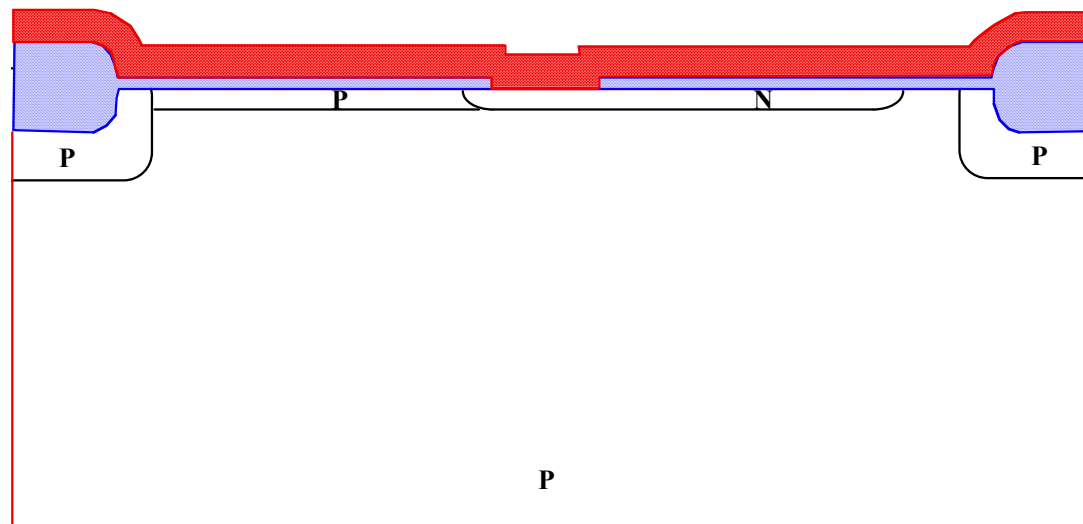
Mask #3 is used to mask the threshold shifting implant for the enhancement transistors. A P type dopant is implanted.



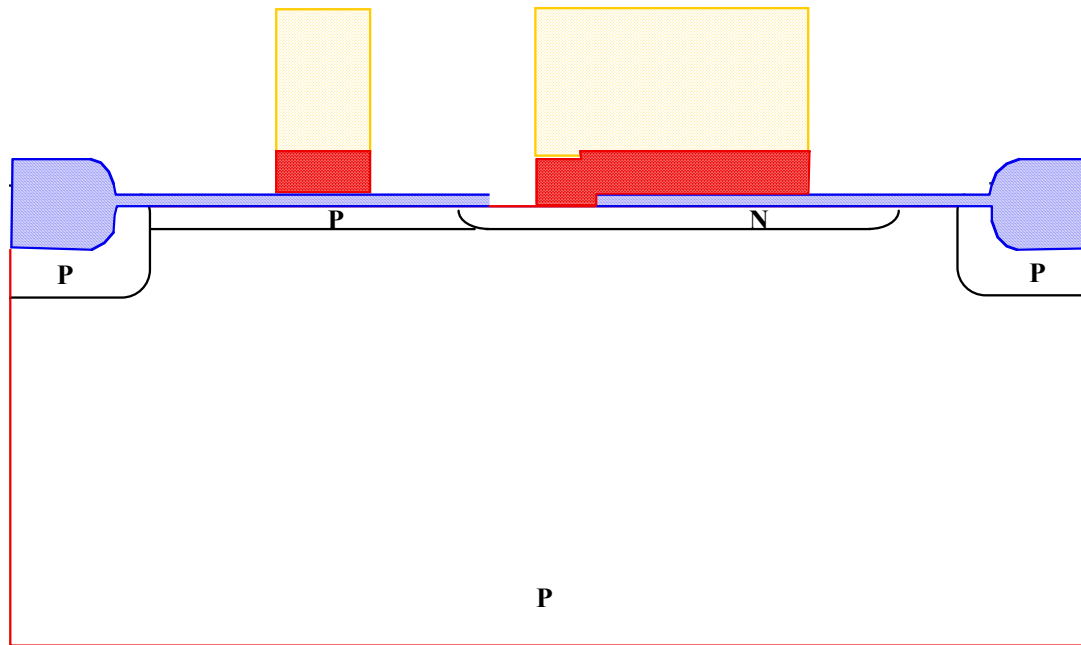
After etching back the thin oxide to bare silicon, the gate oxide is grown for the MOS transistors.



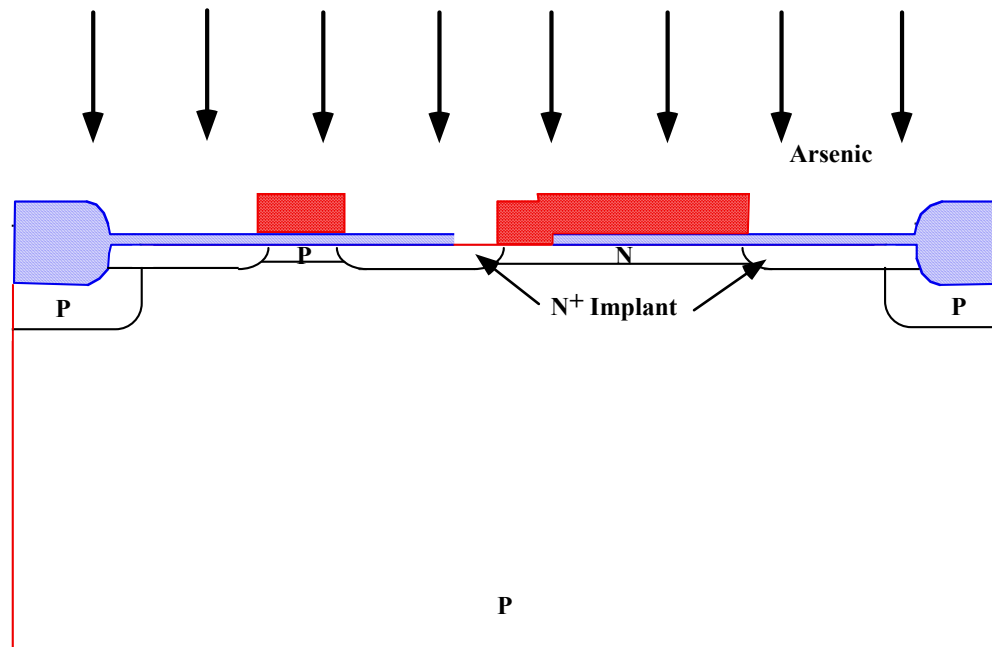
Mask #4 is used to provide the buried contact. The gate oxide is etched where the poly needs to contact the silicon.



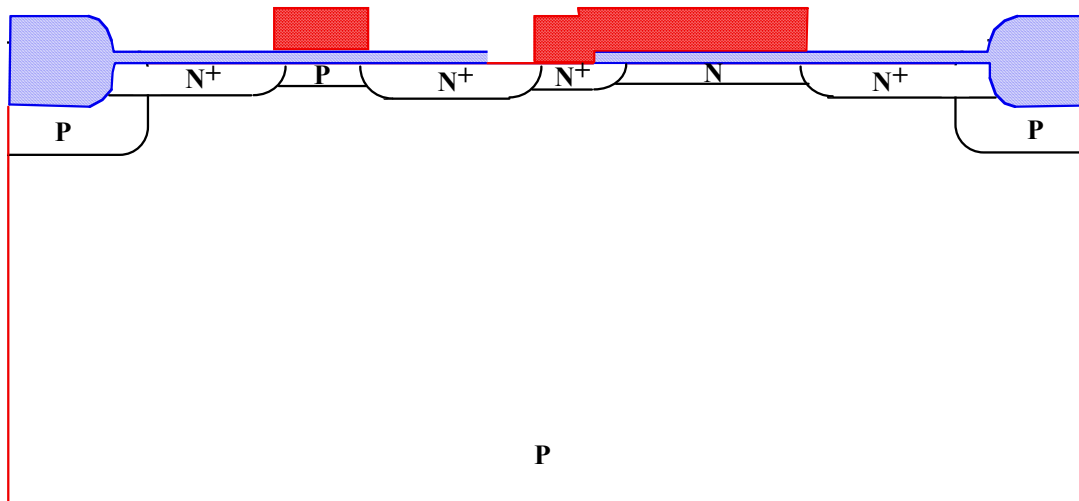
A layer of polysilicon is deposited. Ion implantation of an N type dopant follows the deposition to heavily dope the poly.



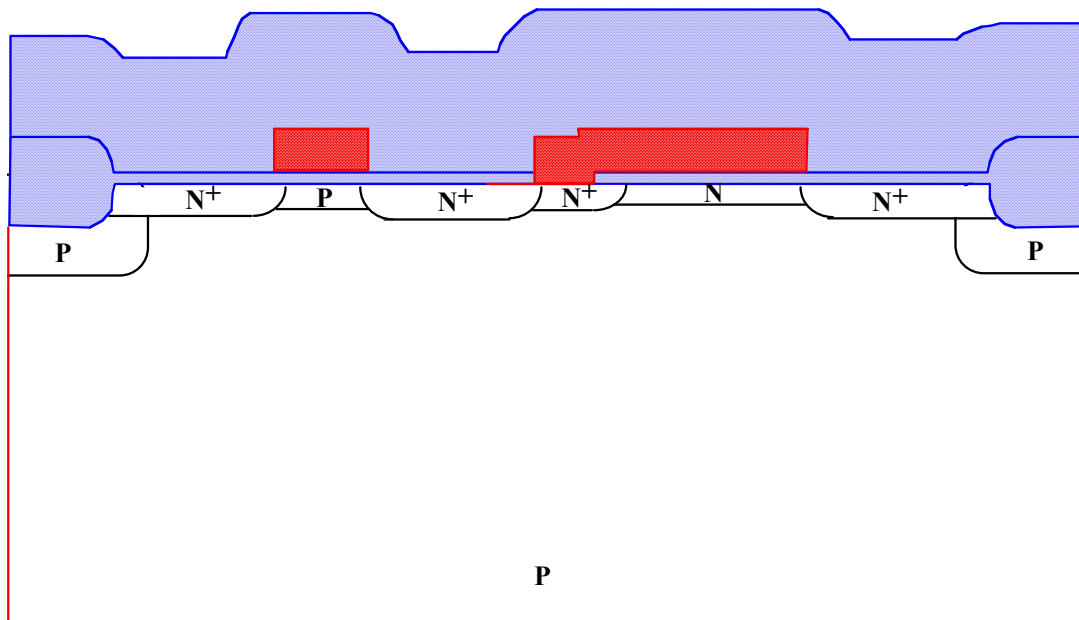
Photoresist is applied and mask #5 is used to define the regions where MOS gates are located. The polysilicon layer is then etched using plasma etching.



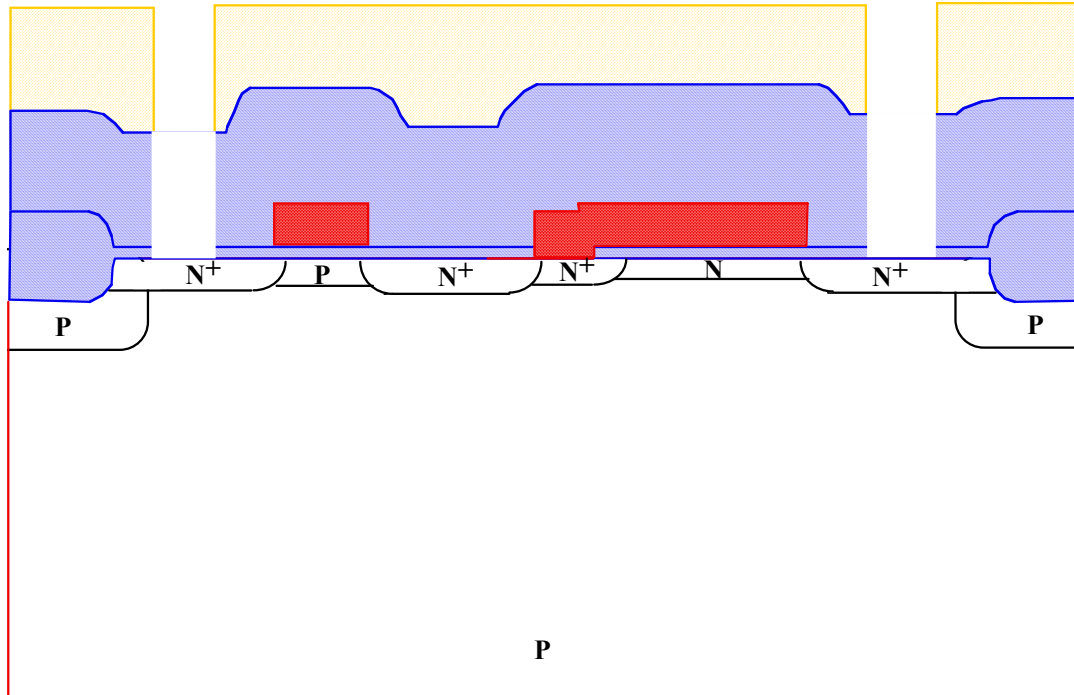
Arsenic is implanted to form the source and drain regions. Note that this can be unmasked because there are only NMOS transistors on the chip.



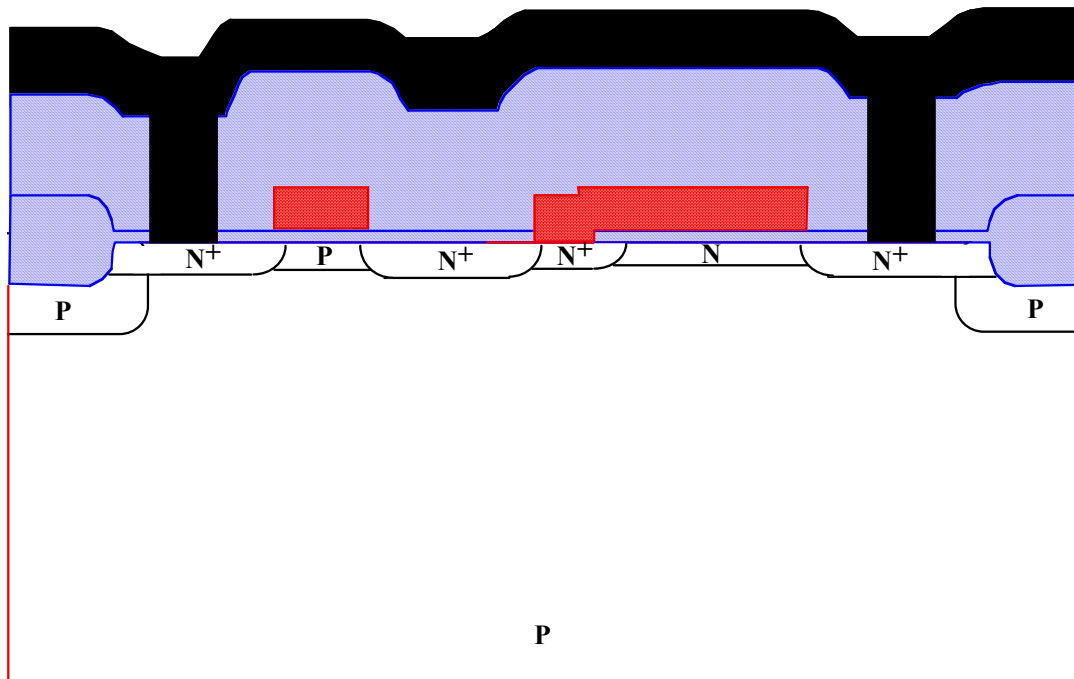
A final high temperature drive-in activates all the implanted dopants and diffuses junctions to their final depth. The N doping in the poly outdiffuses to provide the buried contact.



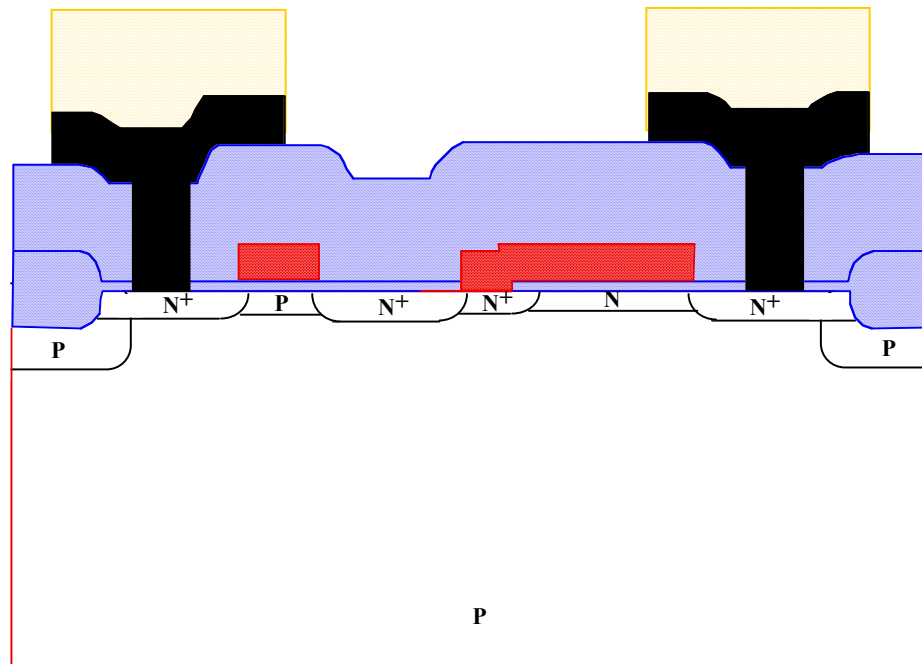
A conformal SiO_2 layer is deposited by LPCVD.



Mask #6 is used to define the contact holes.

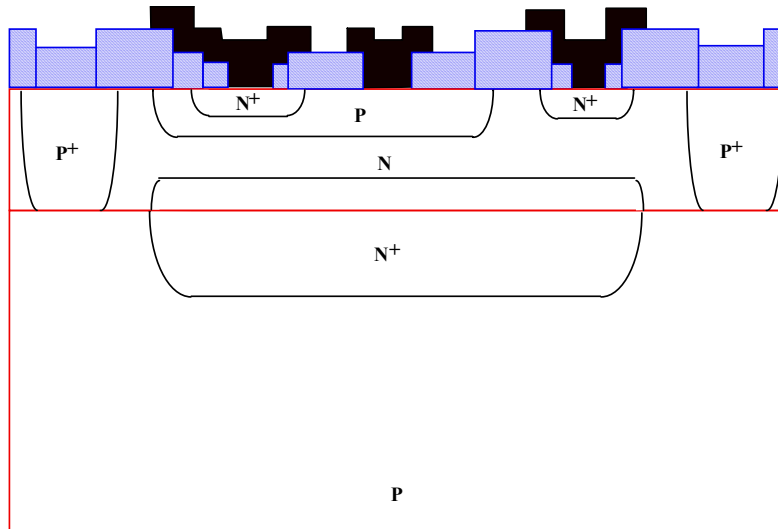


Aluminum is deposited on the wafer.

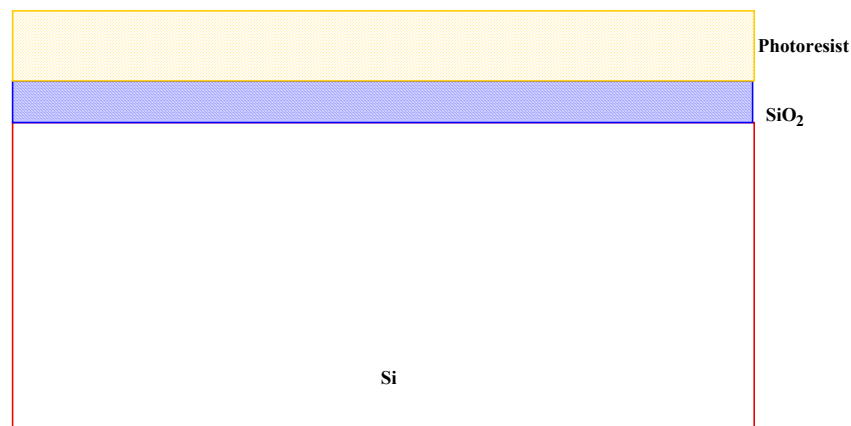


Mask #7 is used to pattern the aluminum. After stripping the resist, the structure is finished to the point shown in the cross-section we started with. In actual practice an additional deposition of a final passivation layer and an additional mask (#8) would be needed to open up the regions over the bonding pads.

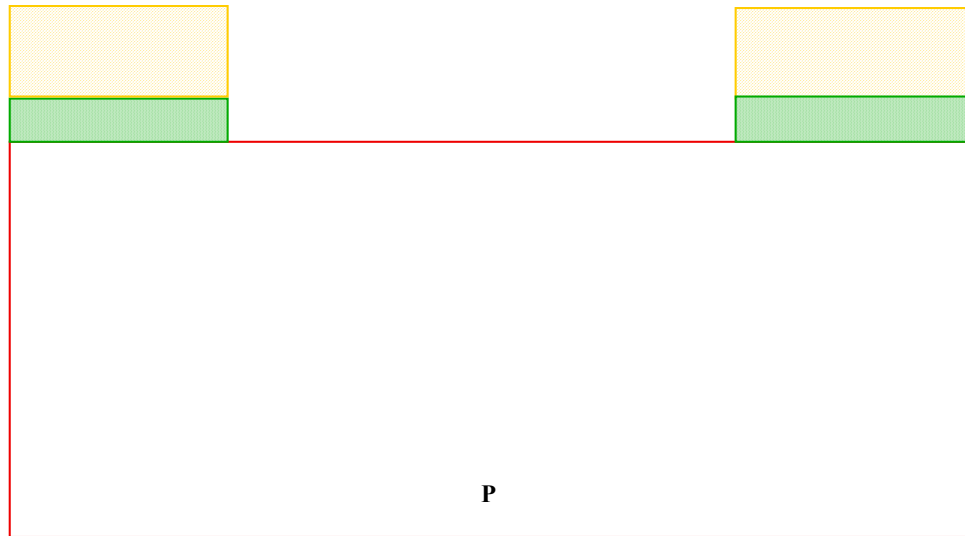
- 2.3. The cross-section below illustrates a simple bipolar transistor fabricated as part of a silicon IC. (See also Figure 1-32.) Design a plausible process flow to fabricate such a structure, following the ideas of the CMOS process flow in this chapter. You do not have to include any quantitative process parameters (times, temperatures, doses etc.) Your answer should be given in terms of a series of sketches of the structure after each major process step, like the figures in this chapter. Briefly explain your reasoning for each step and the order you choose to do things.



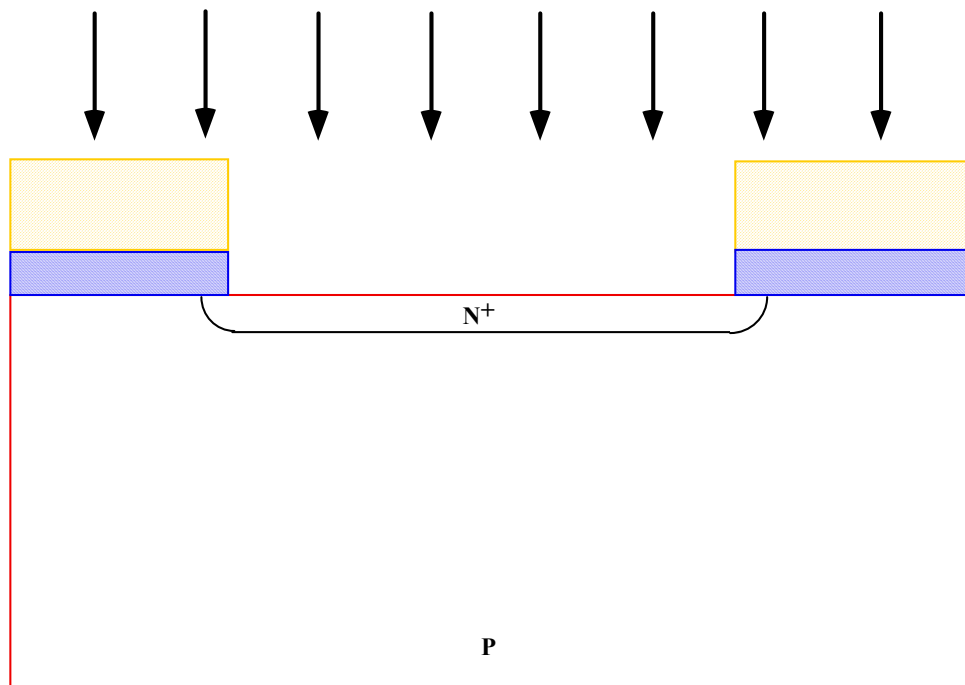
Answer:



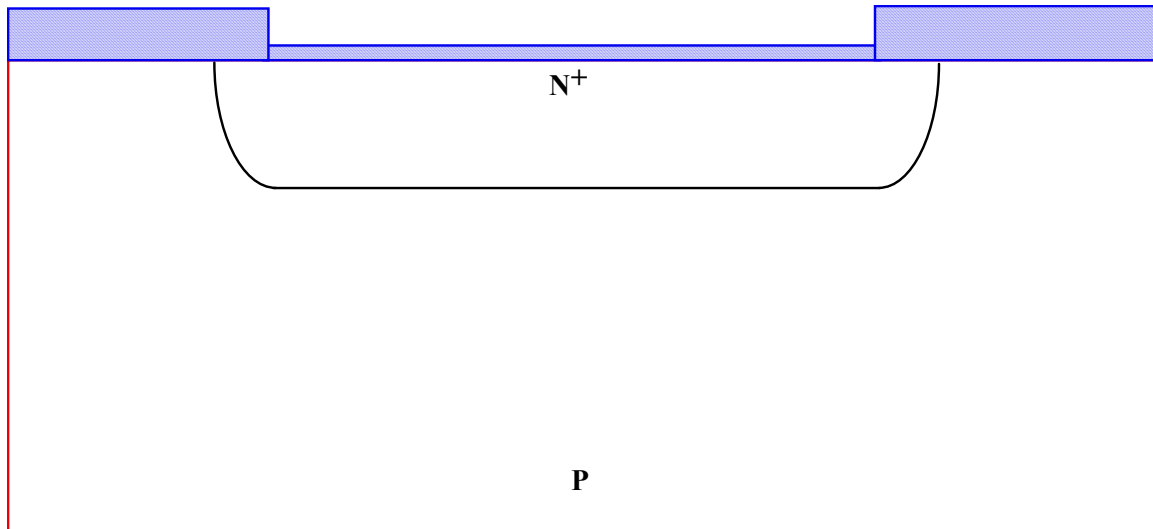
Following initial cleaning, an SiO_2 layer is thermally grown on the silicon substrate. Photoresist is spun on the wafer to prepare for the first masking operation.



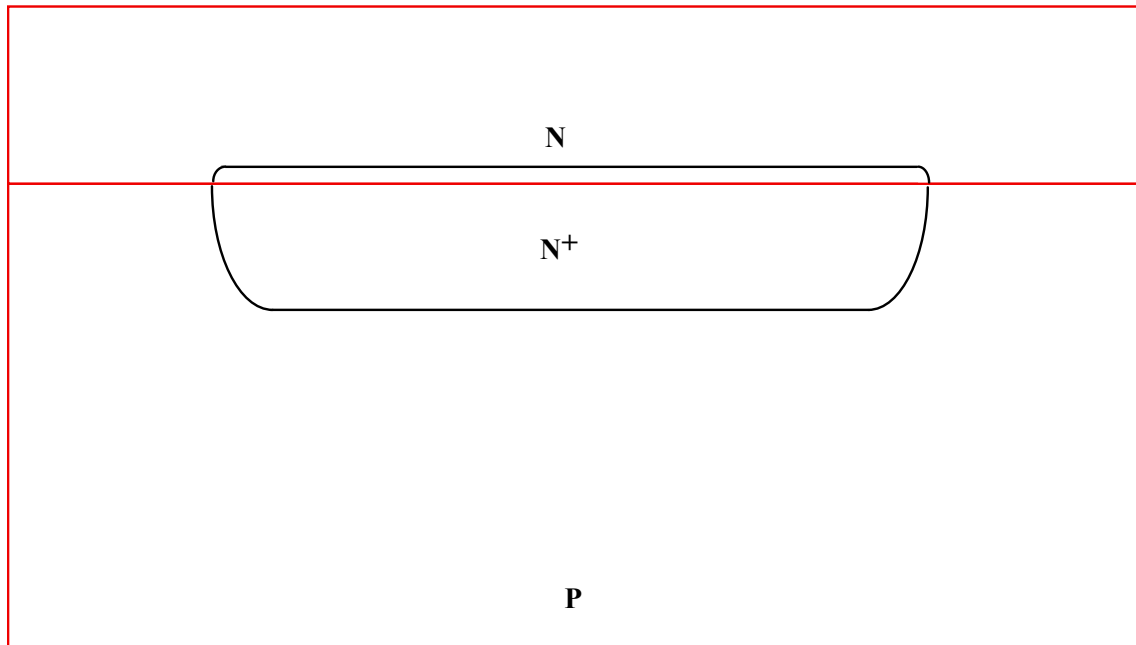
Mask #1 patterns the photoresist. The SiO_2 layer is removed where it is not protected by the photoresist by dry etching.



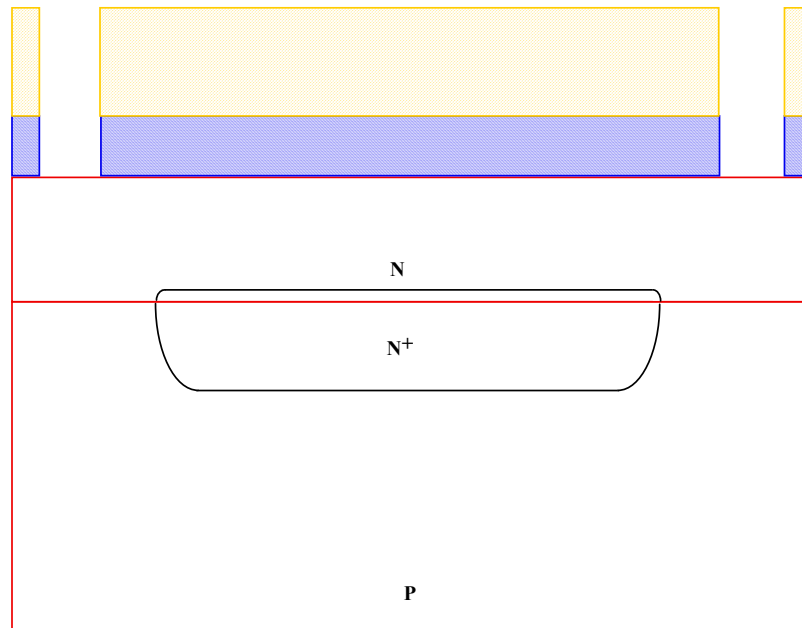
An N^+ implant is performed to dope the buried layer region. As or Sb would typically be used here because they have smaller diffusivities than P.



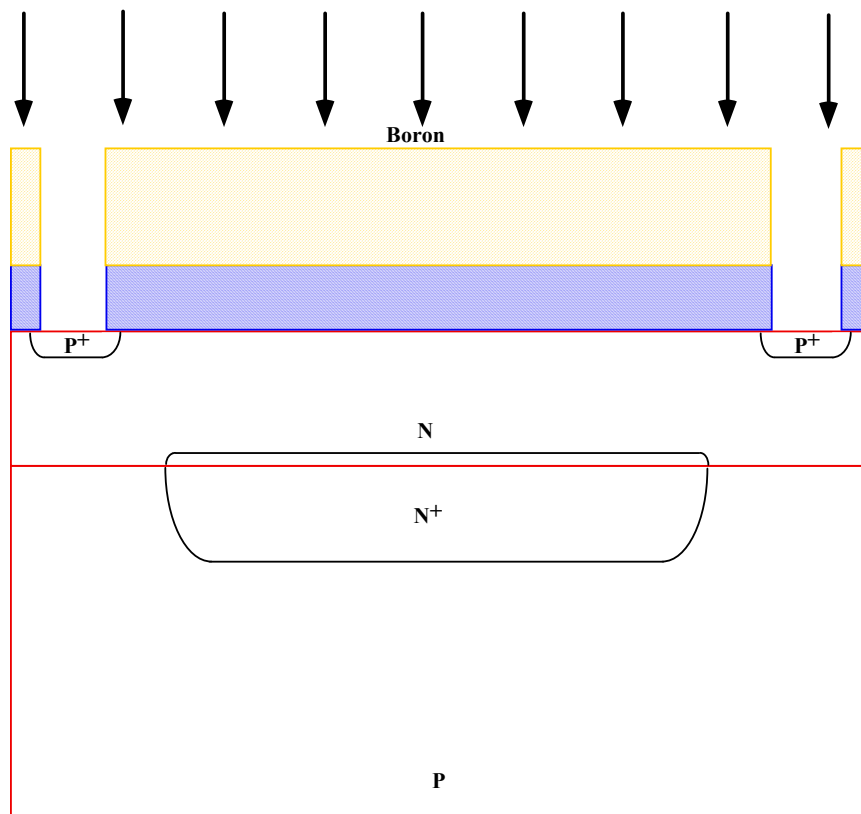
The buried layer is driven in using a high temperature furnace cycle.



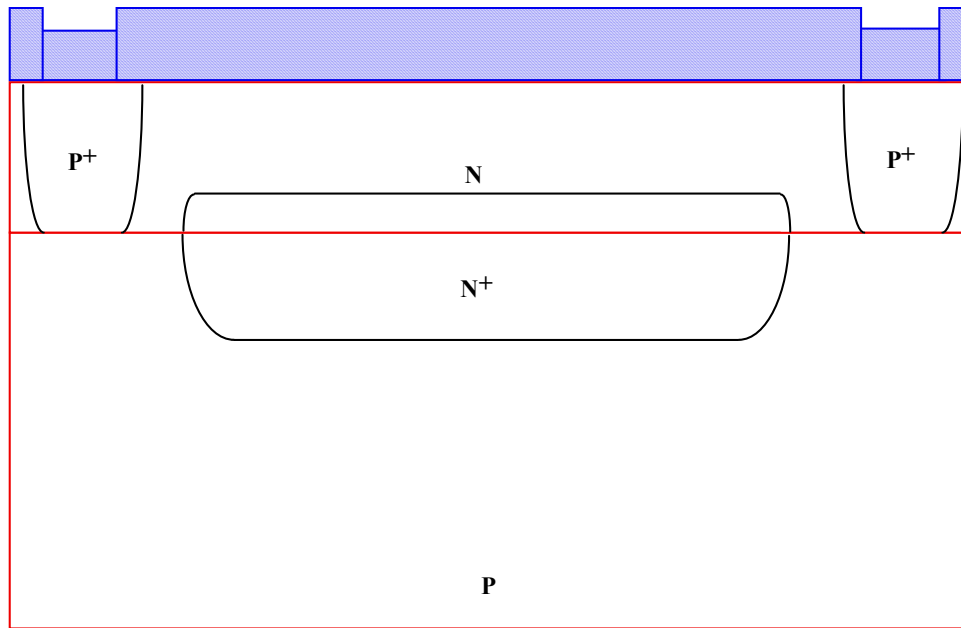
The SiO_2 is etched off the surface and an N type epitaxial layer is grown. Note that during the epi growth, the buried layer diffuses upwards.



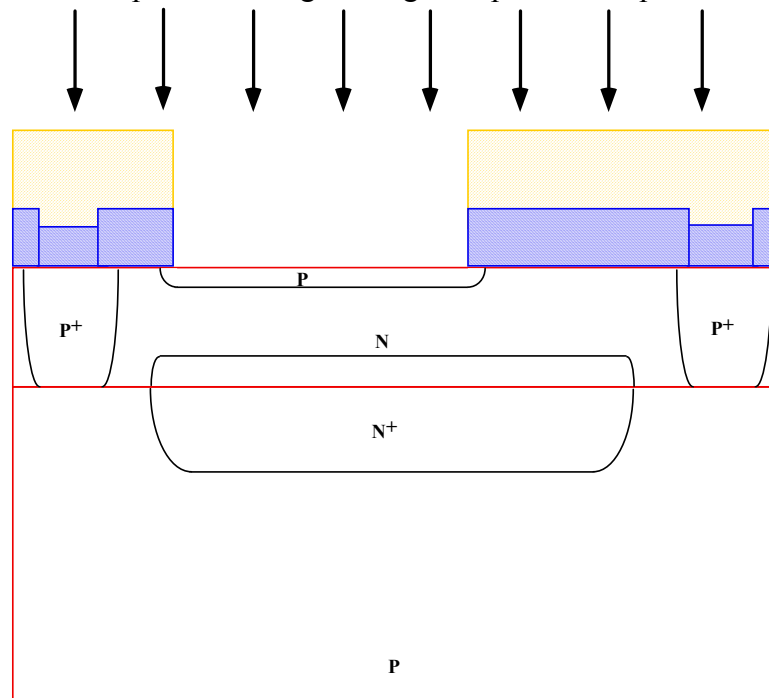
SiO₂ is thermally grown on the surface and photoresist is spun on. Mask #2 is used to define the resist and then the SiO₂ layer is etched using the resist as a mask.



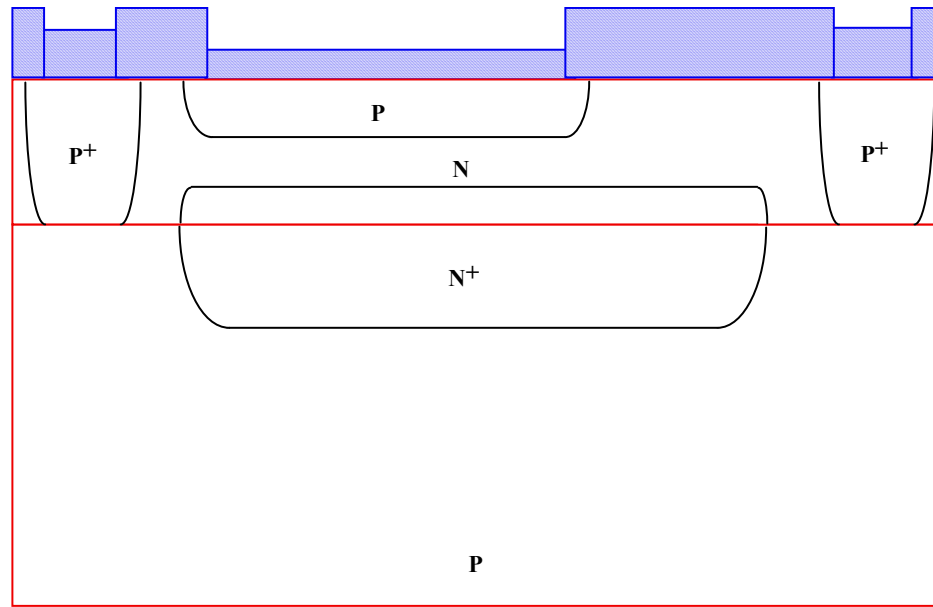
A boron implant dopes the isolation regions.



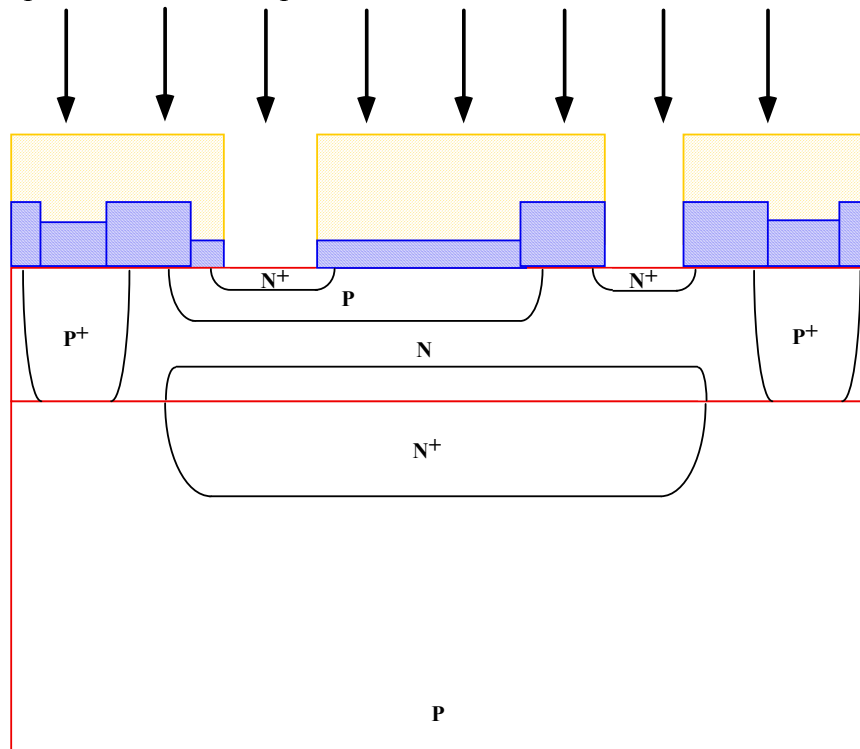
The P^+ isolation regions are driven down to the P substrate to laterally isolate the devices. SiO_2 is grown on the surface during this drive-in. Note that the buried layer continues to diffuse upwards during this high temperature step.



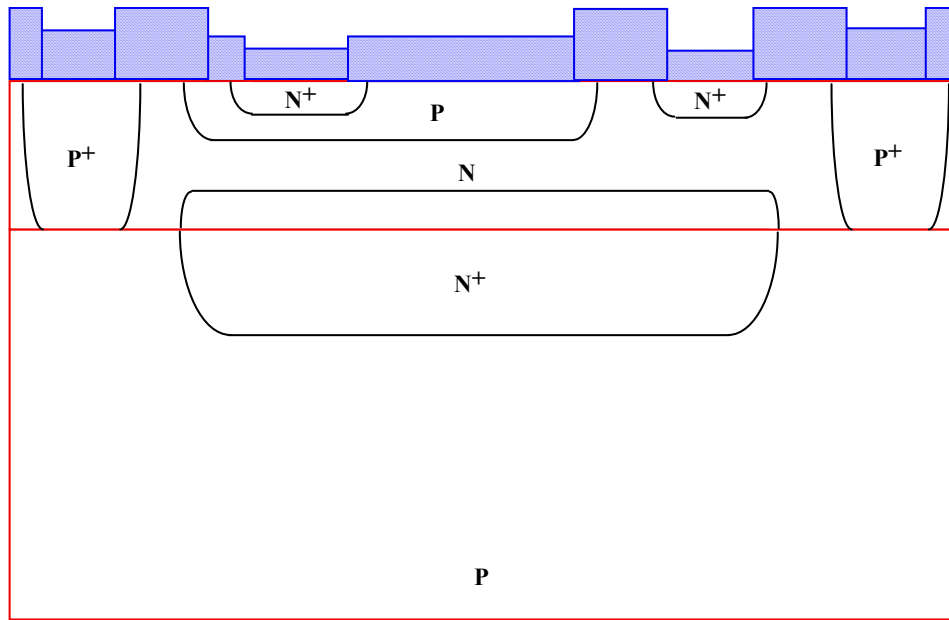
Photoresist is spun onto the wafer and mask #3 is used to define the base regions. The SiO_2 is etched and a boron implant forms the base region.



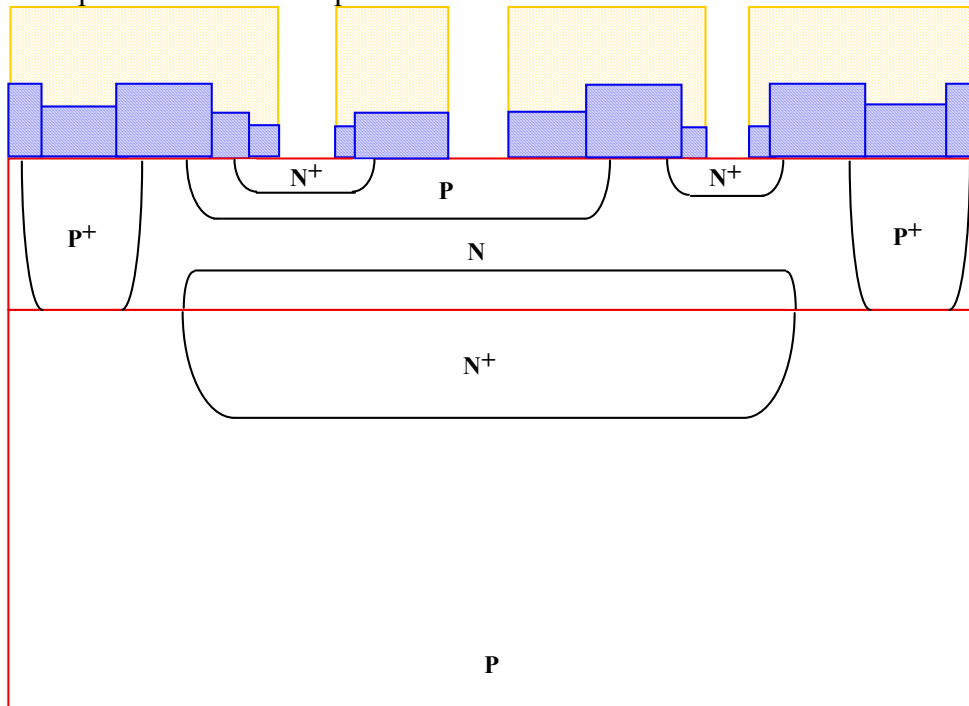
The base region is driven-in to its final junction depth. A surface SiO_2 layer is grown as part of this drive-in process.



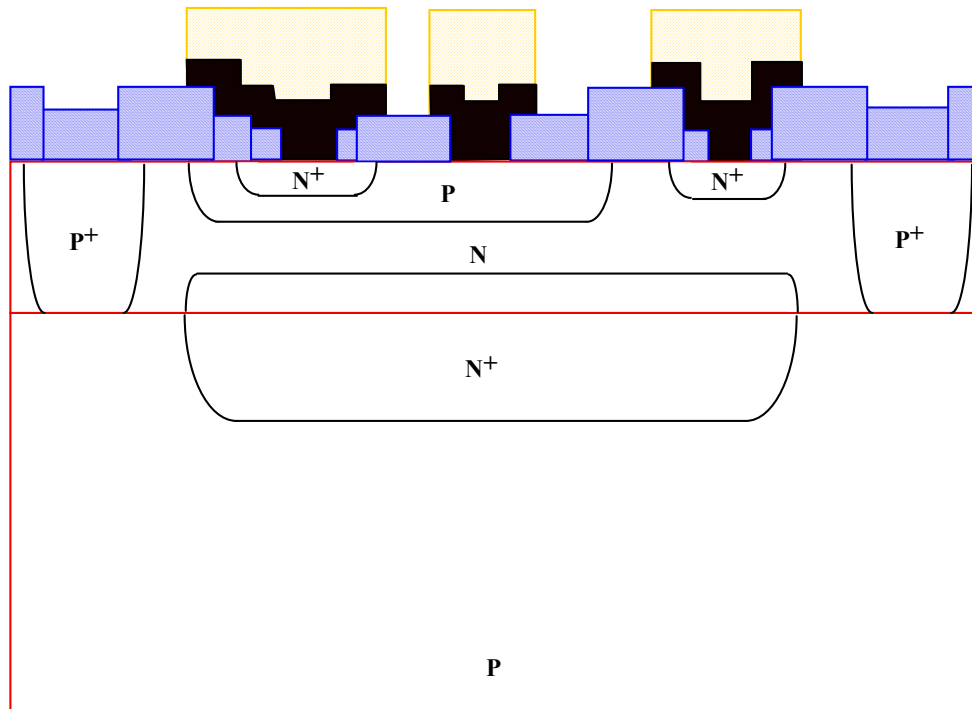
Photoresist is spun onto the wafer and mask #4 is used to define the emitter and collector contact regions. The SiO_2 is etched and an arsenic implant forms the N^+ regions.



The N^+ regions are driven-in to their final junction depth. A surface SiO_2 layer is grown as part of this drive-in process.



Photoresist is spun onto the wafer and mask #5 is used to define the contact regions. The SiO_2 is etched.



Aluminum is deposited on the wafer. Photoresist is then spun onto the wafer and mask #6 is used to define the wiring regions. The Al is then etched. Stripping the photoresist completes the overall process flow.

Chapter 3 Problems

- 3.1. Calculate the temperature difference across a (100) silicon wafer necessary for the silicon to reach its yield strength. This gradient sets an upper bound on the temperature non-uniformity that is acceptable in an RTA heating system.**

Answer:

Using Eqn. 3.1, we have $\sigma = \alpha Y \Delta T$. Thus

$$\Delta T = \frac{\sigma}{\alpha Y} = \frac{0.5 \times 10^9 \text{ dyne cm}^{-2}}{(2.3 \times 10^{-6} \text{ pC}^1)(1.3 \times 10^{12} \text{ dyne cm}^{-1})} = 167 \text{ pC cm}^{-1}$$

- 3.2. A boron-doped crystal pulled by the Czochralski technique is required to have a resistivity of 10 Ω cm when half the crystal is grown. Assuming that a 100 gm pure silicon charge is used, how much 0.01 Ω cm boron doped silicon must be added to the melt? For this crystal, plot resistivity as a function of the fraction of the melt solidified. Assume $k_0 = 0.8$ and the hole mobility $\mu_p = 550 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$.**

Answer:

Using the mobility value given, and $\rho = \frac{1}{q\mu N_A}$ we have:

$$10 \Omega \text{ cm} \Rightarrow N_A = 1.14 \times 10^{15} \text{ cm}^{-3} \text{ and } 0.01 \Omega \text{ cm} \Rightarrow N_A = 1.14 \times 10^{18} \text{ cm}^{-3}$$

From Eqn. 3.38, $C_S = C_0 k_0 (1-f)^{k_0-1}$ and we want $C_S = 1.14 \times 10^{15} \text{ cm}^{-3}$ when $f = 0.5$. Thus, solving for C_0 the initial doping concentration in the melt, we have:

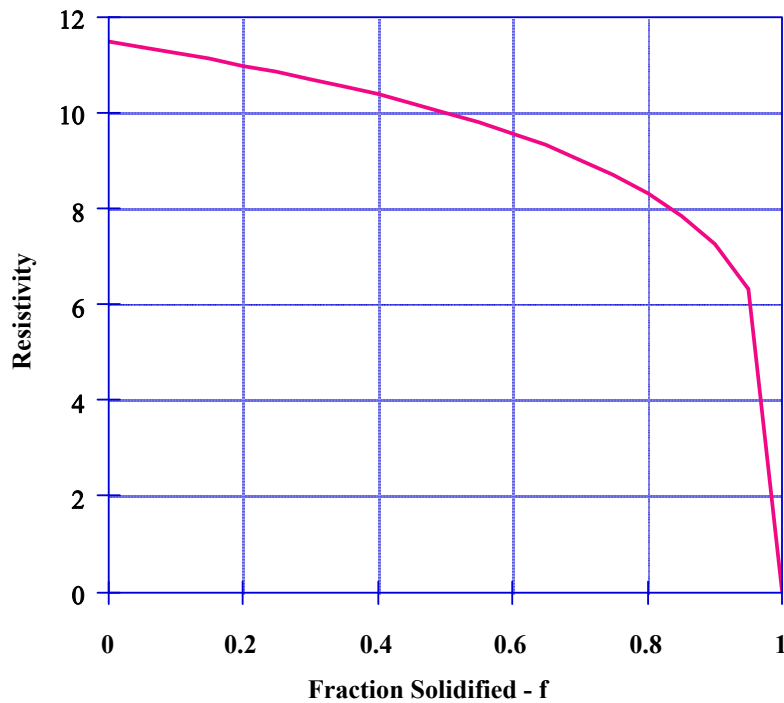
$$C_0 = \frac{1.14 \times 10^{15}}{0.8} (1-0.5)^{0.2} = 1.24 \times 10^{15} \text{ cm}^{-3}$$

$$\text{But } C_0 = \frac{I_0}{V_0} = \frac{\# \text{ of impurities}}{\text{unit vol of melt}} = \frac{(\text{Doping})(\text{Vol. of } 0.01 \Omega \text{ cm})}{\text{Vol } 100 \text{ gm Si}}$$

$$\therefore \text{Wgt added of } 0.01 \Omega \text{ cm Si} = \left(\frac{C_0}{\text{Doping}} \right) (100 \text{ gm}) = 0.109 \text{ gm}$$

The resistivity as a function of distance is plotted below and is given by

$$\rho(x) = \frac{1}{q\mu N_A(x)} = \frac{(1-f)^{1-k_0}}{q\mu C_0 k_0} = 11.5 \Omega \text{ cm} (1-f)^{0.2}$$



3.3. A Czochralski crystal is pulled from a melt containing 10^{15} cm^{-3} boron and $2 \times 10^{14} \text{ cm}^{-3}$ phosphorus. Initially the crystal will be P type but as it is pulled, more and more phosphorus will build up in the liquid because of segregation. At some point the crystal will become N type. Assuming $k_0 = 0.32$ for phosphorus and 0.8 for boron, calculate the distance along the pulled crystal at which the transition from P to N type takes place.

Answer:

We can calculate the point at which the crystal becomes N type from Eqn. 3.38 as follows:

$$C_S (\text{Phos}) = C_0 k_0 (1-f)^{k_0-1} = (2 \times 10^{14}) (0.32) (1-f)^{-0.68}$$

$$C_S (\text{Boron}) = C_0 k_0 (1-f)^{k_0-1} = (10^{15}) (0.8) (1-f)^{-0.2}$$

At the point where the cross-over occurs to N type, these two concentrations will be equal. Solving for f , we find

$$f \cong 0.995$$

Thus only the last 0.5% of the crystal is N type.

- 3.4. A Czochralski crystal is grown with an initial Sb concentration in the melt of $1 \times 10^{16} \text{ cm}^{-3}$. After 80% of the melt has been used up in pulling the crystal, pure silicon is added to return the melt to its original volume. Growth is then resumed. What will the Sb concentration be in the crystal after 50% of the new melt has been consumed by growth? Assume $k_0 = 0.02$ for Sb.**

Answer:

After the initial 80% growth, the concentration in the solid is given by Eqn. 3.38 as

$$C_S = C_O k_0 (1 - f)^{k_0 - 1} = (1 \times 10^{16}) (0.02) (1 - 0.8)^{0.02 - 1} \cong 9.68 \times 10^{14} \text{ cm}^{-3}$$

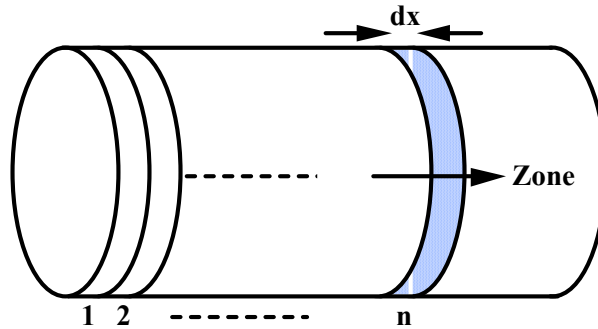
The liquid concentration at that point is given by

$$C_L = \frac{C_S}{0.02} \cong 4.84 \times 10^{16} \text{ cm}^{-3}$$

The volume is now increased by 5X, so the new $C_L \cong 9.68 \times 10^{15} \text{ cm}^{-3}$. After 50% of this new melt is consumed, the concentration in the solid will be

$$C_S = C_O k_0 (1 - f)^{k_0 - 1} = (9.68 \times 10^{15}) (0.02) (1 - 0.5)^{0.02 - 1} \cong 3.82 \times 10^{14} \text{ cm}^{-3}$$

- 3.5. Consider the zone refining process illustrated in Figures 3-7 and 3-19. Set up a simple spreadsheet to analyze this problem and use it to generate plots like those shown in Figure 3-21. Divide the crystal up into n segments as shown below. Consider a zone length dx which steps in increments dx up the crystal during refining. As the zone moves an amount dx , it incorporates an impurity concentration given by $C_S^p(n+1)$ where $n+1$ is the next zone to be melted and $C_S^p(n)$ is the impurity concentration in that zone from the previous pass. The liquid zone also leaves behind an impurity concentration given by $k_0 C_L$ where C_L is the concentration in the liquid during the current pass. Note that the impurity concentration in the liquid consists of C_S^p plus all the impurities “swept up” by the liquid during the current pass. Note that the final zone at the end simply solidifies and can be neglected.**



Answer:

We will take $k_0 = 0.1$ as is done in Fig. 3-21. In setting up the spreadsheet, we start with an initial concentration in each incremental zone of just C_0 or 1. This is column A in the spreadsheet below. On the first pass through the crystal ($i = 1$), the concentration in each liquid zone is given in column B. For example in the zone 2, the concentration is 1 from the solid liquefied in zone 2 + 0.9 from the impurity carried forward from zone 1. Thus the formula for each cell in column B is simply

$$C_L(\text{zone } n) = 1 + 0.9C_L(\text{zone } n - 1)$$

The concentration in the solid after the first pass is simply 0.1 times the C_L values given above. Thus the formula for the cells in column C is simply

$$C_S(\text{zone } n) = 0.1C_L(\text{zone } n)$$

We repeat this process for the second pass. The concentration in the liquid in each zone (column D) is given by the impurity concentration in that cell from the first pass, + 0.9 of the impurity from the preceding cell. Thus the formula for cell D4 for example is

$$D4 = C4 + 0.9D3$$

This formula increments for each cell in column D. Column E is then simply given by 0.1 times column D. For example,

$$E4 = 0.1D4$$

The result is the table below, which is also plotted below. The plot includes additional passes and more zone lengths, all calculated the same way.

	A	B	C	D	E
Zone	$C_S (i = 0)$	$C_L (i = 1)$	$C_S (i = 1)$	$C_L (i = 2)$	$C_S (i = 2)$
1	1	1.000	0.100	0.100	0.010

2	1	1.900	0.190	0.280	0.028
3	1	2.710	0.271	0.523	0.052
4	1	3.439	0.344	0.815	0.081
5	1	4.095	0.410	1.143	0.114
6	1	4.686	0.469	1.497	0.150
7	1	5.217	0.522	1.869	0.187
8	1	5.695	0.570	2.252	0.225
9	1	6.126	0.613	2.639	0.264
10	1	6.513	0.651	3.026	0.303

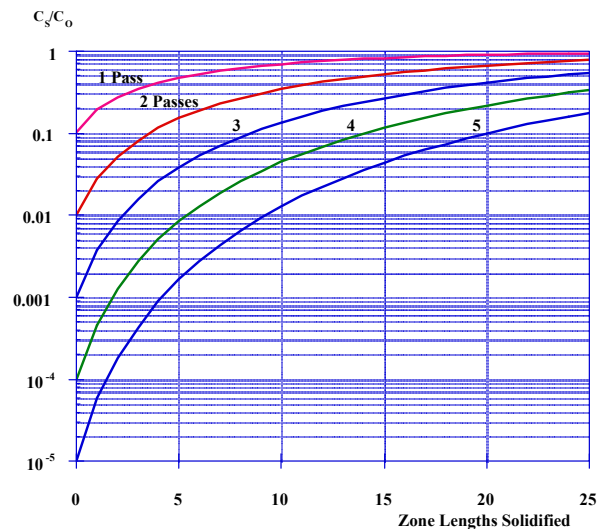


Fig. 3-21: Zone refining with multiple passes. $k_0 = 0.1$ in this example.

3.6. Suppose your company was in the business of producing silicon wafers for the semiconductor industry by the CZ growth process. Suppose you had to produce the maximum number of wafers per boule that met a fairly tight resistivity specification.

- Would you prefer to grow N type or P type crystals? Why?
- What dopant would you use in growing N-type crystals? What dopant would you use in growing P type crystals? Explain

Answer:

a). Boron has the segregation coefficient closest to unity of all the dopants. Thus it produces the most uniform doping along the length of a CZ crystal. Thus P type would be the natural choice.

b). For P type, the obvious (and only real choice) is boron as explained in part a). For N type crystals Fig. 3-18 shows that either P or As would be a reasonable choice since their segregation coefficients are quite close and are better than Sb.

Table 3-2 indicates that P might be slightly preferred over As because its k_0 value is slightly closer to 1.

Chapter 4 Problems

- 4.1. An IC manufacturing plant produces 1000 wafers per week. Assume that each wafer contains 100 die, each of which can be sold for \$50 if it works. The yield on these chips is currently running at 50%. If the yield can be increased, the incremental income is almost pure profit because all 100 chips on each wafer are manufactured whether they work or not. How much would the yield have to be increased to produce an annual profit increase of \$10,000,000?**

Answer:

At 1000 wafers per week, the plant produces 52,000 wafers per year. If each wafer has 50 good die each of which sells for \$50, the plant gross income is simply

$$\text{Income} = (52,000)(50)(\$50) = \$130,000,000 \text{ per year.}$$

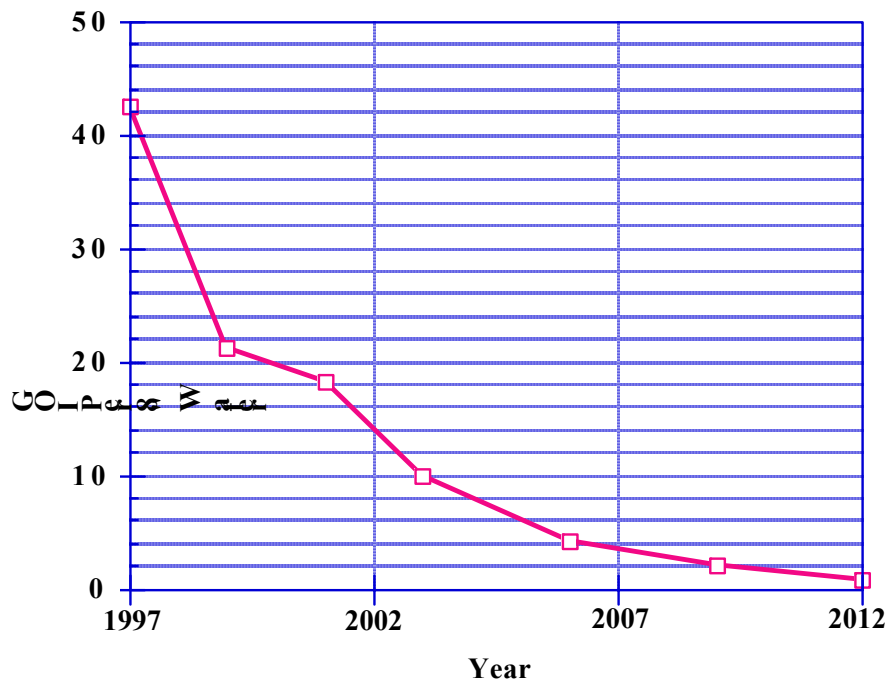
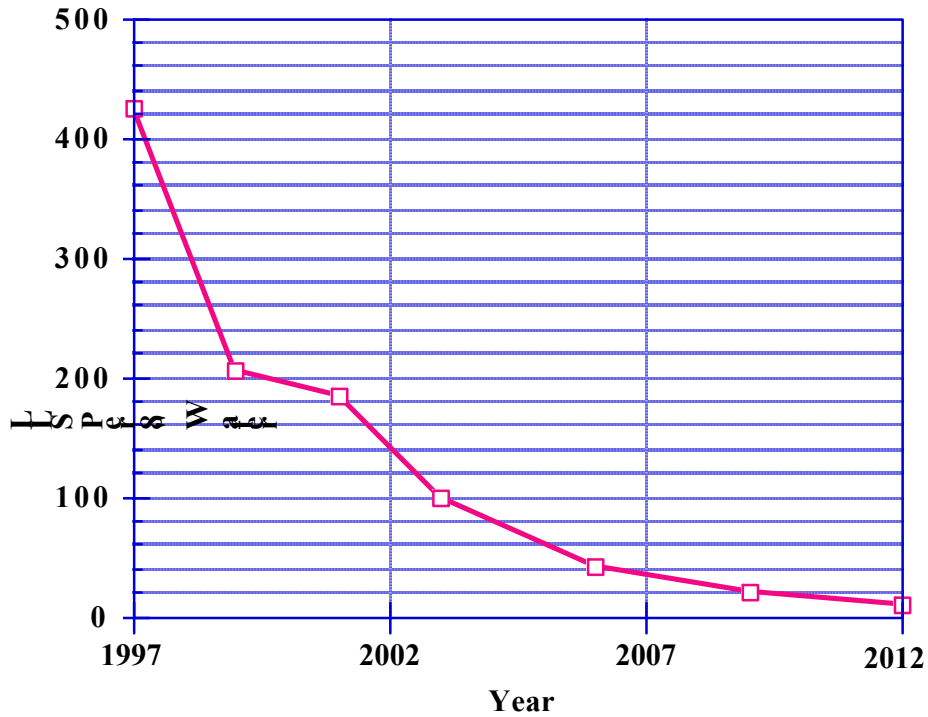
To increase this income by \$10,000,000 requires that the yield increase by

$$\frac{10}{130} \cong 7.7\%$$

- 4.2. Calculate and plot the total number of allowable LLS (localized light scatterers) on an 8" wafer, through 2012, according to the NTRS. Generate a similar plot of the total number of allowed gate oxide defects in DRAMs. What conclusions do you draw about these requirements?**

Answer:

The area of an 8" (300 mm) wafer is simply $\pi(15 \text{ cm})^2 = 706.9 \text{ cm}^2$. The NTRS requirements for LLS and for gate oxide defects are given in Table 4-1 in the text. Plotting these numbers vs. time, we find



Obviously the NTRS is predicated on constantly improving defect densities as time goes on. The numbers are actually as bit deceiving plotted this way because the NTRS also assumes a change in wafer size, from 200 mm (8") to 300 mm (12") in 1999 and from 300 mm (12") to 450 mm (15") in 2009.

- 4.3. As MOS devices are scaled to smaller dimensions, gate oxides must be reduced in thickness.**
- As the gate oxide thickness decreases, do MOS devices become more or less sensitive to sodium contamination? Explain.**
 - As the gate oxide thickness decreases, what must be done to the substrate doping (or alternatively the channel V_{TH} implant, to maintain the same V_{TH} ? Explain.**

Answer:

a). From the text, Na^+ contamination causes threshold voltage instabilities in MOS devices. Also from Eqn. 4.1, the threshold voltage is given by

$$V_{TH} = V_{FB} + 2\phi_f + \frac{\sqrt{2\varepsilon_S q N_A (2\phi_f)}}{C_{OX}} + \frac{qQ_M}{C_{OX}}$$

As the gate oxide thickness decreases, C_{OX} increases, so the same amount of mobile charge Q_M will have less effect on V_{TH} as oxides get thinner. Therefore MOS devices are less sensitive to sodium contamination.

b). Using the same expression for V_{TH} as in part a), we observe that as the oxide thickness decreases, (C_{OX} increases), to maintain the same V_{TH} , N_A will have to increase. N_A will actually have to increase by the square root of the oxide thickness decrease to keep V_{TH} constant.

- 4.4. A new cleaning procedure has been proposed which is based on H_2O saturated with O_2 as an oxidant. This has been suggested as a replacement for the H_2O_2 oxidizing solution used in the RCA clean. Suppose a Si wafer, contaminated with trace amounts of Au, Fe and Cu is cleaned in the new H_2O/O_2 solution. Will this clean the wafer effectively? Why or why not? Explain.**

Answer:

As described in the text, cleaning metal ions off of silicon wafers involves the following chemistry:



The cleaning solution must be chosen so that the reaction is driven to the right because this puts the metal ions in solution where they can be rinsed off. Since driving the reaction to the right corresponds to oxidation, we need an oxidizing solution to clean the wafer.

$\text{H}_2\text{O}/\text{O}_2$ is certainly an oxidizing solution. But whether it cleans effectively or not depends on the standard oxidation potential of the various possible reactions. From Table 4-3 in the text, we have:

Oxidant/ Reductant	Standard Oxidation Potential (volts)	Oxidation-Reduction Reaction
SiO_2/Si	0.84	$\text{Si} + 2\text{H}_2\text{O} \leftrightarrow \text{SiO}_2 + 4\text{H}^+ + 4\text{e}^-$
Fe^{3+}/Fe	0.17	$\text{Fe} \leftrightarrow \text{Fe}^{3+} + 3\text{e}^-$
Cu^{2+}/Cu	-0.34	$\text{Cu} \leftrightarrow \text{Cu}^{2+} + 2\text{e}^-$
$\text{O}_2/\text{H}_2\text{O}$	-1.23	$2\text{H}_2\text{O} \leftrightarrow \text{O}_2 + 4\text{H}^+ + 2\text{e}^-$
Au^{3+}/Au	-1.42	$\text{Au} \leftrightarrow \text{Au}^{3+} + 3\text{e}^-$

The stronger reactions (dominating) are at the bottom.

Thus the $\text{H}_2\text{O}/\text{O}_2$ reaction will clean Fe and Cu, but it will not clean Au off the wafer.

4.5. Explain why it is important that the generation lifetime measurement illustrated in Figure 4-19 is done in the dark.

Answer:

The measurement depends on measuring carriers generated thermally in the silicon substrate (or at the surface). If light is shining on the sample, then absorbed photons can also generate the required carriers. As a result, the extracted generation lifetime with the light on would really be measuring the intensity of the incident light and not a basic property of the silicon material.

4.6. If you wanted to look for trace levels of contaminants left on a wafer surface after cleaning, would XRF or XPS be a better choice? Explain.

Answer:

In each of these techniques, an x-ray is used to provide the input energy. The difference is in the emitted species. In XRF an x-ray is emitted and in XPS an electron is emitted. The key difference is that the escape depth is much larger for x-rays, so if we wanted information about the near surface region, XPS would be a much better choice.

4.7. In the discussion in the text related to Figure 4-12 the importance of the temperature ramp rate between the nucleation and precipitation stages is

described. Is the ramp rate coming back down to room temperature after the precipitation stage important in terms of gettering? Explain.

Answer:

The temperature ramp coming back down to room temperature should not be very critical in terms of its impact on gettering. As the temperature ramps down, precipitation will slow down and stop, but the embryos will already have grown to their final size during the high temperature precipitation step anyway. Once the temperature starts to drop, they will quickly stop growing, but they will be stable at lower temperatures, so there is not a major concern about how fast the temperature ramps down.

Note that there may be other issues that do affect how fast the temperature is ramped down. Generally a fairly slow ramp is used to make certain that large temperature gradients do not develop across the wafers. Such gradients could generate defects in the wafers.

Chapter 5 Problems

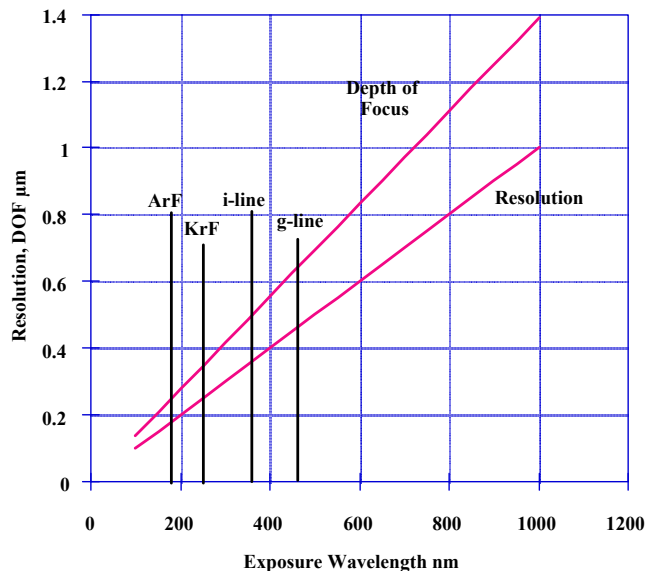
- 5.1. Calculate and plot versus exposure wavelength the theoretical resolution and depth of focus for a projection exposure system with a NA of 0.6 (about the best that can be done today). Assume $k_1 = 0.6$ and $k_2 = 0.5$ (both typical values). Consider wavelengths between 100 nm and 1000 nm (DUV and visible light). Indicate the common exposure wavelengths being used or considered today on your plot (g-line, i-line, KrF and ArF). Will an ArF source be adequate for the 0.13 μm and 0.1 μm technology generations according to these simple calculations?

Answer:

The relevant equations are simply

$$\therefore R = k_1 \frac{\lambda}{\text{NA}} = 0.6 \frac{\lambda}{0.6} \quad \text{and} \quad \text{DOF} = \pm k_2 \frac{\lambda}{(\text{NA})^2} = \pm 0.5 \frac{\lambda}{(0.6)^2}$$

These equations are plotted below. Note that the ArF (193 nm) will not reach 0.13 μm or 0.1 μm resolution according to these simple calculations. In fact, with more sophisticated techniques such as phase shift masks, off axis illumination etc., ArF is expected to reach 0.13 μm and perhaps the 0.1 μm generations.



- 5.2. In a particular positive resist process it is sometimes noticed that there is difficulty developing away the last few hundred angstroms of resist in exposed areas. This sometimes causes etching problems because the resist remains in the areas to be etched. Suggest a possible cause of this problem and therefore propose a solution.

Answer:

The most likely cause is that there are standing waves in the resist during the exposure and a minimum of exposure intensity exists in the bottom region of the resist. Thus the resist is not fully exposed and won't clear properly during the develop cycle. The problem could be solved by changing a number of parameters like the resist thickness, the thickness of underlying layers, or by increasing the exposure time. Alternatively, an ARC layer could be added under the resist to minimize standing wave and reflection effects.

Alternative answers are certainly possible. Perhaps the exposure time was too short and because of bleaching effects, the bottom part of the resist was not fully absorbed. Or perhaps the developing time was too short and hence the resist was not developed out at the bottom.

5.3. An X-ray exposure system uses photons with an energy of 1 keV. If the separation between the mask and wafer is 20 μm , estimate the diffraction limited resolution that is achievable by this system.

Answer:

The equivalent wavelength of 1 keV x-rays is given by

$$E = h\nu = \frac{hc}{\lambda} \quad \therefore \quad \lambda = \frac{hc}{E} = \frac{(4.14 \times 10^{-15} \text{ eV sec})(3 \times 10^{10} \text{ cm sec}^{-1})}{10^3 \text{ eV}}$$

$$= 1.24 \times 10^{-7} \text{ cm} = 1.24 \text{ nm}$$

X-ray systems operate in the proximity printing mode, so that the theoretical resolution is given by Eqn. 5.12:

$$\text{Resolution} = \sqrt{\lambda g} = \sqrt{(1.24 \times 10^{-3} \mu\text{m})(20 \mu\text{m})} = 0.15 \mu\text{m}$$

5.4. Estimate the exposure wavelength that was used in the simulation example in Figure 5.44 in the text. Assume that the index of refraction of the photoresist is 1.68 (typical value).

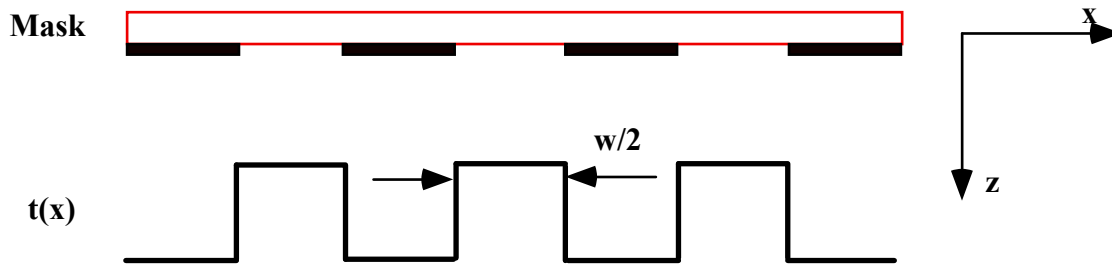
Answer:

The standing waves on the left part of the exposed region have a period of $\approx 0.12 \mu\text{m}$, estimated from Fig. 5.44. From the text, standing waves have a period of $\lambda/2$ where λ is the wavelength of the light in the resist. The wavelength is thus $\approx 0.24 \mu\text{m}$ in the resist. This corresponds to a wavelength in air of $\approx 1.68(0.24 \mu\text{m})$ or

$$\text{Exposure Wavelength} \approx 0.40 \mu\text{m}$$

(Actually an i-line stepper was assumed in the simulation with $\lambda = 0.365\mu\text{m}$.)

- 5.5. In this chapter, we considered an example of an isolated space on a mask and saw that the Fourier transform of this pattern is the $\sin(x)/x$ function. This function describes the spatial variation of the light intensity that the objective lens must capture. An even more illustrative example of the effects of diffraction is given by a mask pattern consisting of a periodic pattern of lines and spaces as shown below.



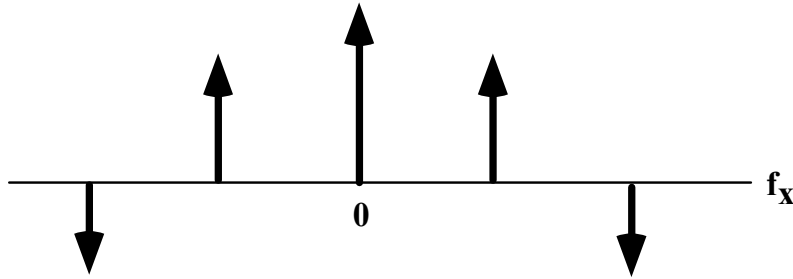
Find the Fourier transform for this mask function and plot it as was done in in the example in this chapter. The objective lens must capture at least the first diffraction order if it is going to resolve features with a pitch p . Use this criteria to derive a result like Eqn. 5.2 for the resolution of such a system.

Answer:

The Fourier transform is found in standard texts or references. For this case, the diffraction pattern is a series of discrete diffraction orders given by

$$F\{t(x)\} = \sum_{n=-\infty}^{n=\infty} \frac{\sin\left(\pi \frac{w}{2} f_x\right)}{\pi f_x} \delta\left(f_x - \frac{n}{w}\right)$$

This is plotted below. The central maximum occurs at a spatial frequency of 0 (passes straight through the objective lens ($f_x = 0$), and can be thought of as the “DC” component of the light. It provides power information, but no spatial information. The first order diffraction pattern occurs at a spatial frequency of $\pm 1/w$. Since this (and higher diffraction orders) occur at a spatial frequency related to the mask spacing w , these diffraction orders carry the spatial information.



The first order diffraction pattern occurs at $f_x = 1/w$. Using Eqn. 9.27 in the text, we have,

$$f_x = \frac{1}{w} = \frac{NA}{\lambda}$$

The minimum feature size or resolution is $w/2$ and thus we have finally that

$$\text{Resolution} = 0.5 \frac{\lambda}{NA}$$

5.6. Assume that Fig. 5-46 in the text was experimentally determined for a $0.6 \mu\text{m}$ thick resist with an index of refraction of 1.68. Estimate the Dill resist parameters from the data in this figure.

Answer:

The Dill resist parameters are defined in the text as follows:

$$A = \frac{1}{D} \ln \left(\frac{T_\infty}{T_0} \right)$$

$$B = -\frac{1}{D} \ln(T_\infty)$$

$$C = \frac{A + B}{AT_0(1 - T_0)T_{12}} \left. \frac{dT}{dE} \right|_{E=0}$$

where

$$T_{12} = 1 - \left(\frac{n_{\text{resist}} - 1}{n_{\text{resist}} + 1} \right)^2$$

From Fig. 5-46, we have that $T_0 \approx 0.375$, $T_\infty \approx 0.9$ and $D = 0.6 \mu\text{m}$. Thus we have

$$A = \frac{1}{D} \ln\left(\frac{T_\infty}{T_0}\right) = \frac{1}{0.6} \ln\left(\frac{0.9}{0.375}\right) = 1.46 \mu\text{m}^{-1}$$

$$B = -\frac{1}{D} \ln(T_\infty) = -\frac{1}{0.6} \ln(0.9) = 0.176 \mu\text{m}^{-1}$$

$$T_{12} = 1 - \left(\frac{n_{\text{resist}} - 1}{n_{\text{resist}} + 1}\right)^2 = 1 - \left(\frac{1.68 - 1}{1.68 + 1}\right)^2 = 0.935$$

$$C = \frac{A + B}{AT_0(1 - T_0)T_{12}} \frac{dT}{dE}\bigg|_{E=0} = \frac{1.46 + 0.176}{(1.46)(0.375)(0.625)(0.935)} \frac{0.9 - 0.375}{250} = 0.011 \mu\text{m}^{-1}$$

- 5.7. Lithography often has to be done over underlying topography on a silicon chip. This can result in variations in the resist thickness as the underlying topography goes up and down. This can sometimes cause some parts of the photoresist image to be underexposed and/or other regions to be overexposed. Explain in terms of the chemistry of the resist exposure process why these underexposure and overexposure problems occur.**

Answer:

Photoresist exposure occurs with a bleaching process. What this means is that the photons are initially absorbed by the PAC in the top layers of the resist and that is therefore where the initial exposure occurs. As the PAC is reacted in these top layers, the resist becomes more transparent (bleaches) and the light then penetrates to the deeper portions of the resist to expose those regions. The time required for complete exposure all the way to the bottom of the resist will thus depend on the resist thickness. Thus if there are thickness variations, there will be problems with overexposure in the thin regions if the exposure is long enough to expose the thick regions. Alternatively, there will be underexposure problems in the thick regions if the exposure is only long enough to expose the resist in the thin regions.

- 5.8. As described in this chapter, there are no clear choices for lithography systems beyond optical projection tools based on 193-nm ArF excimer lasers. One possibility is an optical projection system using a 157-nm F₂ excimer laser.**
- a. Assuming a numerical aperture of 0.8 and $k_1 = 0.75$, what is the expected resolution of such a system using a first order estimate of resolution?**

b. Actual projections for such systems suggest that they might be capable of resolving features suitable for the 2009 0.07 μm generation. Suggest three approaches to actually achieving this resolution with these systems.

Answer

a). The simple formula for resolution is

$$R = k_1 \frac{\lambda}{\text{NA}} = 0.75 \frac{0.157 \mu\text{m}}{0.8} = 0.147 \mu\text{m}$$

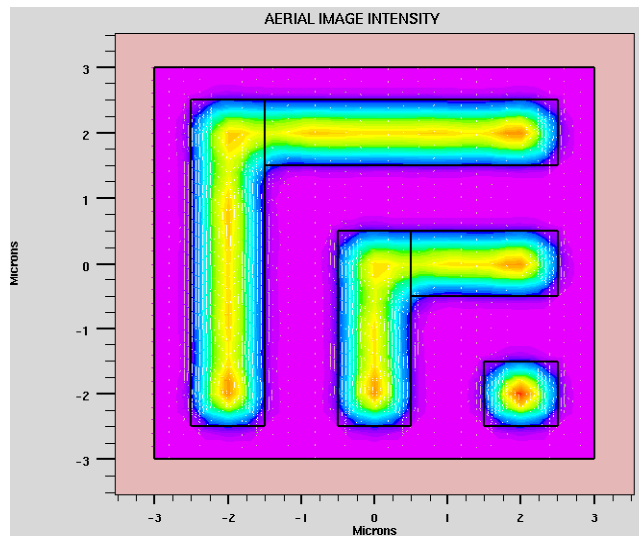
b). The calculated resolution in part a is a factor of two larger than required for the 0.07 μm generation. Therefore some “tricks” will have to be used to actually achieve such resolution. There are a number of possibilities:

1. Use of phase-shift masks. This technique, discussed in this chapter, has the potential for significant resolution improvements. It works by designing a more sophisticated mask. Simple masks are digital - black or white. Phase shifting adds a second material to the mask features, usually at the edges which shifts the optical phase and sharpens up the aerial image. Sophisticated computer programs are required to design such masks.

2. Use of optical proximity correction in the mask design. This is another approach to designing a better mask and as discussed in class, can also improve resolution significantly. The approach involves adding extra features to the mask, usually at corners where features are sharp, to compensate for the high frequency information lost to diffraction effects.

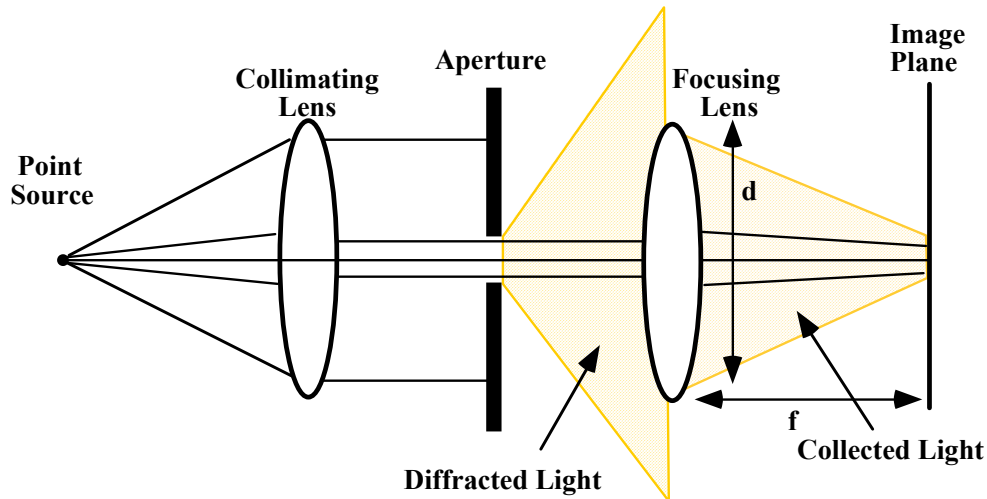
3. Off-axis illumination. This allows the optical system to capture some of the higher order diffracted light and hence can improve resolution.

- 5.9. Current optical projection lithography tools produce diffraction limited aerial images. A typical aerial image produced by such a system is shown in the simulation below where a square and rectangular mask regions produce the image shown. (The mask features are the black outlines, the calculated aerial image is the grayscale inside the black rectangles.) The major feature of the aerial image is its rounded corners compared to the sharp square corners of the desired pattern. Explain physically why these features look the way they do, using diffraction theory and the physical properties of modern projection optical lithography tools.



Answer:

Modern optical projection lithography systems are limited in the resolution they can achieve by diffraction effects. The finite size of the focusing lens means that the high order diffraction components are “lost” and are therefore not available to help in printing a replica of the mask image. But the high frequency spatial components are exactly the components that contain information about “sharp” features, i.e. corners etc. Thus the projected aerial image loses this information and corners become rounded. The only ways to improve the image are by using shorter wavelength light, or a higher NA lens.



5.10. Future optical lithography systems will likely use shorter exposure wavelengths to achieve higher resolution and they will also likely use planarization techniques to provide “flat” substrates on which to expose the resist layers. Explain why “flat” substrates will be more important in the future than they have been in the past.

Answer:

As the wavelength of the exposure system decreases, the depth of focus of the exposure system also decreases. Thus it will be necessary to make sure that the resist in which the image is to be exposed, is flat and does not require much depth of focus. Planarization techniques will be required to accomplish this. This could mean CMP to planarize the substrate before the resist is applied, or it could mean using a spun on resist which planarizes the substrate and which is then covered with a thin, uniform imaging resist layer.

Chapter 6 Problems

- 6.1. A spherically shaped piece of silicon is cut and polished from a Czochralski single crystal ingot and oxidized in a thermal oxidation furnace. Upon pulling the silicon sphere from the furnace, the color is observed to vary significantly over the surface. Why?**

Answer:

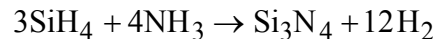
A sphere of silicon has all possible planes exposed to the oxidizing ambient. The faster oxidizing planes (111) will grow a thicker oxide than the lower index planes. The different oxide thickness around the sphere will cause the color to vary.

- 6.2. Using Eqns. 6.1 and 6.2, explain why oxides thinner than about 50 nm do not have any characteristic color associated with them. You do not need to do any calculations. Simply explain the physical reason for the lack of color in thin oxides.**

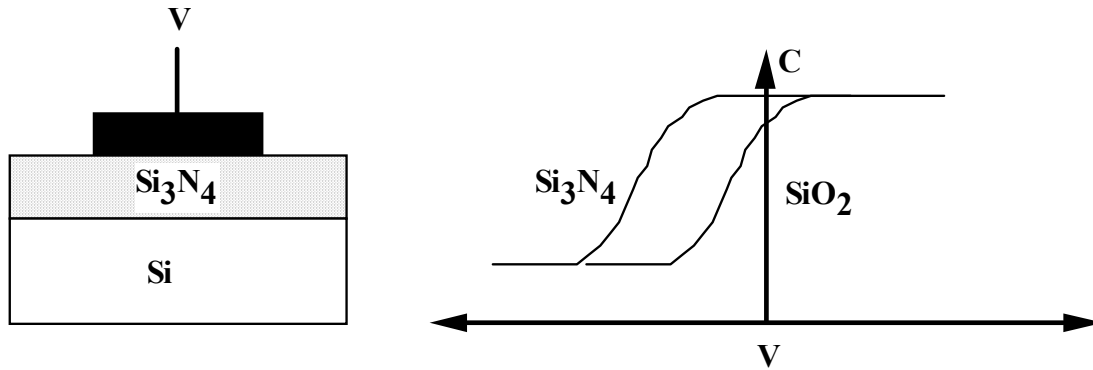
Answer:

The color charts correspond to white light being incident vertically on the SiO₂/Si structure. Light is reflected at both the air/SiO₂ interface and at the Si/SiO₂ interface. Destructive interference can occur between these two reflected waves, if the oxide thickness is a specific fraction of a wavelength. The color arises because when this destructive interference occurs, the incident white light loses some of its color components. For very thin oxides, the light wavelength that corresponds to destructive interference is outside the visible range (λ is too short).

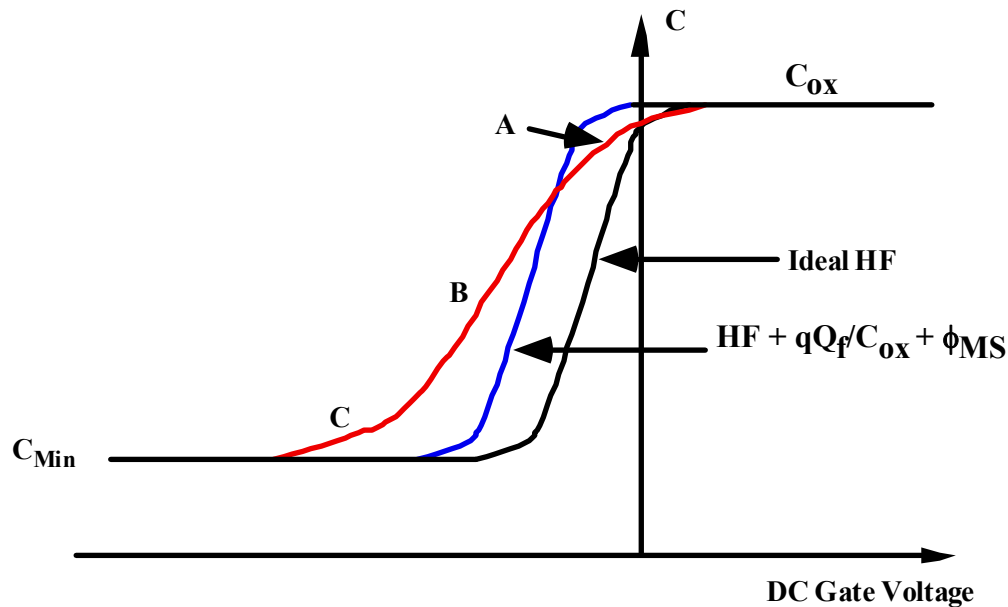
- 6.3. An experimental MIS (I=insulator) structure is fabricated by depositing Si₃N₄ (silicon nitride) on a silicon substrate. The nitride is deposited by directing a jet of silane and ammonia at the surface**



A metal electrode is deposited and a C-V plot is made as shown below. A representative C-V plot is also shown for an identical structure except with thermally grown SiO₂ as the insulator. Explain the lateral shift in the C-V curve of the Si₃N₄.



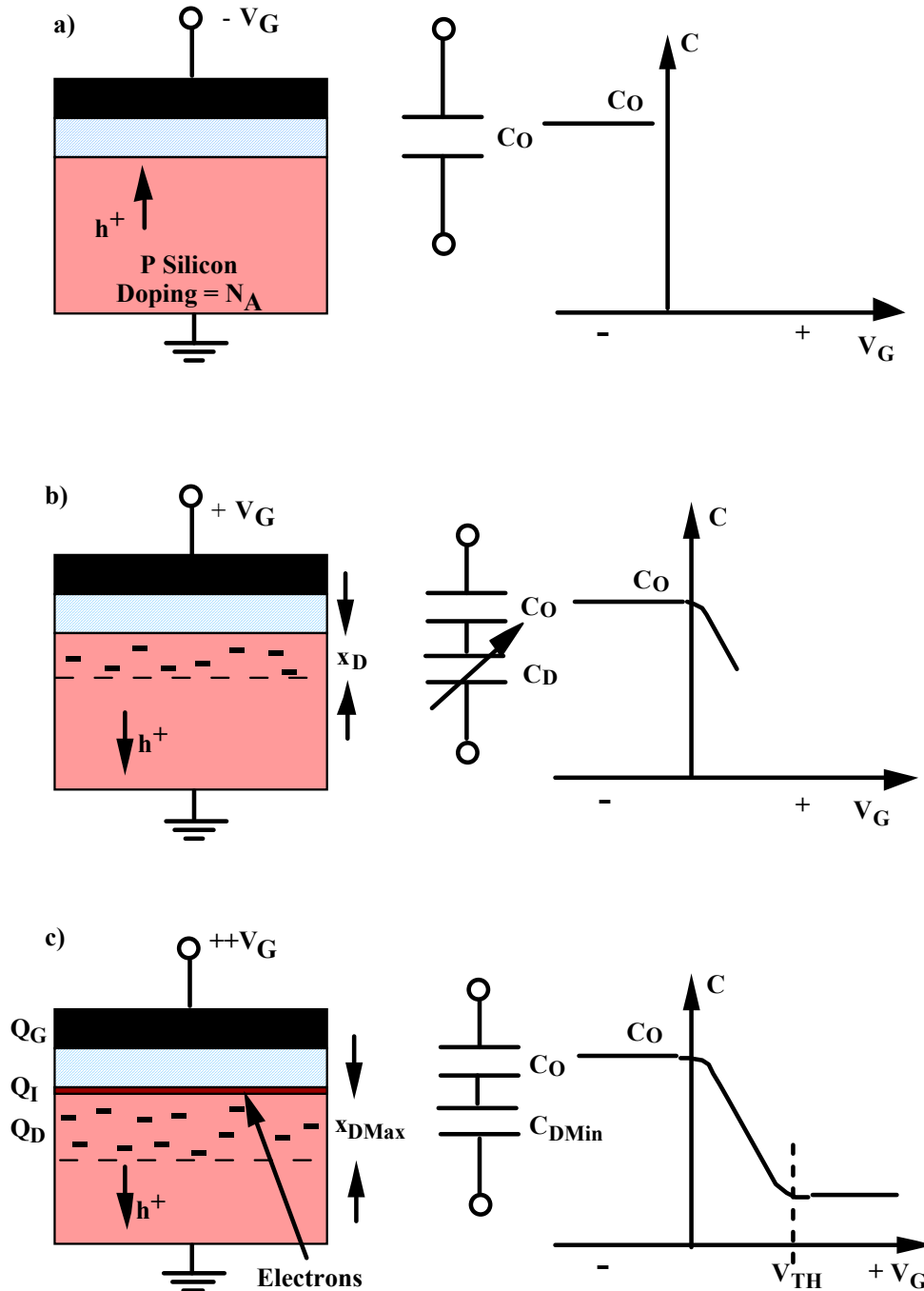
Answer:



The plot above (from Fig. 6.13) looks like the difference between the oxide C-V curve (ideal HF) and the nitride C-V curve (with interface charges). Thus, it is likely that the deposited nitride layer has unterminated bonds at the interface that give rise to fixed interface charges, shifting the C-V curve as shown.

- 6.4. Construct a HF CV plot for a P-type silicon sample, analogous to Fig. 6-9. Explain your plot based on the behavior of holes and electrons in the semiconductor in a similar manner to the discussion in the text for Fig. 6.9.

Answer:

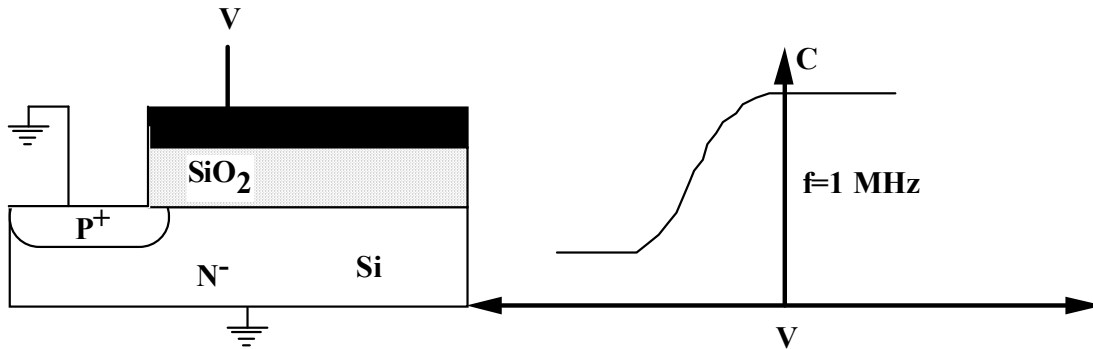


The C-V plot looks basically the same as the N substrate example in the text, that we discussed in class, except that the horizontal axis is flipped. For negative applied gate voltages, the majority carrier holes in the substrate are attracted to the surface. This is the accumulation region a) above. We measure just C_{OX} for the capacitance since there is no depletion in the substrate. For $+V_G$, the holes are driven away from the surface creating first a depletion region as in b) and finally an inversion

layer of electrons as in c). The measured capacitance drops as we move into depletion and finally reaches a minimum value after an inversion layer forms.

The C-V curves shown are high frequency curves. As discussed in the text, the capacitance remains at its minimum value for $+V_G$ values greater than V_{TH} because the inversion layer electrons cannot be created or destroyed as fast as the signal is changing. Hence the small AC signal must “wiggle” the bottom of the depletion region to balance ΔV_G .

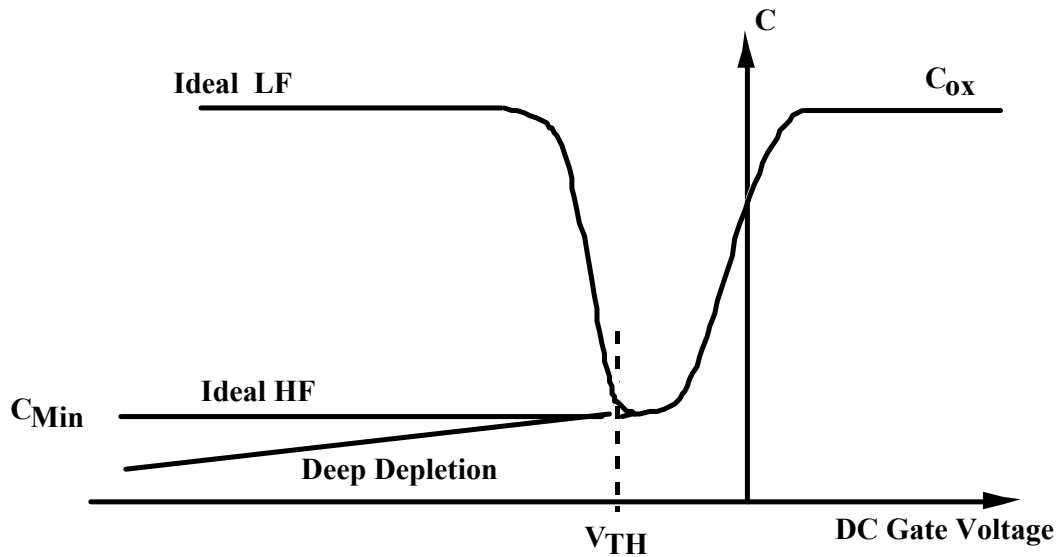
- 6.5. A MOS structure is fabricated to make C-V measurements as shown below. The C-V plot shows the result if the P^+ diffusion is NOT present. Sketch the expected shape of the C-V plot with the P^+ diffusion. Explain.



Answer:

The C-V curve shown looks just like the ideal high frequency curve from Fig. 6.11. Without the P^+ diffusion, C reaches a minimum under inversion conditions because the minority carriers in the inversion layer cannot recombine or be generated at a rate fast enough to follow the 1MHz signal. Therefore, ΔQ_G must be balanced by a change in the depletion layer charge i.e. the depletion region must widen and shrink with the AC signal.

The addition of the P^+ diffusion will allow carriers to be supplied rapidly to the inversion layer that forms on the N-type substrate at high negative voltages, even though the AC frequency is high. Because the inversion layer charge Q_I follows the gate charge Q_G , the capacitance measured in the structure will be just the oxide capacitance because the top and bottom plates of the capacitor where the charge is changing are on the two edges of the oxide. Thus, the curve will look exactly like the low frequency curve (“ideal LF”), where the generation and recombination processes can keep up with the AC signal and the inversion layer charge can follow the changes in the gate charge.



- 6.6. In a small MOS device, there may be a statistical variation in V_T due to differences in Q_F from one device to another. In a $0.13\ \mu\text{m}$ technology minimum device (gate oxide area = $0.1\ \mu\text{m} \times 0.1\ \mu\text{m}$) with a 2.5nm gate oxide, what would the difference in threshold voltage be for devices with 0 or 1 fixed charge in the gate oxide?

Answer:

The oxide capacitance is

$$C_{\text{ox}} = \frac{\epsilon A}{d} = \frac{(3.9 \times 8.854 \times 10^{-14}) (0.1 \times 10^{-4}) (0.1 \times 10^{-4})}{2.5 \times 10^{-7}} = 1.38 \times 10^{-16}$$

The change in threshold voltage is given by

$$\Delta V_T = \frac{qQ_F}{C_{\text{ox}}} = \frac{(1.6 \times 10^{-19})(1)}{1.38 \times 10^{-16}} = 1.1\ \text{mV}$$

This shows that a single electron trap in a gate oxide will have a negligible effect on the threshold voltage at this technology generation.

- 6.7. Why is steam oxidation more rapid than dry O_2 oxidation?

Answer:

The effective diffusivities of both O_2 and H_2O are on the same order (about $5 \times 10^3\ \mu\text{m}^2\ \text{hr}^{-1}$ at 1100°C).

At 1100°C, typical values for C^* are $\approx 5 \times 10^{16} \text{ cm}^{-3}$ for dry O_2 and $\approx 3 \times 10^{19} \text{ cm}^{-3}$ for H_2O . As a result, both rate constants, B and B/A are much larger for H_2O than for O_2 .

Thus, the main reason that steam oxidation is faster than dry oxidation is because the solubility of the oxidant species is higher by three orders of magnitude.

6.8. Under what conditions is the thermal growth rate of SiO_2 linearly proportional to time?

Answer:

The oxide growth is in the linear regime for small values of the oxide thickness, or more precisely, when $k_S x_O / D \ll 1$. The thickness at which $k_S x_O / D \approx 1$ varies with temperature since both k_S and D change with temperature but is generally in the range of 50-200nm. The oxide growth will stay in the linear regime for a longer time at low temperatures and for dry rather than wet oxidations.

6.9. According to the Deal Grove model, oxidation kinetics start out linear and become parabolic as the oxidation proceeds. Calculate the oxide thickness at which this transition takes place and plot this versus oxidation temperature.

Answer:

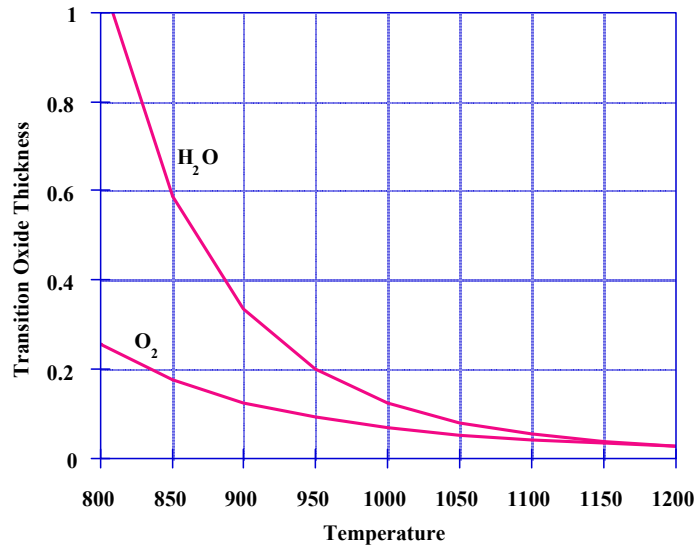
The transition from linear to parabolic growth is a gradual one and therefore there is no specific thickness at which an abrupt transition takes place. The text “defines” the transition (pg. 315), as occurring when

$$\frac{k_S x_o}{D} \approx 1 \quad \text{or when} \quad x_o = \frac{D}{k_S} = \frac{B}{2(B/A)}$$

Taking the values for dry O_2 as an example, from Table 6-2, we have

$$x_o = \frac{B}{2(B/A)} = \frac{7.72 \times 10^2 \exp\left(-\frac{1.23 \text{ eV}}{kT}\right)}{(2)6.23 \times 10^6 \exp\left(-\frac{2 \text{ eV}}{kT}\right)} = 0.619 \times 10^{-4} \exp\left(\frac{0.77 \text{ eV}}{kT}\right)$$

This equation is plotted below along with the H_2O result also using values from Table 6-2. The values in Table 6-2 apply to (111) silicon as does the graph below. (100) curves are easily generated by dividing the respective B/A values by 1.68.



6.10. Does the oxide thickness at which there is a transition from linear to parabolic rates change if we perform an oxidation at a pressure of 20 atmospheres rather than at 1 atmosphere.

Answer:

The transition from linear to parabolic rates occurs when

$$\frac{k_S x_0}{D} \approx 1 \quad \text{or when} \quad x_0 = \frac{D}{k_S} = \frac{B}{2(B/A)}$$

In the standard Deal-Grove model, both B and B/A are linearly proportional to pressure. This is a consequence of assuming that Henry's law holds during the oxidant absorption into the oxide and that only a molecular species is involved in the transport through the oxide and reaction at the Si/SiO_2 interface.

Thus, a change in pressure will not affect the thickness at which the transition occurs.

6.11. A MOS device requires a gate oxide of $10\text{nm} \pm 0.5\text{nm}$. Assume the growth is done at 900°C in dry O_2 . Neglect any effect of the anomalous initial growth. Derive a simple expression which gives the sensitivity of the oxide thickness to growth temperature (dx/dT). Evaluate this expression to see how well controlled the furnace T must be in order to obtain $10\text{nm} \pm 0.5\text{nm}$ at 900°C .

Answer:

We are asked for a simple expression for the temperature sensitivity. From Problem 6.9, we see that the transition thickness at 900°C for dry oxygen is greater than 10nm, so we can use the linear growth rate approximation.

At 900°C,

$$kT = 0.101\text{eV}$$

$$kT^2 = 118.605$$

$$B = 7.72 \times 10^2 \exp\left(-\frac{1.23}{kT}\right) = 0.004$$

$$\frac{B}{A} = 6.23 \times 10^6 \exp\left(-\frac{2.0}{kT}\right) = 0.016$$

$$A = 0.25$$

The maximum growth rate at 900°C is

$$\frac{B}{A} = 0.016$$

The growth rate at 900°C after growing 100nm of oxide is

$$\frac{dx}{dt} = \frac{B}{2x + A} = 0.015$$

Thus, the growth rate is approximately constant and we are justified in using the linear approximation:

$$x = \left(\frac{B}{A}\right)t = C_2 \exp\left(-\frac{E_2}{kT}\right)t$$

$$\frac{dx}{dT} = C_2 \exp\left(-\frac{E_2}{kT}\right) \left(-\frac{E_2}{k}\right) \left(-\frac{1}{T^2}\right) (t)$$

$$\frac{dx}{dT} = \left(\frac{B}{A}\right) \left(\frac{E_2}{kT^2}\right) (t)$$

Continuing to use the linear approximation, we have

$$x = \left(\frac{B}{A}\right)t$$

so that

$$\Delta T = \left(\frac{\Delta x}{x} \right) \left(\frac{kT^2}{E_2} \right) = \left(\frac{5}{100} \right) \left(\frac{8.62 \times 10^{-5} (900 + 273)^2}{2.0} \right)$$

$$\Delta T = 2.96 \text{ K}$$

6.12 A silicon wafer is covered by an SiO₂ film 0.3 μm thick.

a. What is the time required to increase the thickness by 0.5 μm by oxidation in H₂O at 1200°C?

b. Repeat for oxidation in dry O₂ at 1200°C.

Answer:

We will perform the calculation for <111> silicon wafers. For <100> wafers, the linear rate constant should be divided by 1.68.

a. At 1200°C, in H₂O

$$B = 3.86 \times 10^2 \exp\left(-\frac{0.78}{kT}\right) = 0.829 \text{ μm}^2 / \text{hr}$$

$$\frac{B}{A} = 1.63 \times 10^8 \exp\left(-\frac{2.05}{kT}\right) = 15.86 \text{ μm} / \text{hr}$$

$$A = 0.052 \text{ μm}$$

The initial oxide, if grown at 1200°C would have taken this long to grow

$$\tau = \frac{x_i^2 + Ax_i}{B} = \frac{(0.3)^2 + (0.052)(0.3)}{0.829} = 0.127 \text{ hr}$$

The time required to grow 0.8 μm at 1200°C is

$$\tau = \frac{x_i^2 + Ax_i}{B} = \frac{(0.8)^2 + (0.052)(0.8)}{0.829} = 0.822 \text{ hr}$$

Thus, the time required to add 0.5 μm to an existing 0.3 μm film is
 $0.822 - 0.127 = 0.695 \text{ hr}$ or 41.7 minutes.

b. At 1200°C, in dry oxygen

$$B = 7.72 \times 10^2 \exp\left(-\frac{1.23}{k(1200 + 273)}\right) = 0.048 \text{ μm}^2 / \text{hr}$$

$$\frac{B}{A} = 6.23 \times 10^6 \exp\left(-\frac{2.0}{kT}\right) = 0.899 \mu\text{m/hr}$$

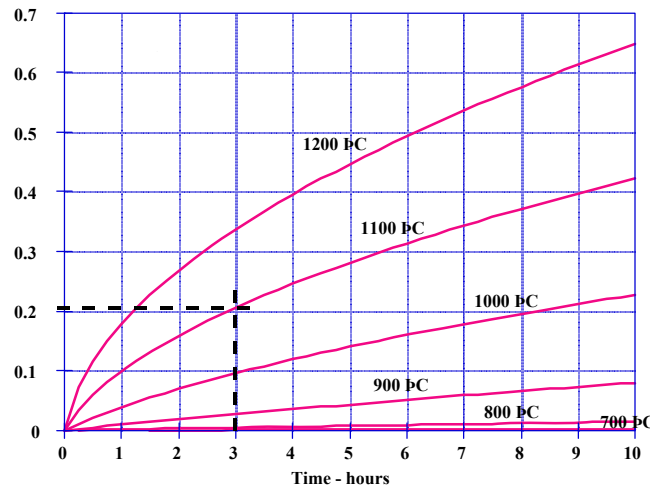
$$A = 0.053 \mu\text{m}$$

The initial oxide would have taken 2.206 hours to grow in dry oxygen, it would require 14.217 hours to grow 0.8 μm , thus would require an additional 12 hours to add 0.5 μm to an existing 0.3 μm film.

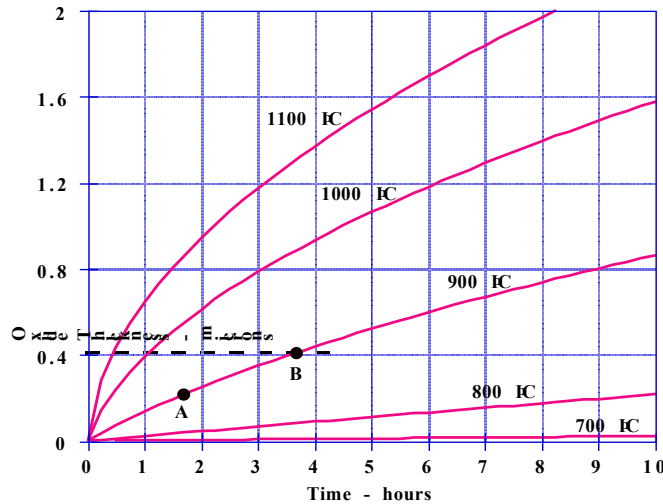
- 6.13. Suppose an oxidation process is used in which (100) wafers are oxidized in O_2 for three hrs. at 1100°C , followed by two hrs. in H_2O at 900°C , followed by two hrs in O_2 at 1200°C . Use Figs. 6-19 and 6-20 in the text to estimate the resulting final oxide thickness. Explain how you use these figures to calculate the results of a multi-step oxidation like this.**

Answer:

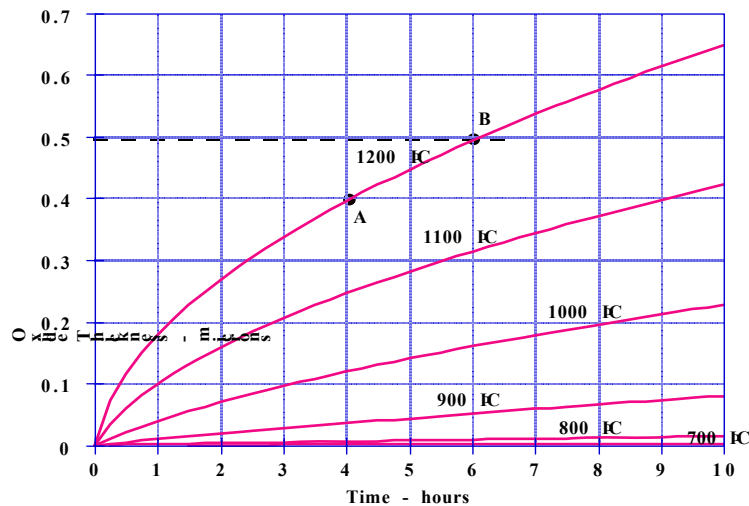
We can use these figures to estimate the oxide thickness as follows. First, we use Fig. 6-19 for the first dry oxidation cycle



A three hour oxidation at 1100°C produces an oxide thickness of about 0.21 μm . We next use Fig. 6-20 for the wet oxidation as shown below. The oxidation is 2 hrs in H_2O at 90°C . We start by finding the point on the 900°C curve that corresponds to 0.21 μm since this is the starting oxide thickness. This is point A. We then move along the 900°C curve by two hours to point B. This corresponds to a thickness of about 0.4 μm which is the thickness at the end of the wet oxidation.



We now go back to Fig. 6-19 for the final dry O_2 cycle. This process is 2 hrs at $1200^\circ C$. We start by finding the point on the $1200^\circ C$ curve that corresponds to a starting oxide thickness of $0.4 \mu m$. This is point A below. We then increment the time by 2 hrs along the $1200^\circ C$ curve, to arrive at a final oxide thickness of about $0.5 \mu m$.



6.14. What is the approximate oxide thickness (use charts) after a 100 minute dry O_2 oxidation followed by a 35 minute H_2O oxidation at $900^\circ C$?

Answer:

The charts are really not very accurate for these thin oxides. We will do the calculation based on the Deal-Grove formulas.

We will perform the calculation for $\langle 111 \rangle$ silicon wafers. For $\langle 100 \rangle$ wafers, the linear rate constant should be divided by 1.68.

At 900°C, in dry oxygen

$$B = 7.72 \times 10^2 \exp\left(-\frac{1.23}{k(900 + 273)}\right) = 0.004 \mu\text{m}^2 / \text{hr}$$

$$\frac{B}{A} = 6.23 \times 10^6 \exp\left(-\frac{2.0}{kT}\right) = 0.016 \mu\text{m} / \text{hr}$$

$$A = 0.251 \mu\text{m}$$

At 900°C, in H₂O

$$B = 3.86 \times 10^2 \exp\left(-\frac{0.78}{kT}\right) = 0.172 \mu\text{m}^2 / \text{hr}$$

$$\frac{B}{A} = 1.63 \times 10^8 \exp\left(-\frac{2.05}{kT}\right) = 0.255 \mu\text{m} / \text{hr}$$

$$A = 0.674 \mu\text{m}$$

We will ignore the initial fast oxidation regime in the following calculations.

The thickness after 100 minutes in dry oxygen at 900°C is

$$x_{\text{O}} = \frac{A}{2} \left\{ \sqrt{1 + \frac{t + \tau}{A^2 / 4B}} - 1 \right\} = \frac{0.251}{2} \left\{ \sqrt{1 + \frac{100/60}{(0.251)^2 / 4(0.004)}} - 1 \right\} = 0.024 \mu\text{m}$$

This thickness corresponds to a time of

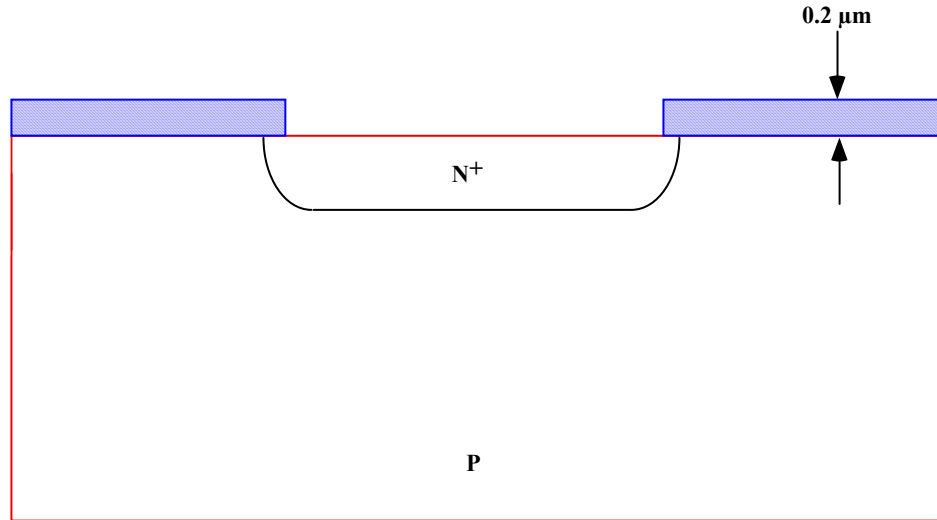
$$\tau = \frac{x_i^2 + Ax_i}{B} = \frac{(0.024)^2 + (0.674)(0.024)}{0.172} = 0.097 \text{ hr}$$

in H₂O at 900°C. This corresponds to the “initial time” that should be used for the wet oxidation step, giving the thickness after an additional 35 minutes in wet oxide as

$$x_{\text{O}} = \frac{A}{2} \left\{ \sqrt{1 + \frac{t + \tau}{A^2 / 4B}} - 1 \right\} = \frac{0.674}{2} \left\{ \sqrt{1 + \frac{35/60 + 0.097}{(0.674)^2 / 4(0.172)}} - 1 \right\} = 0.143 \mu\text{m}$$

6.15. The structure shown below is formed by oxidizing a silicon wafer ($x_0 = 200$ nm), and then using standard masking and etching techniques to remove the SiO₂ in the center region. An N⁺ doping step is then used to produce the

structure shown. The structure is next placed in an oxidation furnace and oxidized at 900°C in H₂O. The oxide will grow faster over the N⁺ region than it will over the lightly doped substrate. Assume that B/A is enhanced by 4X over the N⁺ region. Will the growing oxide over the N⁺ region ever catch up in thickness to the other oxide? If so, when and at what thickness. Use the Deal Grove model for the oxidation kinetics.



Answer:

Using the Deal Grove model for oxidation, the growth is described by

$$\frac{X_O^2 - X_i^2}{B} + \frac{X_O - X_i}{B/A} = t$$

If the oxide over the N⁺ region catches up with the oxide over the P region at some time t , then at that time we will have

$$\frac{X_O^2 - X_i^2}{B} + \frac{X_O - X_i}{B/A} = \frac{X_O^2}{B} + \frac{X_O}{4B/A}$$

where $x_i = 0.2 \mu\text{m}$. This gives

$$x_0 = \frac{4}{3} \left(\frac{x_i^2}{A} + x_i \right)$$

At 900°C, in H₂O

$$B = 3.86 \times 10^2 \exp\left(-\frac{0.78}{kT}\right) = 0.172 \mu\text{m}^2 / \text{hr}$$

$$\frac{B}{A} = 1.63 \times 10^8 \exp\left(-\frac{2.05}{kT}\right) = 0.255 \mu\text{m/hr}$$

$$A = 0.674 \mu\text{m}$$

These values apply to (111) silicon. For (100) silicon, B/A is 1.68 times smaller. Thus we have,

$$A_{(111)} = 0.674 \mu\text{m}$$

$$A_{(100)} = 1.13 \mu\text{m}$$

Finally, the N^+ oxide catches up with the P oxide at an oxide thickness of

$$X_{(111)} = 0.346 \mu\text{m}$$

$$X_{(100)} = 0.314 \mu\text{m}$$

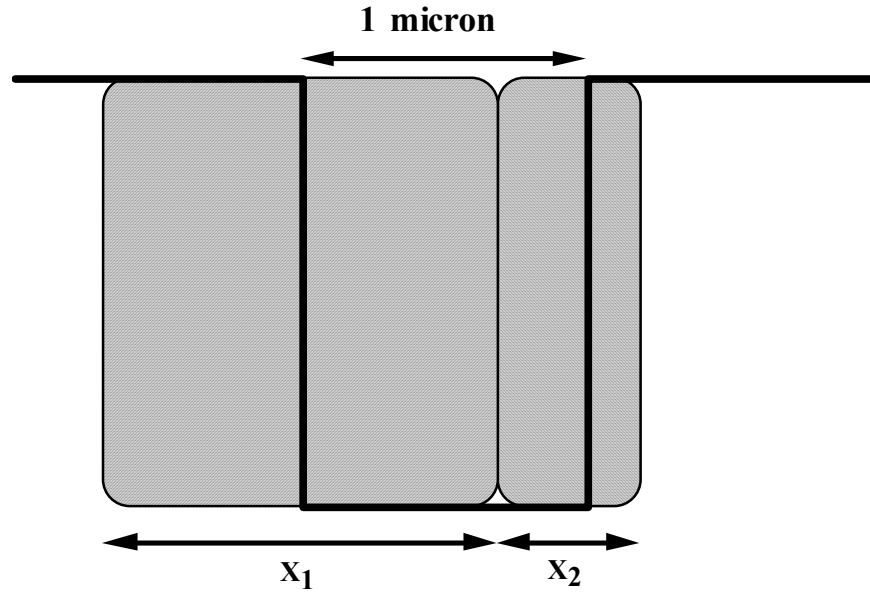
This occurs at a time given by

$$t_{(111)} = \frac{(0.346)^2 - (0.2)^2}{.172} + \frac{0.346 - 0.2}{0.255} = 1.04 \text{ hrs}$$

$$t_{(100)} = \frac{(0.314)^2 - (0.2)^2}{.172} + \frac{0.314 - 0.2}{0.152} = 1.09 \text{ hrs}$$

6.16. A $1\mu\text{m}$ wide trench is etched in a $\langle 100 \rangle$ silicon wafer, so that the sides of the trench are $\langle 110 \rangle$ planes. An angled implant is performed, doping the sidewall N^+ and thereby enhancing the linear rate constant by a factor of 4. The structure is then oxidized in steam at 1100°C . At what time during the oxidation will the groove be filled with SiO_2 ? Assume the appropriate oxidation coefficients scale as $[(111 : 110 : 100) = (1.68 : 1.2 : 1.0)]$.

Answer:



$$X_1 + X_2 = \frac{1.0}{0.55} = 1.82 \mu\text{m}$$

$$\Rightarrow X_2 = 1.82 - X_1$$

The linear growth rate is 4 times larger on the doped side of the trench, but the parabolic growth rates are the same, so that

$$\frac{X_1^2}{B} + \frac{X_1}{4(B/A)} = t = \text{oxidation time} = \frac{X_2^2}{B} + \frac{X_2}{(B/A)}$$

$$\Rightarrow \frac{1}{B} (X_1^2 - X_2^2) + \frac{1}{(B/A)} \left(\frac{X_1}{4} + X_2 \right) = 0$$

$$\frac{1}{B} (X_1 + X_2)(X_1 - X_2) + \frac{1}{(B/A)} (0.25X_1 - 1.82 + X_1) = 0$$

$$\frac{1}{B} (1.82)(X_1 - 1.82 + X_1) + \frac{1}{(B/A)} (1.25X_1 - 1.82) = 0$$

$$\left(\frac{3.64}{B} + \frac{1.25}{B/A} \right) X_1 = \left(\frac{3.31}{B} + \frac{1.82}{B/A} \right)$$

For the <110> surface:

$$B_{\langle 110 \rangle} = B_{\langle 100 \rangle}$$

$$\left(\frac{B}{A}\right)_{\langle 110 \rangle} = 1.2 \left(\frac{B}{A}\right)_{\langle 100 \rangle}$$

At 1100°C, $kT=0.1183\text{eV}$

$$B_{\langle 110 \rangle} = 3.86 \times 10^2 \exp\left(-\frac{0.78}{0.1183}\right) = 0.528$$

$$\left(\frac{B}{A}\right)_{\langle 110 \rangle} = 1.2 \times 1.63 \times 10^8 \exp\left(-\frac{2.05}{0.1183}\right) = 5.83$$

$$\Rightarrow X_1 = 0.926 \mu\text{m} \quad , \quad X_2 = 0.892 \mu\text{m}$$

$$\Rightarrow t = 1.66 \text{ h}$$

Note: The n^+ doping affects only the linear rate constant, and since the oxide is thick, the effect of the sidewall doping is not very significant.

6.17. A uniform oxide layer of $0.4\mu\text{m}$ thickness is selectively etched to expose the silicon surface in some locations on a wafer surface. A second oxidation at 1000°C in H_2O grows $0.2\mu\text{m}$ on the bare silicon.

(a) Sketch a cross-section of the SiO_2 in all locations on the wafer and the position of the Si/SiO_2 interface.

(b) Would your picture be the same if the second oxidation grew the $0.2\mu\text{m}$ at a different temperature? Explain.

Answer:

We will perform the calculation for $\langle 111 \rangle$ silicon wafers. For $\langle 100 \rangle$ wafers, the linear rate constant should be divided by 1.68.

At 1000°C , in H_2O

$$B = 3.86 \times 10^2 \exp\left(-\frac{0.78}{kT}\right) = 0.316 \mu\text{m}^2 / \text{hr}$$

$$\frac{B}{A} = 1.63 \times 10^8 \exp\left(-\frac{2.05}{kT}\right) = 1.255 \mu\text{m} / \text{hr}$$

$$A = 0.252 \mu\text{m}$$

The time to grow 0.2 μm is then

$$\tau = \frac{x_i^2 + Ax_i}{B} = \frac{(0.2)^2 + (0.252)(0.2)}{0.316} = 0.286 \text{ hr}$$

The time to grow the initial 0.4 μm was

$$\tau = \frac{x_i^2 + Ax_i}{B} = \frac{(0.4)^2 + (0.252)(0.4)}{0.316} = 0.825 \text{ hr}$$

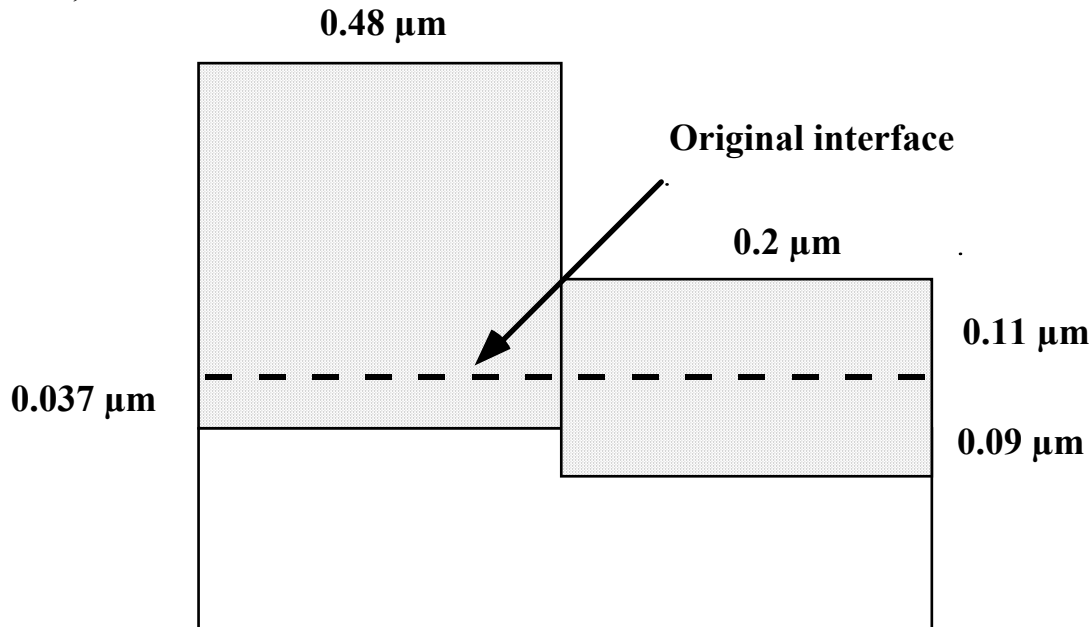
The extra time to grow 0.2 μm in the bare areas will grow some extra oxide where the initial 0.4 μm oxide existed

$$x_O = \frac{A}{2} \left\{ \sqrt{1 + \frac{t + \tau}{A^2 / 4B}} - 1 \right\} = \frac{0.252}{2} \left\{ \sqrt{1 + \frac{0.286 + 0.825}{(0.252)^2 / 4(0.316)}} - 1 \right\} = 0.480 \mu\text{m}$$

The growth of the extra 0.080 μm will consume $0.46 \times 0.080 = 0.037 \mu\text{m}$ of silicon under the thick oxide.

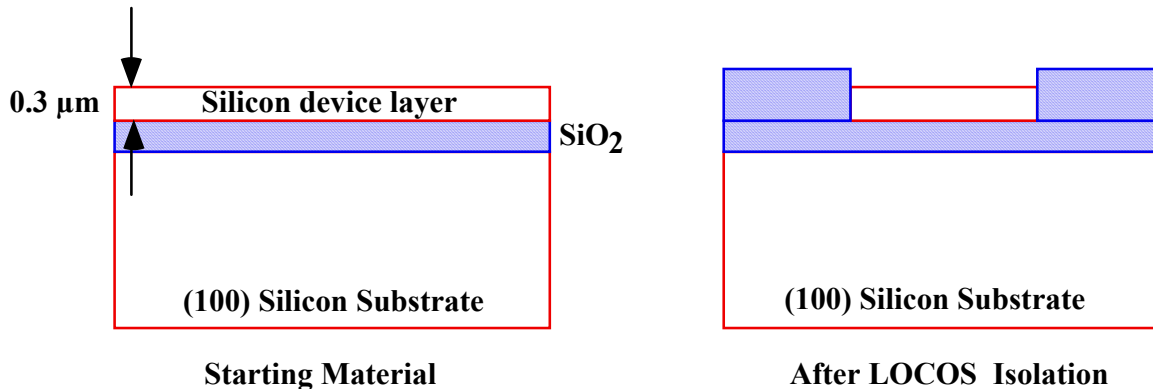
The growth of 0.2 μm of oxide will consume $0.46 \times 0.2 = 0.092 \mu\text{m}$ of silicon.

Thus, a sketch of the old and new oxide-silicon interface is shown below:



6.18. Silicon on Insulator or SOI is a new substrate material that is being considered for future integrated circuits. The structure, shown below, consists of a thin single crystal silicon layer on an insulating (SiO_2) substrate. The silicon below

the SiO_2 provides mechanical support for the structure. One of the reasons this type of material is being considered, is because junctions can be diffused completely through the thin silicon layer to the underlying SiO_2 . This reduces junction capacitances and produces faster circuits. Isolation is also easy to achieve in this material, because the thin Si layer can be completely oxidized, resulting in devices completely surrounded by SiO_2 . A LOCOS process is used to locally oxidize through the silicon as shown on the right below. Assuming the LOCOS oxidation is done in H_2O at 1000°C , how long will it take to oxidize through the $0.3\text{ }\mu\text{m}$ silicon layer? Calculate a numerical answer using the Deal Grove model.



Answer:

To oxidize completely through a $0.3\text{ }\mu\text{m}$ silicon layer, we will need to grow $(2.2)(0.3\text{ }\mu\text{m}) = 0.66\text{ }\mu\text{m}$ of SiO_2 . At 1000°C in H_2O , the Deal Grove rate constants are given by (Table 6-2):

$$B = 3.86 \times 10^2 \exp\left(-\frac{0.78\text{eV}}{kT}\right) = 0.316\text{ }\mu\text{m}^2\text{hr}^{-1}$$

$$\frac{B}{A} = \frac{1.63 \times 10^8}{1.68} \exp\left(-\frac{2.05\text{eV}}{kT}\right) = 0.747\text{ }\mu\text{mhr}^{-1}$$

$$\therefore t = \frac{(0.66)^2}{0.316} + \frac{0.66}{0.747} \cong 2.25\text{ hours}$$

6.19. As MOS devices are scaled to smaller dimensions, gate oxides must be reduced in thickness. a). As the gate oxide thickness decreases, do MOS devices become more or less sensitive to sodium contamination? Explain. b). As the gate oxide thickness decreases, what must be done to the substrate doping (or alternatively the channel V_{TH} implant, to maintain the same V_{TH} ?

Answer:

(a) For thinner gate oxides, the capacitance of the gate oxide per unit area increases

$$C_{OX} = \frac{\epsilon}{d}$$

Thus, the change in threshold caused by mobile ions will be smaller for large gate oxide capacitances

$$\Delta V_T \propto \frac{qQ_m}{C_{OX}}$$

So thinner oxides will be less sensitive to sodium contamination.

(b) Since

$$V_T \propto \frac{1}{C_{OX}}$$

an increase in the oxide capacitance caused by a thinner oxide will tend to give a lower threshold voltage if the doping in the substrate is unchanged. Thus, to keep the same threshold voltage, a higher doping level in the substrate is required.

6.20. A SiO₂ layer is thermally grown at 1000°C using a 10-20-30 min dry-wet-dry oxidation cycle. Upon pulling the (100) wafer from the furnace, the oxide color is observed to be tan. Is this right? If not, suggest what might have gone wrong in the experiment.

Answer:

Most of the oxide should grow during the 20 min wet cycle since the growth rate is so much faster in H₂O. Looking at the oxidation curve in Fig. 6.19, 20 min at 1000°C should have grown ≈ 150 nm of SiO₂. (A more accurate estimate could be obtained using the formulas for the Deal Grove model. Looking at the color charts in the Appendix, we do not find any tan color near 150 nm, so something must have gone wrong.

The only tan on charts is at around 50 nm, much thinner than we expected. There are thus a number of possible explanations. Perhaps the furnace temperature was much lower than the 1000 °C expected. A more likely reason is apparent if we look at the dry O₂ chart (Fig. 6.18) and observe that a 60 min 1000 °C dry O₂ oxidation should produce ≈ 40 nm of SiO₂. This is in the range for the tan color. So probably what happened is that the H₂O did not come on during the wet cycle.

- 6.21. In the simulation example in Fig. 6.35, the difference in oxidation rate between the heavily doped and lightly doped regions is much more pronounced at low temperatures (the 800°C example) than at high temperatures (the 1000°C example). Explain physically why this is the case using the mechanisms in the Deal Grove model and the behavior of point defect (V) concentrations versus temperature.

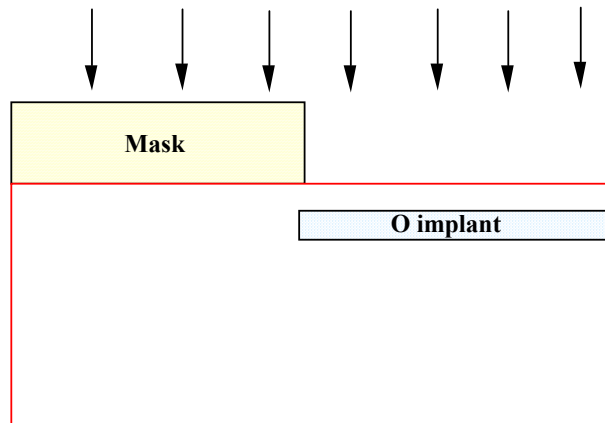
Answer:

The high doping affects the linear rate constant. The microscopic mechanism depends on how the vacancy concentration in the substrate compares to the intrinsic vacancy concentration. The intrinsic value for the linear rate constant is modified as follows for an n-type sample

$$\frac{B}{A} \propto \left(\frac{B}{A}\right)^i \left(\frac{C_V}{C_V^*}\right) \propto \left(\frac{B}{A}\right)^i \left(\frac{n}{n_i}\right)$$

At lower temperatures, the doping level causes the sample to be more extrinsic, so the difference in oxidation rates between an intrinsic region and an extrinsic region becomes larger.

- 6.22. The structure shown below is implanted with oxygen using a $1 \times 10^{18} \text{ cm}^{-2}$ implant at 200 keV ($R_P = 0.35 \mu\text{m}$). The left hand side is masked from the implant. Following the implant, a high temperature anneal is performed which forms stoichiometric SiO_2 in a buried layer on the right side. Calculate the structural dimensions on the right side following this anneal (oxide thickness, distance from the surface and all other important dimensions). You can assume the silicon atomic density is $5 \times 10^{22} \text{ atoms cm}^{-3}$, and the Si lattice planes are 0.25 nm apart. State any other assumptions you make.



Answer:

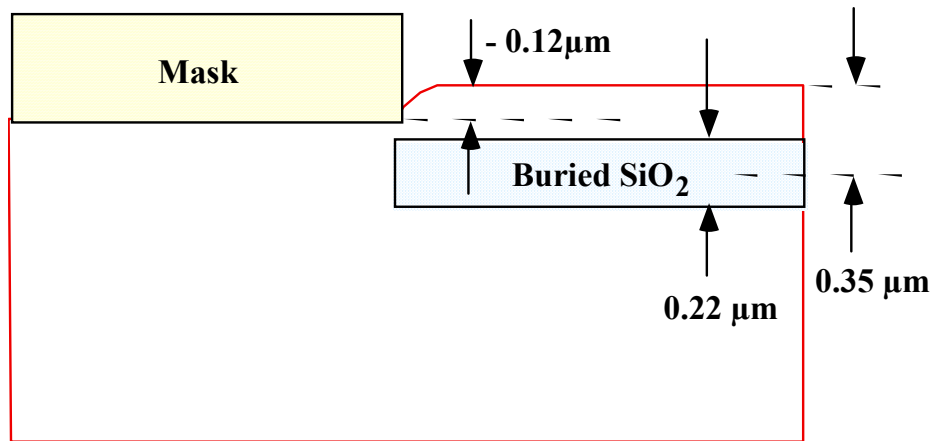
Given the Si lattice density and the lattice plane spacing, we have:

$$\# \text{ of Si atoms per lattice plane} = (5 \times 10^{22}) (2.8 \times 10^{-8} \text{ cm}) = 1.25 \times 10^{15} \text{ cm}^{-2}$$

Thus a $1 \times 10^{18} \text{ cm}^{-2}$ oxygen implant will consume $(1 \times 10^{18}) / (2)(1.25 \times 10^{15}) = 400$ planes of Si atoms in making SiO_2 . This corresponds to 100 nm ($0.1 \mu\text{m}$) of Si being consumed.

Because of the volume expansion in forming SiO_2 , 100 nm of Si will form 220 nm of SiO_2 . Thus the buried SiO_2 layer will be centered $0.35 \mu\text{m}$ below the surface (at R_p), and extend 110 nm up and down from that point.

The final issue to be considered is how the volume expansion is accommodated. The Si above the buried SiO_2 will be pushed up and the Si below the SiO_2 will be pushed down. Since the buried SiO_2 is closest to the top surface, the biggest effect will be there, resulting in the structure shown below. The 120 nm distance should be regarded as approximate and would require a detailed mechanical calculation to compute accurately.



6.23. As part of an IC process flow, a CVD SiO_2 layer $1.0 \mu\text{m}$ thick is deposited on a $\langle 100 \rangle$ silicon substrate. This structure is then oxidized at 900°C for 60 minutes in an H_2O ambient. What is the final SiO_2 thickness after this oxidation? Calculate an answer, do not use the oxidation charts in the text.

Answer:

At 900°C in H_2O , the oxidation rate constants are given by:

$$B = 3.86 \times 10^2 \exp \left(- \frac{0.78}{(8.62 \times 10^{-5}) (1173)} \right) \mu\text{m}^2 \text{ hr}^{-1} = 0.17 \mu\text{m}^2 \text{ hr}^{-1}$$

$$\frac{B}{A} = \frac{1.63 \times 10^8}{1.68} \exp \left(- \frac{2.05}{(8.62 \times 10^{-5})(1173)} \right) \mu\text{m hr}^{-1} = 0.152 \mu\text{m hr}^{-1}$$

The initial oxide on the wafer is 1.0 μm thick. This corresponds to a τ of

$$\tau = \frac{(1)^2 + (1) \left(\frac{0.17}{0.152} \right)}{0.17} = 12.46 \text{ hours}$$

Thus the final oxide thickness is given by

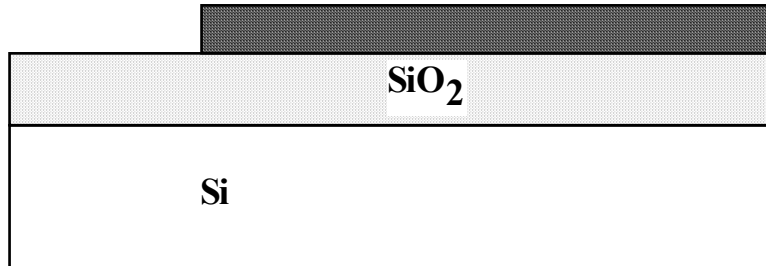
$$x_o = \frac{0.17}{(2)(0.152)} \left\{ \sqrt{1 + \frac{13.46}{\frac{(1.11)^2}{(4)(0.17)}}} - 1 \right\} = 1.064 \mu\text{m}$$

Thus not much additional oxide grows.

Chapter 7 Problems

7.1. A resistor for an analog integrated circuit is made using a layer of deposited polysilicon 0.5 μm thick, as shown below.

Polysilicon



- (a) The doping the polysilicon is $1 \times 10^{16} \text{ cm}^{-3}$. The carrier mobility $\mu = 100 \text{ cm}^2 \text{ V}^{-1} \text{ sec}^{-1}$ is low because of scattering at grain boundaries. If the resistor has $L=100\mu\text{m}$, $W=10\mu\text{m}$, what is its resistance in Ohms?
- (b) A thermal oxidation is performed on the polysilicon for 2 hours at 900°C in H_2O . Assuming B/A for polysilicon is $2/3$ that of $\langle 111 \rangle$ silicon, what is the polysilicon thickness that remains.
- (c) Assuming that all of the dopant remains in the polysilicon (i.e. does not segregate to oxide), what is the new value of the resistor in (a). Assume the mobility does not change.

Answer:

(a)

$$\rho = \frac{1}{nq\mu} = \frac{1}{(1 \times 10^{16})(1.6 \times 10^{-19})(100)} = 6.25 \Omega\text{cm}$$

$$\rho_S = \frac{\rho}{x_j} = \frac{6.25}{0.5 \times 10^{-4}} = 125 \text{ k}\Omega$$

$$R = \frac{100}{10} \rho_S = 1.25 \text{ M}\Omega \quad (10 \text{ squares})$$

(b) The linear rate coefficient at 900°C is

$$\left(\frac{B}{A}\right)_{\text{poly}} = \frac{2}{3} \left(1.63 \times 10^8 \exp\left(-\frac{2.05}{kT}\right) \right) = 0.170 \quad \mu\text{m hr}^{-1}$$

The parabolic rate constant for poly is unchanged:

$$B_{\text{poly}} = 3.86 \times 10^2 \exp\left(-\frac{0.78}{kT}\right) = 0.172 \mu\text{m}^2 \text{ hr}^{-1}$$

$$A_{\text{poly}} = 1.01 \mu\text{m}$$

The oxide thickness is

$$x_o = \frac{A}{2} \left\{ \sqrt{1 + \frac{t}{A^2/4B}} - 1 \right\}$$

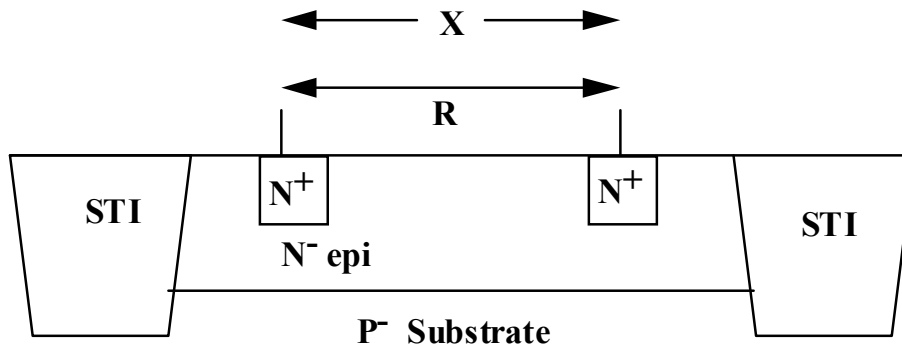
$$x_o = \frac{1.01}{2} \left\{ \sqrt{1 + \frac{2}{(1.01)^2/4(0.172)}} - 1 \right\} = 0.27 \mu\text{m}$$

This oxide consumes a silicon thickness of $0.45 \times 0.27 = 0.12 \mu\text{m}$, leaving a remaining polysilicon thickness of $0.5 - 0.12 = 0.38 \mu\text{m}$ and contains all the dopant with a concentration of

$$\left(1 \times 10^{16}\right) \frac{0.5}{0.38} = 1.31 \times 10^{16} \text{ cm}^{-3}$$

(c) Since the concentration has gone up and the thickness has gone down by the same factor, the polysilicon resistivity and hence the resistance of the line remains the same.

- 7.2. A resistor is made as part of a high frequency analog integrated circuit as shown below. The N^- epi layer forms the body of the resistor. If the width of the resistor in the direction into the paper is $2.5 \mu\text{m}$, what should the length X be to give a resistor of approximately $50 \text{ k}\Omega$. The epilayer is doped with phosphorus at a concentration of $1 \times 10^{15} \text{ cm}^{-3}$ and is $3 \mu\text{m}$ thick.**



Answer:

From Fig. 1-18, the resistivity of a $1 \times 10^{15} \text{ cm}^{-3}$ n-type region is

$$\rho = 4 \Omega\text{cm}$$

The sheet resistance of the layer is

$$\rho_S = \frac{4}{(3 \times 10^{-4})} = 7500 \, \Omega/\text{square}$$

The number of squares to get the required resistance is

$$n = \frac{50 \times 10^3}{7500} = 6.66$$

so the required length is

$$X = 6.66 \times 2.5 = 16.66 \, \mu\text{m}$$

7.3. A p-type (boron) diffusion is performed as follows:

Pre-dep: 30 minutes, 900°C, solid solubility

Drive-in: 60 minutes, 1000°C

(a) What is the deposited Q?

(b) If the substrate is doped $1 \times 10^{15} \, \text{cm}^{-3}$ phosphorus, what is x_J .

(c) What is the sheet resistance of the diffused layer?

Answer:

The boron diffusion coefficient is

$$\text{At } 900^\circ\text{C: } D_B^{900} = 1.0 \exp\left(-\frac{3.5}{k(900 + 273)}\right) = 9.27 \times 10^{-16} \, \text{cm}^2\text{s}^{-1}$$

$$\text{At } 1000^\circ\text{C: } D_B^{1000} = 1.0 \exp\left(-\frac{3.5}{k(1000 + 273)}\right) = 1.41 \times 10^{-14} \, \text{cm}^2\text{s}^{-1}$$

(a) The predep is performed at 900°C where the boron solid solubility from Table 7.4 is

$$C_S = 1.2 \times 10^{20} \, \text{cm}^{-3}$$

The dose introduced is then

$$Q = \frac{2C_S}{\sqrt{\pi}} \sqrt{Dt} = \frac{2(1.2 \times 10^{20})}{\sqrt{\pi}} \sqrt{(9.27 \times 10^{-16})(30 \times 60)} = 1.75 \times 10^{14} \, \text{cm}^{-2}$$

(b) The implicit assumption we will make is that the drive-in for 60 minutes at 1000°C is sufficient to make the initial predeposition profile look like a delta-function. We will treat the predep profile as a delta function and calculate the junction based on a one-sided gaussian diffusion near a surface. We can check this assumption now:

$$(Dt)_{900} = 30 \times 60 \times 9.27 \times 10^{-16} = 1.67 \times 10^{-12} \ll$$

$$(Dt)_{1000} = 60 \times 60 \times 1.4 \times 10^{-14} = 5.0 \times 10^{-11}$$

which shows that the initial predep has a (Dt) product 30 times less than the final drive in. This means that the initial profile looks like a delta-function compared to the final profile and we are justified in treating the final profile as a gaussian. Then

$$x_j = \sqrt{(4Dt) \ln \left(\frac{Q}{C_B \sqrt{\pi Dt}} \right)} =$$

$$\sqrt{(4 \times 1.41 \times 10^{-14} \times 60 \times 60) \ln \left(\frac{1.75 \times 10^{14}}{1 \times 10^{15} \sqrt{\pi \times 1.41 \times 10^{-14} \times 60 \times 60}} \right)}$$

$$= 0.44 \times 10^{-4} \text{ cm} = 0.44 \mu\text{m}$$

(c) A p-type Gaussian profile with a peak surface concentration of $1.2 \times 10^{20} \text{ cm}^{-3}$ has an effective conductivity from Fig.7.17 of approximately

$$\bar{\sigma} = \frac{1}{\rho_S x_j} = 300 (\Omega \cdot \text{cm})^{-1}$$

Thus, the sheet resistance of the layer is

$$\rho_S = \frac{1}{300 \times 0.44 \times 10^{-4}} = 75.75 \Omega/\text{square}$$

7.4. Suppose we perform a solid solubility limited predeposition from a doped glass source which introduces a total of Q impurities / cm^2 .

- (a) **If this predeposition was performed for a total of t minutes, how long would it take (total time) to predeposit a total of $3Q$ impurities / cm^2 into a wafer if the predeposition temperature remained constant.**
- (b) **Derive a simple expression for the $(Dt)_{\text{drive-in}}$ which would be required to drive the initial predeposition of Q impurities / cm^2 sufficiently deep so that the final surface concentration is equal to 1% of the solid solubility concentration. This can be expressed in terms of $(Dt)_{\text{predep}}$ and the solid solubility concentration C_S .**

Answer:

(a)

$$Q = \frac{2C_S}{\sqrt{\pi}} \sqrt{Dt} \Rightarrow Q \propto \sqrt{t}$$

$$\therefore 3Q \Rightarrow 9t$$

(b)

$$C(0,t)_{\text{drive-in}} = \frac{Q}{\sqrt{\pi Dt}} = 0.01 C_S$$

$$Q = \frac{2C_S}{\sqrt{\pi}} \sqrt{(Dt)_{\text{predep}}}$$

$$\therefore \frac{2}{\sqrt{\pi}} \frac{\sqrt{(Dt)_{\text{predep}}}}{\sqrt{(Dt)_{\text{drive-in}}}} = 0.01$$

$$\therefore (Dt)_{\text{drive-in}} = \left(\frac{200}{\pi} \right)^2 (Dt)_{\text{predep}}$$

7.5. A diffused region is formed by an ultra-shallow implant followed by a drive-in. The final profile is Gaussian. Derive a simple expression for the sensitivity of x_j to the implant dose Q . Is x_j more sensitive to Q at high or low doses?

Answer:

$$x_j = \left\{ 4Dt \ln \left(\frac{Q}{\sqrt{\pi Dt} C_B} \right) \right\}^{\frac{1}{2}}$$

$$\therefore \frac{dx_j}{dQ} = \sqrt{4Dt} \left(\frac{1}{2} \right) \left\{ \ln \left(\frac{Q}{\sqrt{\pi Dt} C_B} \right) \right\}^{-\frac{1}{2}} \left(\frac{Q}{\sqrt{\pi Dt} C_B} \right)^{-1} \left(\frac{1}{\sqrt{\pi Dt} C_B} \right)$$

$$\frac{dx_j}{dQ} = \frac{2Dt}{x_j Q}$$

$$\Delta x_j = \frac{2Dt}{x_j Q} \Delta Q$$

so that the junction depth is more sensitive to Q for small values of Q .

7.6. From a process control point of view, predeposition times > 10 min are required. From an economic point of view, times < 10 hours are required. Equipment limitations restrict $700^{\circ}\text{C} < T < 1200^{\circ}\text{C}$. You need only consider simple Gaussian and erfc type profiles in this problem.

a). Is it possible to dope an MOS channel region by predeposition to shift threshold voltages? The required dose is $5 \times 10^{11} \text{ cm}^{-2}$, boron.

b). What would be a reasonable schedule (T, t) for an MOS source drain predep? The required dose is $5 \times 10^{15} \text{ cm}^{-2}$, arsenic. The junction depth cannot be deeper than $0.2 \mu\text{m}$ for device reasons. Assume the channel doping is $1 \times 10^{18} \text{ cm}^{-3}$.

c). What sheet resistance would your profile in b) have if the channel doping is $1 \times 10^{18} \text{ cm}^{-3}$ P type?

Answer:

(a) The dose required to shift the threshold voltage is very low, so we will pick that lower limit of the diffusion temperatures to deposit the dose where the solubility from Table 7.4 is

$$C_s^{700} \approx 4 \times 10^{19} \text{ cm}^{-3}$$

The predep Dt is then

$$(Dt) = \left(\frac{\sqrt{\pi} Q}{2C_s} \right)^2 = \left(\frac{\sqrt{\pi} 5 \times 10^{11}}{2 \times 4 \times 10^{19}} \right)^2 = 1.23 \times 10^{-16} \text{ cm}^2$$

The diffusivity of boron at 700°C is

$$D_B = 1.0 \exp \left(- \frac{3.5}{(8.62 \times 10^{-5})(700 + 273)} \right) = 7.5 \times 10^{-19} \text{ cm}^2 \text{ sec}^{-1}$$

The time required is thus

$$t = \frac{1.23 \times 10^{-16}}{7.5 \times 10^{-19}} = 163 \text{ sec}$$

which is far too short for a furnace operation. It is not possible to predeposit the required dose at the solid solubility limit and in practice ion implantation is the alternative used.

(b) We have a choice in this design problem, so we can pick a temperature of 1000°C and see if the time we get is reasonable.

The arsenic diffusivity at 1000°C is

$$D_{As} = 9.17 \exp\left(-\frac{3.99}{8.62 \times 10^{-5}(1000 + 273)}\right) = 1.48 \times 10^{-15} \text{ cm}^2 \text{ sec}^{-1}$$

The solid solubility from Table 7.4 is

$$C_S \approx 2 \times 10^{21} \text{ cm}^{-3}$$

and the (Dt) required is

$$(Dt) = \left(\frac{\sqrt{\pi} Q}{2C_S}\right)^2 = \left(\frac{\sqrt{\pi} 5 \times 10^{15}}{2 \times 2 \times 10^{21}}\right)^2 = 4.9 \times 10^{-12} \text{ cm}^2$$

giving a time of

$$t = \frac{4.9 \times 10^{-12}}{1.48 \times 10^{-15}} = 3311 \text{ sec} = 55 \text{ mins}$$

This is a reasonable time for a furnace operation at 1000°C and will deposit the required dose.

Checking the junction depth when the background doping level is $1 \times 10^{18} \text{ cm}^{-3}$ and since the surface is held at solid solubility, the error function formula is appropriate:

$$C(x, t) = 1 \times 10^{18} = C_S \operatorname{erfc} \frac{x_j}{2\sqrt{Dt}}$$

$$\frac{x_j}{2\sqrt{Dt}} = \operatorname{erf}^{-1}\left(1 - \frac{1 \times 10^{18}}{2 \times 10^{21}}\right) = \operatorname{erf}^{-1}(0.9995) \quad \therefore x_j = 0.109 \text{ } \mu\text{m}$$

(c) From Irvin's curves for an n-type, complementary error function profile when the surface concentration is

$$\bar{\sigma} = \frac{1}{\rho_S x_j} = 1300$$

giving

$$\rho_S = 70.6 \text{ } \Omega/\text{square}$$

7.7. A boron diffusion is performed in silicon such that the maximum boron concentration is $1 \times 10^{18} \text{ cm}^{-3}$. For what range of diffusion temperatures will electric field effects and concentration dependent diffusion coefficients be important?

Answer:

Electric field effects and concentration dependent diffusion are both important when the doping concentration exceeds the intrinsic electron (or hole) concentration. The intrinsic or background electron concentration is n_i which increases with higher temperature. This provides a background sea of electrons or holes in the lattice at a given temperature. If the doping exceeds this concentration, then these extrinsic effects are important.

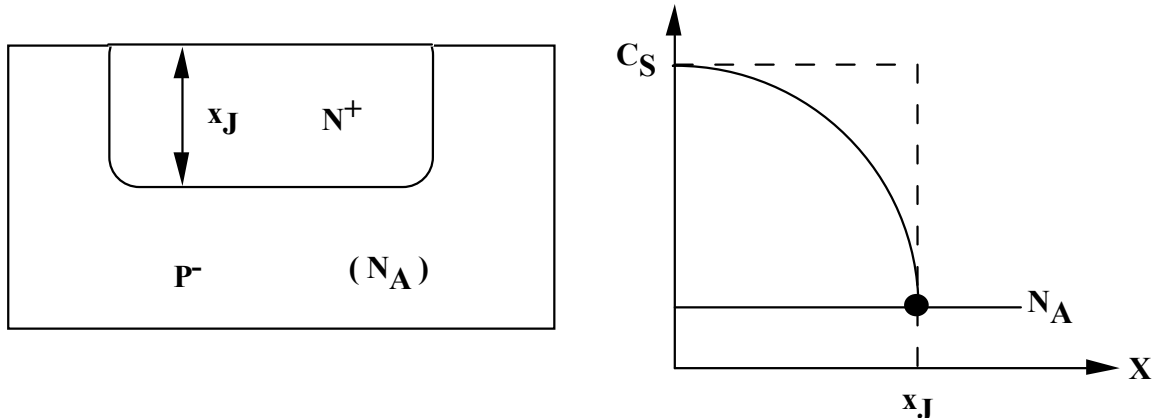
When the temperature is below the temperature where $n_i = 1 \times 10^{18} / \text{cm}^3$, these effects will become dominant since they often depend on n / n_i (where $n = N_A$ or $n = N_D$ to a first approximation).

$$n_i = 3.9 \times 10^{16} T^{3/2} \exp\left(-\frac{0.605}{kT}\right)$$

By trial and error, $n_i = 1 \times 10^{18} / \text{cm}^3$ at $T=720^\circ\text{C}$.

Therefore, extrinsic effects become important below 720°C .

- 7.8. An N^+ region is formed in a P^- substrate with a junction depth is x_J as shown below. For the device being fabricated it is important to minimize the sheet resistance of the N^+ region.



- If the ideal "box" profile shown could be obtained, it would provide an absolute minimum limit on ρ_s . Derive an approximate expression for this lower bound on ρ_s .
- If As had been used as the dopant and a metastable doping at the solid solubility of $2 \times 10^{21} \text{cm}^{-3}$ was obtained in a box-shaped profile by laser melting, estimate the absolute minimum ρ_s for an x_J of $0.1 \mu\text{m}$ if the limiting carrier mobility in N^+ silicon is $85 \text{cm}^2 \text{volt}^{-1} \text{sec}^{-1}$
- If a normal error function profile was used with the surface at the metastable solid solubility, what value of ρ_s would be realized with an x_J of $0.1 \mu\text{m}$.

- (d) Repeat the calculation in (c) if the arsenic had deactivated to its normal "electrical solubility" in silicon.

Answer:

- (a) The sheet resistance for a uniform box-shaped profile is given by

$$\rho_s = \frac{1}{Nq\mu x_j}$$

- (b) The sheet resistance is given by

$$\rho_s = \frac{1}{(2 \times 10^{21} \text{ cm}^{-3})(1.6 \times 10^{-19} \text{ C})(85 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1})(0.1 \times 10^{-4} \text{ cm})} = 3.67 \Omega/\text{square}$$

- (c) If an error function profile was used, we can estimate the average conductivity of the layer from Irvin's n-type complementary error function curves to be (extrapolating)

$$\bar{\sigma} = \frac{1}{\rho_s x_j} = 1500 (\Omega \text{cm})^{-1}$$

so that

$$\rho_s = \frac{1}{(1500)(0.1 \times 10^{-4})} = 66.7 \Omega/\text{square}$$

- (d) The "electrical solubility" of arsenic in silicon is approximately an order of magnitude lower than the solid solubility, because of the formation of Arsenic-Vacancy clusters which inactivate the arsenic. From Irvin's curves at an active surface concentration of $2 \times 10^{20} \text{ cm}^{-3}$

$$\bar{\sigma} = 400 (\Omega - \text{cm})^{-1}$$

so that

$$\rho_s = \frac{1}{(400)(0.1 \times 10^{-4})} = 250 \Omega/\text{square}$$

which is typical for a high concentration S/D diffusion.

- 7.9. A bipolar transistor is fabricated by implanting the boron base with a dose of $Q_B = 2 \times 10^{13} \text{ cm}^{-2}$. Then an in-situ arsenic doped polysilicon emitter region doped at 10^{20} cm^{-3} is deposited on the surface. A drive-in is performed for 60 min at 1100°C . Assume that the implant can be treated as a delta function. The substrate is 10^{15} cm^{-3} N type. Calculate the base width of the final NPN bipolar transistor. You can neglect all second order effects like concentration dependent diffusion, ϵ field effects, segregation, rapid diffusion in polysilicon etc.

Answer:

The arsenic diffuses from a semi-infinite source (the doped polysilicon) and thus diffuses like a complementary error function

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

The boron implant looks like a delta function that diffuses in a semi-infinite medium (assuming no difference in diffusivity in the polysilicon layer and the silicon layer) and is thus a gaussian

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

The appropriate diffusion coefficients are (ignoring any concentration dependence)

$$D_{\text{As}}^{1100\text{C}} = 9.17 \exp \left(-\frac{3.99}{8.62 \times 10^{-5} \times (1100 + 273)} \right) = 2.1 \times 10^{-14} \text{ cm}^2 \text{ s}^{-1}$$

$$D_{\text{B}}^{1100\text{C}} = 1.0 \exp \left(-\frac{3.5}{kT} \right) = 1.5 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1}$$

The junction between the arsenic and boron occurs when the error function equals the gaussian function, giving

$$\frac{C_{\text{As}}}{2} \left[\operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right) \right] - \frac{Q_B}{2\sqrt{\pi Dt}} \exp \left(-\frac{x^2}{4Dt} \right) = 0$$

$$\frac{10^{20}}{2} \left[\operatorname{erfc} \left(\frac{x}{2\sqrt{2.1 \times 10^{-14} \times 3600}} \right) \right] - \frac{2 \times 10^{13}}{2\sqrt{\pi \times 1.5 \times 10^{-13}}} \exp \left(-\frac{x^2}{4 \times 1.5 \times 10^{-13} \times 3600} \right) = 0$$

Calculating, and using units of microns gives

$$\operatorname{erfc}\left(\frac{x}{0.174}\right) - 4.8 \times 10^{-3} \exp\left(-\frac{x^2}{0.22}\right) = 0$$

Choosing values of x and iterating a few times:

$$x_j^{\text{EB}} = 0.37 \mu\text{m}$$

The junction between the base and collector occurs where the boron base meets the substrate doping

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

$$1 \times 10^{15} = \frac{2 \times 10^{13}}{2\sqrt{\pi \times 1.5 \times 10^{-13}}} \exp\left(-\frac{x^2}{4 \times 1.5 \times 10^{-13} \times 3600}\right)$$

$$x_j^{\text{BC}} = \sqrt{\ln\left(\frac{Q}{C_{\text{SUB}} \times 2\sqrt{\pi Dt}}\right) \times 4Dt}$$

$$x_j^{\text{BC}} = 1.09 \mu\text{m}$$

The base width is

$$W_B = x_j^{\text{BC}} - x_j^{\text{EB}} = 1.09 - 0.37 = 0.72 \mu\text{m}$$

7.10. A special twin-well (twin-tub) CMOS technology requires that the wells have precisely the same depth at the substrate concentration of $1 \times 10^{15} \text{ cm}^{-3}$, with arsenic used for the n-tub and boron used for the p-tub. A shallow implant dose of $1 \times 10^{14} \text{ cm}^{-2}$ is used for both and the slow diffusing arsenic is introduced first and partially driven-in. Then the boron is introduced and the rest of the anneal is performed until both junctions reach 2.5 microns. Calculate all of the drive-in times and temperatures used.

Answer:

To reach the same depth,

$$D_B t_2 = D_{\text{As}}(t_1 + t_2)$$

Both dopants are diffusing from a finite source, being implanted near the surface, so are characterized by a one-sided gaussian

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

For boron,

$$1 \times 10^{15} = \frac{10^{14}}{\sqrt{\pi Dt}} \exp\left(-\frac{(2.5 \times 10^{-4})^2}{4Dt}\right)$$

Iterating gives

$$D_B t_2 = 2.2 \times 10^{-9} \text{ cm}^2$$

If we assume a temperature of 1100°C for the high temperature well drive in, then

$$D_B^{1100\text{C}} = 1.0 \exp\left(-\frac{3.5}{kT}\right) = 1.5 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1}$$

so that

$$t_2 = \frac{2.2 \times 10^{-9}}{1.5 \times 10^{-13}} = 4 \text{ hours}$$

At this temperature of 1100°C,

$$D_{As}^{1100\text{C}} = 9.17 \exp\left(-\frac{3.99}{8.62 \times 10^{-5} \times (1100 + 273)}\right) = 2.1 \times 10^{-14} \text{ cm}^2 \text{ s}^{-1}$$

$$D_{As}(t_1 + t_2) = 2.2 \times 10^{-9} \text{ cm}^2$$

$$t_1 + t_2 = \frac{2.2 \times 10^{-9}}{2.1 \times 10^{-14}} = 5.3 \text{ hours}$$

$$t_1 = 5.3 - t_2 = 1.3 \text{ hours}$$

- 7.11. A process engineer on the day shift started a boron isolation diffusion for a structure in which the boron diffusion needs to penetrate completely through a 6 μm thick N-type epitaxial layer that is lightly doped with phosphorus ($N_D = 1 \times 10^{15} \text{ cm}^{-3}$) on a P type substrate ($N_A = 1 \times 10^{14} \text{ cm}^{-3}$). The purpose of the diffusion is to provide isolation between different N-type regions. The day shift engineer left no information on what he or she did. On a monitor wafer, you make measurements and find that probes to the N-type layer still conduct and the resistance of a patterned layer that is 100 μm long by 5 μm wide is 400 kΩ.**

The sheet resistance of the p-type diffusion measured using a 4-point probe is 250 Ω /square. What time and temperature are needed to complete the isolation diffusion?

Answer:

Given the resistance of 20 squares (100 x 5 microns) is 400 k Ω , we have the sheet resistance

$$\rho_s = 20 \times 10^3 \text{ Ohms / Square}$$

This is the resistance of a thin layer of uniformly doped epitaxial material of thickness x_j : the resistivity of the N-type material uniformly doped at $1 \times 10^{15} / \text{cm}^3$ is given from Table 1.18 in Chapter 1 as 4 Ohm-cm, so that

$$\rho = 4.0 = \rho_s \times x_j = 20 \times 10^3 x_j$$

$$x_j = 2.0 \mu\text{m}$$

Thus, the P-type layer has diffused down 4 microns into the epilayer. The sheet resistance of the p-type layer is 25 Ohms/Square, so the average conductivity of the layer (needed to use Irvin's curves) is

$$\bar{\sigma} = \frac{1}{\rho_s x_j} = \frac{1}{25 \times 4 \times 10^{-4}} = 100 \text{ Ohm} \cdot \text{cm}$$

Using Irvin's curves for a p-type Gaussian diffusion gives a surface concentration of

$$C_s = 4 \times 10^{19} / \text{cm}^3$$

The p-type diffusion is now completely characterized by its surface concentration, junction depth and sheet resistance. Our implicit assumption that it was a Gaussian seems to be valid - the surface concentration is below the solid solubility for boron and the junction depth is deep, suggesting a layer that has received a substantial drive-in. Using the formula for a one-sided surface Gaussian diffusion,

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

At the surface,

$$C_s = \frac{Q}{\sqrt{\pi Dt}}$$

so that at the junction depth of 4 microns,

$$1 \times 10^{15} = 4 \times 10^{19} \exp\left(-\frac{(4 \times 10^{-4})^2}{4Dt}\right)$$

giving

$$Dt = 3.8 \times 10^{-9} \text{ cm}^2$$

thus, the initial dose is

$$Q = C_s \sqrt{\pi Dt} = 4 \times 10^{19} \sqrt{\pi \times 3.8 \times 10^{-9}} = 4.3 \times 10^{15} \text{ cm}^{-2}$$

Knowing the initial dose, we can now calculate the Dt needed to form a junction 6um deep, using the one-sided Gaussian formula

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

$$1 \times 10^{15} = \frac{4.3 \times 10^{15}}{\sqrt{\pi Dt}} \exp\left(-\frac{(6 \times 10^{-4})^2}{4Dt}\right)$$

Iterating using an initial guess of $3.8 \times 10^{-9} \times (6/4)^2 = 8.55 \times 10^{-9}$ gives

$$\begin{aligned} Dt &= 8.55 \times 10^{-9} > 0 \\ Dt &= 8.90 \times 10^{-9} < 0 \\ Dt &= 8.80 \times 10^{-9} > 0 \\ Dt &= 8.85 \times 10^{-9} \text{ converged} \end{aligned}$$

Thus, the extra Dt required is $5.05 \times 10^{-9} \text{ cm}^2$

Picking a temperature like 1100C for the diffusion gives a boron diffusivity of

$$D_B^{1100\text{C}} = 1.0 \exp\left(-\frac{3.5}{kT}\right) = 1.5 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1}$$

Thus another 9.35 hours at 1100C are required at 1100°C.

7.12. A bipolar transistor is fabricated by implanting the base and emitter regions and then driving them in together. Boron is used for the base and arsenic for the emitter. The implants are $Q_B = 2 \times 10^{13} \text{ cm}^{-2}$ and $Q_E = 2 \times 10^{15} \text{ cm}^{-2}$. The

drive-in is 60 min at 1100 °C. Assume that the implants can be treated as delta functions at the surface. The substrate is 10^{15} cm^{-3} N type.

(a) Calculate and plot the resulting impurity profiles. Assume that the B and As diffusivities are the values from Table 7.5. You can neglect all second order effects like concentration dependent diffusion, E field effects etc.

(b) A parameter of interest in bipolar transistors is the base Gummel number which is the total dose (atoms cm^{-2}) in the base region under the emitter. Estimate this parameter for the device in part a).

Answer:

(a) Since only the Q's are given and fixed doses of dopants that have been introduced, we can assume that the subsequent diffusion will be a one-sided gaussian.

The equation describing each of the diffusion profiles will be

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

The diffusivity of the dopants is given by

$$\begin{aligned} D_B &= 0.95 \exp\left(-\frac{3.5}{(8.62 \times 10^{-5})(1100 + 273)}\right) + 0.05 \exp\left(-\frac{3.5}{kT}\right) \\ &= 1.5 \times 10^{-13} \text{ cm}^2 \text{ s}^{-1} \end{aligned}$$

$$\begin{aligned} D_{As} &= 0.011 \exp\left(-\frac{3.44}{kT}\right) + 31.0 \exp\left(-\frac{4.15}{kT}\right) \\ &= 2 \times 10^{-14} \text{ cm}^2 \text{ s}^{-1} \end{aligned}$$

In each case, the junction with the background doping occurs when

Boron Profile:

$$1 \times 10^{15} = \frac{2 \times 10^{13}}{\sqrt{\pi(1.5 \times 10^{-13})(3600)}} \exp\left(-\frac{x_j^2}{4(1.5 \times 10^{-13})(3600)}\right)$$

$$x_j = 1.16 \times 10^{-4} \text{ cm} = 1.16 \mu\text{m}$$

$$C_S = \frac{2 \times 10^{13}}{\sqrt{\pi(1.5 \times 10^{-13})(3600)}} = 4.8 \times 10^{17} \text{ cm}^{-3}$$

As Profile:

$$1 \times 10^{15} = \frac{2 \times 10^{15}}{\sqrt{\pi(2 \times 10^{-14})(3600)}} \exp\left(-\frac{x_j^2}{4(2 \times 10^{-14})(3600)}\right)$$

$$x_j = 0.595 \times 10^{-4} \text{ cm} = 0.595 \mu\text{m}$$

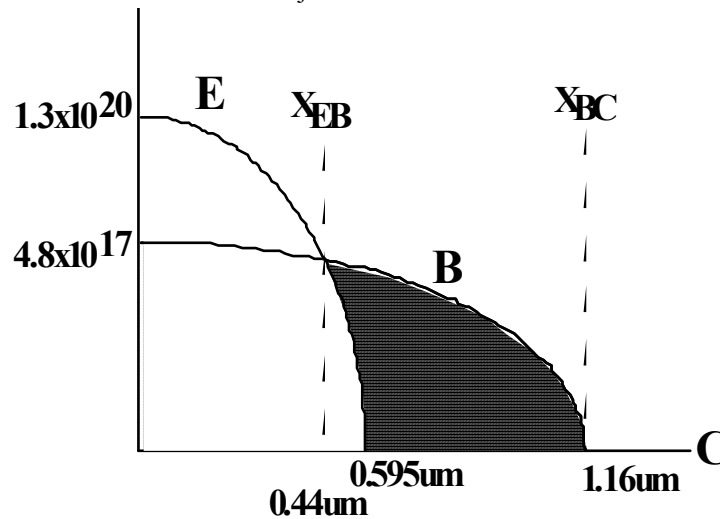
$$C_S = \frac{2 \times 10^{15}}{\sqrt{\pi(2.09 \times 10^{-14})(3600)}} = 1.3 \times 10^{20} \text{ cm}^{-3}$$

The emitter-base junction occurs when the Boron and Arsenic profile intersect

$$4.8 \times 10^{17} \exp\left(-\frac{x_j^2}{4(1.5 \times 10^{-13})(3600)}\right) = 1.3 \times 10^{20} \exp\left(-\frac{x_j^2}{4(2 \times 10^{-14})(3600)}\right)$$

Solving for x gives

$$x_j^{EB} = 0.44 \mu\text{m}$$



To find the base dose, we need to integrate the net dopant (the difference between the n and p doping) between the emitter-base junction and the base-collector junction. To simplify (we were asked to estimate in the problem), we will integrate between the emitter-base junction and infinity, because the background collector doping is so low.

We could simplify further and simply integrate the p-type doping in the base, ignoring the slight tail of the emitter that penetrates the base region.

Here we do the problem the first way, which will show us how good an approximation the second method would be.

The n-type doping that penetrates the base is given by the integral between the emitter-base junction at $0.44 \mu\text{m}$ and the emitter-collector junction at $0.595 \mu\text{m}$ (we will integrate to infinity, in effect ignoring the small amount of arsenic dose contained between $0.595 \mu\text{m}$ and infinity). We know

$$\text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-u^2) du$$

so the dose between u and infinity is given by

$$Q_{EB} = Q_{EMITTER} - \int_0^{x_{EB}} \frac{Q_{EMITTER}}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) dx$$

Let

$$u = \frac{x}{\sqrt{4Dt}}$$

$$dx = \sqrt{4Dt} du$$

When

$$x = 0, \quad u = 0$$

$$x = X_{EB}, \quad u = \frac{X_{EB}}{\sqrt{4Dt}}$$

Thus

$$\begin{aligned} Q_{EB} &= Q_{EMITTER} - \frac{Q_{EMITTER}}{\sqrt{\pi Dt}} \int_0^{X_{EB}/\sqrt{4Dt}} \exp(-u^2) \sqrt{4Dt} du \\ &= Q_{EMITTER} \left[1 - \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} \left(\text{erf}\left(\frac{X_{EB}}{\sqrt{4Dt}}\right) - \text{erf}(0) \right) \right] = Q_{EMITTER} \text{erfc}\left(\frac{X_{EB}}{\sqrt{4Dt}}\right) \end{aligned}$$

Similarly,

$$Q_{BC} = Q_{BASE} \text{erfc}\left(\frac{X_{EB}}{\sqrt{4Dt}}\right)$$

The difference gives the net doping in the base

$$Q_{\text{BASE}} = Q_{\text{BC}} - Q_{\text{EB}} = 2 \times 10^{13} \operatorname{erfc}\left(\frac{0.44\mu\text{m}}{0.468\mu\text{m}}\right) - 2 \times 10^{15} \operatorname{erfc}\left(\frac{0.44\mu\text{m}}{0.173\mu\text{m}}\right)$$

$$\begin{aligned}\therefore Q_{\text{BASE}} &= 2 \times 10^{13}(1 - 0.816) - 2 \times 10^{15}(1 - 0.9997) \\ &= 3.68 \times 10^{12} - 6 \times 10^{11} = 3.08 \times 10^{12} \text{ cm}^{-2}\end{aligned}$$

Note that we have derived the useful relationship

$$\operatorname{erfc}(u) = \frac{2}{\sqrt{\pi}} \int_u^{\infty} \exp(-u^2) du$$

and that the contribution from the emitter dopant past the emitter-base junction is small, but not quite negligible ($6 \times 10^{11} \text{ cm}^{-2}$).

7.13. All semiconductor devices are really “unstable” at device operating temperatures because dopants will continue to diffuse and hence, given long enough, profiles will change and devices will stop functioning correctly. How long would it take at a 150°C operating temperature for the base width in the device in problem 7.12 to double? Assume the As does not move, just the boron moves.

Answer:

The base width from the problem 7.12 is

$$W_B = 1.16 - 0.44 = 0.72\mu\text{m}$$

When the base width doubles, the base collector junction will be approximately at

$$X_{\text{BC}} = 0.44 + (2 \times 0.72) = 1.88\mu\text{m}$$

assuming that the emitter does not move and the emitter-base junction depth does not change much as a result.

The diffusivity of boron at 150°C is

$$\begin{aligned}D_B &= 0.95 \exp\left(-\frac{3.5}{(8.62 \times 10^{-5})(150 + 273)}\right) + 0.05 \exp\left(-\frac{3.5}{kT}\right) \\ &= 2 \times 10^{-42} \text{ cm}^2 / \text{sec}\end{aligned}$$

The junction depth is obtained from

$$1 \times 10^{15} = \frac{2 \times 10^{13}}{\sqrt{\pi(4.66 \times 10^{-42})t}} \exp\left(-\frac{(1.88 \times 10^{-4})^2}{4(4.66 \times 10^{-42})t}\right)$$

Iteratively solving for the time t (seconds) (by trial and error) gives

$$t \approx 10^{25} \text{ years}$$

This shows that devices are stable at the highest operating temperatures, at least in terms of dopant diffusion. The dopants won't diffuse around and affect the device characteristics at normal operating temperatures. (There may still be problems with device stability from mobile ionic charges in the gate oxide or from oxide reliability due to hot electron effects.)

7.14

An epitaxial layer is grown with uniform concentrations of B and P impurities ($B = 2 \times 10^{15} \text{ cm}^{-3}$, $P = 1 \times 10^{15} \text{ cm}^{-3}$) so that the net doping is p-type. The wafer is oxidized at 1000°C for 60 minutes. Approximately (use $2\sqrt{Dt}$ as a measure of the profile motion) sketch the resulting impurity profiles for both boron and phosphorus under the oxide which result from the redistribution that occurs.

Answer:

Phosphorus and boron have similar diffusion coefficients at 1000°C , approximately $1 \times 10^{-14} \text{ cm}^2\text{s}^{-1}$. The profiles in the substrate will change over a distance of a few \sqrt{Dt} , so that while the oxide grows the profile will change because of segregation to the growing oxide over a distance

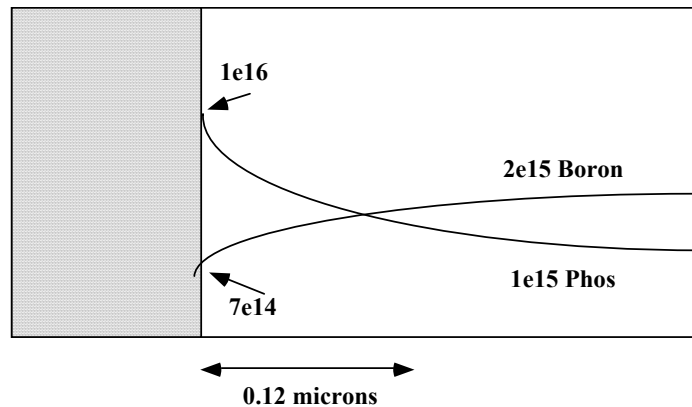
$$2\sqrt{1 \times 10^{-14} \times 3600} = 1.2 \times 10^{-5} \text{ cm} = 0.12 \text{ } \mu\text{m}$$

The surface concentration will tend towards the value in the oxide, so is related to the bulk concentration by the segregation coefficient

$$\left(\frac{C_S}{C_B}\right)_B = 0.3$$

$$\left(\frac{C_S}{C_B}\right)_P = 10$$

so that an approximate sketch of the dopant distribution after oxidation is as follows:



- 7.15. A silicon wafer is uniformly doped with boron ($2 \times 10^{15} \text{ cm}^{-3}$) and phosphorus ($1 \times 10^{15} \text{ cm}^{-3}$) so that it is net P type. This wafer is then thermally oxidized to grow about $1 \mu\text{m}$ of SiO_2 . The oxide is then stripped and a measurement is made to determine the doping type of the wafer surface. Surprisingly it is found to be N type. Explain why the surface was converted from P to N type. Hint: Consider the segregation behavior of dopants when silicon is oxidized.

Answer:

The boron segregates preferentially into the growing oxide, thus depleting the surface concentration in the silicon. The phosphorus on the other hand preferentially segregates (piles-up) on the silicon side of the interface. Both of these effects act in the same direction and tend to make the surface of the silicon more N-type.

It is for this reason that a P-type “channel stop” implant is almost always needed under a locally oxidized lightly doped P-type region, to prevent depletion of the P-type dopant in the substrate and in the worst case to prevent an N-type channel from forming.

- 7.16 An engineer wants to use analytical solutions to diffusion equations in a programmable calculator to make rapid estimates for process changes on junction depths. Consider the following possible diffusion regimes. Which of them are most appropriate for analytical solutions (i.e. which would minimize E-field or concentration dependent effects). Explain.

Answer:

- (a) high temperatures

High temperature makes the material appear more intrinsic for given doping levels, so should help minimize E-field or concentration dependent effects.

(b) low temperatures

Low temperature makes the material appear more extrinsic for given doping levels, so will lead to inaccurate solutions.

(c) long times

Long times - for a fixed dose, longer anneal times will reduce the surface concentration and tend to make the material intrinsic, so will help minimize the overall effect of E-fields or concentration dependent diffusion on the profiles.

For a constant surface concentration type diffusion (erf type profiles), longer times won't help.

(d) short times

Short times - won't help if the material is initially extrinsic. Short times also cause large TED effects.

Thus the profiles will likely be most inaccurate for low temperature, short time diffusions.

7.17. A shallow phosphorus implant with a dose of $1 \times 10^{14} \text{ cm}^{-2}$ is covered in some regions with a deposited layer of inert nitride. An anneal is performed at 1000°C in dry O_2 and a junction depth below the original surface is measured in the inert region of $0.5 \mu\text{m}$ and of $1.2 \mu\text{m}$ under the oxidizing region. What is that diffusivity enhancement that the phosphorus in the oxidizing region experiences. If f_i for phosphorus is 0.9 and I-V recombination is efficient at 1000°C , what is the interstitial supersaturation that is generated by the oxidation.

Answer:

We can assume that both profiles are modeled by Gaussian profiles, since a fixed dose was introduced at the surface and the junctions are deep relative to the initial implant. A one-sided Gaussian solution is valid

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

Assuming a typical substrate doping of 10^{15} cm^{-3} , then in the nitride (masked or inert) region,

$$10^{15} = \frac{10^{14}}{\sqrt{\pi Dt}} \exp\left(-\frac{(0.5 \times 10^{-4})^2}{4Dt}\right)$$

Iterating gives

$$D_N t = 7.1 \times 10^{-11} \text{ cm}^2$$

$$10^{15} = \frac{10^{14}}{\sqrt{\pi Dt}} \exp\left(-\frac{(1.2 \times 10^{-4})^2}{4Dt}\right)$$

Iterating gives

$$D_O t = 4.55 \times 10^{-10} \text{ cm}^2$$

The diffusion enhancement is given by

$$\frac{D}{D^*} = \frac{D_N t}{D_O t} = \frac{4.55 \times 10^{-10}}{7.1 \times 10^{-11}} = 6.4$$

If I-V recombination is efficient, then

$$C_I C_V = C_I^* C_V^*$$

$$\frac{D}{D^*} = f_I \frac{C_I}{C_I^*} + f_V \frac{C_V}{C_V^*}$$

$$6.4 = 0.9 \frac{C_I}{C_I^*} + 0.1 \frac{C_I^*}{C_I}$$

$$0.9 \left(\frac{C_I}{C_I^*} \right)^2 - 6.4 \frac{C_I}{C_I^*} + 0.1 = 0$$

$$\frac{C_I}{C_I^*} = 7.09$$

7.18. Rapid Thermal Annealing (RTA) systems are becoming common for activating dopants. Silicon has quite a high thermal diffusivity of $0.88 \text{ cm}^2 \text{ sec}^{-1}$, which describes how fast heat flows through silicon. Calculate the time required for a silicon wafer of thickness $500 \text{ }\mu\text{m}$ to reach a constant temperature if a thin surface layer absorbs all the incident light and quickly reaches a steady state

temperature of T_s i.e. when does the center reach $T/T_s = 0.5$, since the wafer is heated from both sides.

Answer:

Heat flow and particle flow are analogous, and here we rely on the fact that the same physical equations have the same solutions. The problem is set up like a constant surface concentration that is maintained constant during the diffusion, and thus the error function solution from a constant surface concentration is appropriate. There is a direct analogy between temperature (T) and concentration (C).

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

with a solution of

$$T(x,t) = T_s \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$

We are asked when $T/T_s = 0.5$ at the center of the wafer where $x = 250\mu\text{m}$ so that

$$0.5 = 1 - \operatorname{erf}\left(\frac{250 \times 10^{-4}}{2\sqrt{0.88t}}\right)$$

$$\operatorname{erf}(x) = 0.5 \Rightarrow x = 0.48$$

$$\frac{250 \times 10^{-4}}{2\sqrt{0.88t}} = 0.48$$

$$\text{Hence, } t = 0.77 \text{ millisecond}$$

Thus, the wafer heats up in the millisecond timeframe during rapid thermal anneals.

7.19. By using Eqn. 7.72 ($C_I C_V = C_I^* C_V^*$) and letting $C_V/C_V^* = x$, solve for x and use the fact that the solution must be real and positive to derive Eqn. 7.82. Using the fact that $f_I = 1 - f_V$ and completing the square simplifies the solution.

Alternatively, use Eqn 7.72 to substitute for C_V/C_V^* in Eqn. 7.66 and take the derivative with respect to C_I/C_I^* and find an extreme (maximum or minimum).

Answer:

For a given f we have

$$\Delta R = \frac{D}{D^*} = f_I \frac{C_I}{C_I^*} + f_V \frac{C_I^*}{C_I}$$

when

$$\frac{C_V}{C_V^*} = \frac{C_I}{C_I^*}$$

The extreme (minimum or maximum) value occurs when

$$\frac{\partial \Delta R}{\partial \left(\frac{C_I}{C_I^*} \right)} = f_I - f_V \frac{1}{\left(\frac{C_I}{C_I^*} \right)^2} = 0$$

so that

$$\frac{C_I}{C_I^*} = \sqrt{\frac{f_V}{f_I}}$$

Thus,

$$\Delta R = f_I \sqrt{\frac{f_V}{f_I}} + f_V \sqrt{\frac{f_I}{f_V}} = 2\sqrt{f_I f_V}$$

The retardation is the largest when

$$f_V(1 - f_V) = \frac{\Delta R^2}{4}$$

so that

$$f_V = \frac{1 \pm \sqrt{1 - \Delta R^2}}{2}$$

The observed retardation is not necessarily the maximum retardation, so that the lower bound is actually

$$f_V = 0.5 + 0.5\sqrt{1 - \Delta R^2}$$

7.20. Fig. 7.38 shows that a wet oxidation produces a significantly higher C_I/C_I^* than does a dry O_2 oxidation. Explain quantitatively why this should be the case.

Answer:

Because

$$C_I \propto \sqrt{\frac{dx}{dt}}$$

a faster oxidation rate produces a higher interstitial supersaturation. Thus, wet oxidation produces a higher C_I/C_I^* than dry oxidation, for the same time at the same temperature.

Chapter 8 Problems

- 8.1. Arsenic is implanted into a lightly doped p-type Si substrate at an energy of 75keV. The dose is $1 \times 10^{14} / \text{cm}^2$. The Si substrate is tilted 7° with respect to the ion beam to make it appear amorphous. The implanted region is assumed to be rapidly annealed so that complete electrical activation is achieved. What is the peak electron concentration produced?**

Answer:

From Fig. 8-3, the range and standard deviation for 75 keV arsenic are

$$R_P = 0.05\mu\text{m} \quad \Delta R_P = 0.02\mu\text{m}$$

The peak concentration is

$$C_P = \frac{Q}{\sqrt{2\pi} \Delta R_P} = \frac{1 \times 10^{14}}{\sqrt{2\pi} (0.02 \times 10^{-4})} = 2 \times 10^{19} \text{ cm}^{-3}$$

Assuming all the dose is active, then the peak electron concentration is equal to the peak dopant concentration.

- 8.2. An engineer worried about avoiding punchthrough in an NMOS device decides to perform a deep punchthrough implant while keeping the surface concentration at a maximum value of $1 \times 10^{17} \text{ cm}^{-3}$. What is the maximum implant dose that would be suitable if the peak of the profile is aligned at twice the depth of the deep source/drain contact junction in a 180 nm technology? (See NTRS Table 7.1 for parameter values.)**

Answer:

This is an NMOS device with a highly N doped source and drain. We want to prevent the fields from the drain from punching through (reaching through) to the source and inadvertently turning on the transistor. One way to do this is to increase the substrate (p-type) doping. However, this would also shift the threshold voltage of the device. A better way is to insert a deeper p-type implant to prevent this punchthrough.

From the NTRS, the contact junction depth is 70-140nm, so we can choose 100nm as a reasonable junction depth for the deep source/drain. We want the peak of the boron doping to be at 200nm=0.2microns, so an energy of 60keV is appropriate from the range graph. This has a corresponding $\Delta R_P = 0.07\mu\text{m}$ from the graph.

$$C(0) = C_B = C_P \exp\left(-\frac{(x - R_P)^2}{2\Delta R_P^2}\right)$$

$$1 \times 10^{17} = C_P \exp\left(-\frac{(0.2 \times 10^{-4})^2}{2 \times (0.07 \times 10^{-4})^2}\right)$$

$$C_P = 5.9 \times 10^{18} \text{ cm}^{-3}$$

We know that

$$Q = \sqrt{2\pi} \Delta R_P C_P$$

hence, the dose that causes the surface concentration to reach 10^{17} cm^{-3} is

$$Q = 1 \times 10^{14} \text{ cm}^{-2}$$

8.3. We want to design an implant step which will implant phosphorus ions through 50nm of SiO₂ into an underlying silicon substrate such that the peak concentration in the substrate is $1 \times 10^{17} \text{ cm}^{-3}$ and the concentration at the SiO₂/Si interface is $1 \times 10^{15} \text{ cm}^{-3}$. What energy and dose would you use to achieve these conditions. Assume that the stopping power of SiO₂ is the same as that of silicon. Neglect channeling effects.

Answer:

We have an equation relating concentration to peak doping:

$$C(x) = C_P \exp\left(-\frac{(x - R_P)^2}{\Delta R_P^2}\right)$$

So that

$$2 \ln(100) = \frac{(0.05 - R_P)^2}{\Delta R_P^2}$$

Alternatively, we can use the masking formula

$$x_m = R_P + \Delta R_P \sqrt{2 \log \frac{C_P}{C_B}} \approx R_P + 3.0 \Delta R_P$$

since

$$\frac{C_P}{C_B} = 100$$

Thus, the oxide-silicon interface lies approximately $3\Delta R_p$ away from the peak of the implanted profile. We know the energy of the phosphorus implant must be at least 30keV to penetrate 50nm. By trial and error from the graph of the range for phosphorus we have

$$\begin{array}{llll} 80\text{keV} & R_p = 0.11 & \Delta R_p = 0.040 & \text{too small} \\ 200\text{keV} & R_p = 0.23 & \Delta R_p = 0.07 & \text{too small} \\ 400\text{keV} & R_p = 0.46 & \Delta R_p = 0.14 & \text{OK} \end{array}$$

where we extrapolated linearly from the end of the graph.

This gives a value for $\Delta R_p = 0.14\mu\text{m}$ and hence we can find the dose

$$Q = \sqrt{2\pi} \Delta R_p C_p = 3.5 \times 10^{12} \text{ cm}^{-2}$$

8.4. How thick does a mask have to be to reduce the peak doping of an implant by a factor of 10,000 at the mask/substrate boundary. Provide an equation in terms of the Range and the Standard Deviation of the implant profile.

Answer:

We want to reduce the peak doping N_p^* in the mask at range R_p^* by 10,000 at the mask/substrate boundary. We will use the equation which describes the profile of an implant in a mask layer

$$N^*(d) = N_p^* \left[\exp - \frac{(d - R_p^*)^2}{2\Delta R_p^{*2}} \right]$$

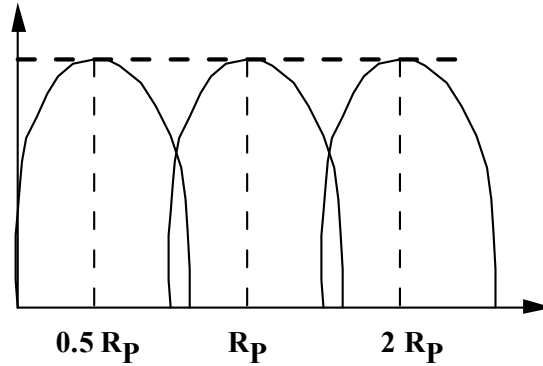
When

$$\frac{N^*(d)}{N_p^*} = 10^{-4}$$

we have

$$d = R_p^* + 4.3\Delta R_p^*$$

8.5. In a particular application, it is important to produce a fairly flat profile over an extended distance by ion implantation, as indicated below. One way to do this is to superimpose several implants at different energies. If phosphorus implants with energies of $0.5R_p$, R_p , $2R_p$ are used with a dose of $1 \times 10^{14} \text{ cm}^{-2}$ in the middle peak, approximately what doses should be used in the adjacent peaks if the initial peak concentrations are to be the same.



Answer:

Empirically, from the implant tables, the range R_P and standard deviation ΔR_P scale similarly with energy. Thus, to maintain the same peak concentration given by

$$C_P = \frac{Q}{\sqrt{2\pi} \Delta R_P}$$

we need to scale the dose Q with ΔR_P which empirically increases with R_P , so the doses should be

$$Q_{0.5R_P} = 0.5Q$$

$$Q_{1.5R_P} = 1.5Q$$

8.6. The equations below provide a reasonable analytical description for some of the diffusion processes indicated schematically in the diagrams on the following page. Put the equation number (a-f) on each figure that is the best match. Equations may be reused, or multiple equations may describe the same figure. A brief explanation is required for each figure.

(a) $C(x, t) = \frac{C}{2} \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$

(b) $C(x, t) = \frac{Q}{\sqrt{\pi(D_1 t_1 + D_2 t_2)}} \exp\left(-\frac{x^2}{4(D_1 t_1 + D_2 t_2)}\right)$

(c) $C(x, t) = \frac{Q}{2\sqrt{\pi Dt}} \exp\left[-\left(\frac{x}{2\sqrt{Dt}}\right)^2\right]$

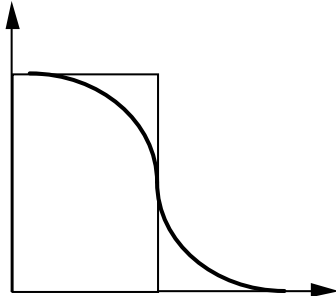
(d) $C(x, t) = C \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$

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

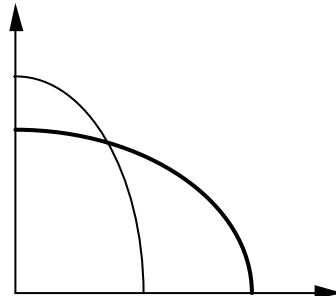
(f) $C(x, t) = \frac{Q}{\sqrt{2\pi(\Delta R_P^2 + 2Dt)}} \exp\left(-\frac{(x - R_P)^2}{2(\Delta R_P^2 + 2Dt)}\right)$

Answer:

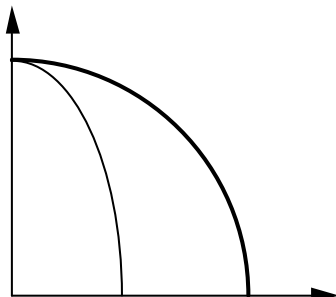
- (a) Describes a complementary error function in-diffusion, where the surface concentration is half the peak concentration in the background.
- (f) Describes a one-sided Gaussian profile that has some additional diffusion time.
- (g) Describes a delta function that diffuses as a Gaussian profile in an infinite medium.
- (h) Describes a complementary error function diffusion from a constant surface concentration.
- (i) Describes a one-sided Gaussian diffused profile near a reflecting surface.
- (j) Describes an implanted profile in an infinite medium that has some extra diffusion time.



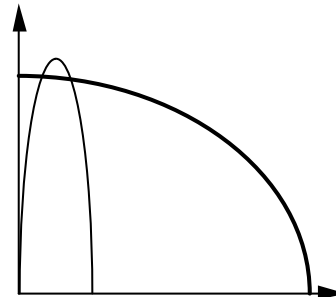
$$(a) \quad C(x, t) = \frac{C}{2} \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$



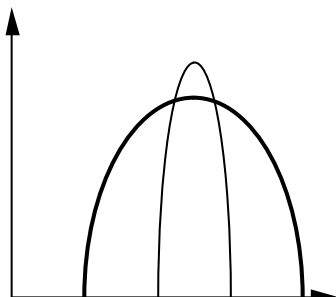
$$C(x, t) = \frac{Q}{\sqrt{\pi(D_1 t_1 + D_2 t_2)}} \exp\left(-\frac{x^2}{4(D_1 t_1 + D_2 t_2)}\right)$$



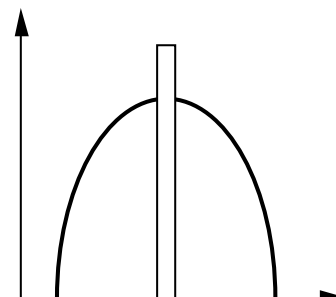
$$C(x, t) = C \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$



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



$$C(x, t) = \frac{Q}{\sqrt{2\pi(\Delta R_p^2 + 2Dt)}} \exp\left(-\frac{(x - R_p)^2}{2(\Delta R_p^2 + 2Dt)}\right)$$



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

8.7. An 80 keV, $5 \times 10^{13} \text{ cm}^{-2}$ boron implant is performed into bare silicon. A subsequent anneal at 950°C ($D_B = 4.24 \times 10^{-15} \text{ cm}^2 \text{ sec}^{-1}$) is performed for 60 min.

(a) Can the annealed profile evolution be described by the formula

$$C(x, t) = \frac{Q}{\sqrt{\pi D_B t}} \exp\left(-\frac{x^2}{4 D_B t}\right) \quad ?$$

(b) Assume that all the dopant remains in the silicon and that none evaporates to the ambient. By considering a virtual or imaginary image profile on the ambient side of the interface (a reflecting boundary), calculate the surface concentration after a 60 min, 950°C anneal.

Answer:

(a) This formula applies only when the initial dose is introduced as a delta-function near the surface, or receives sufficient diffusion (Dt) that the delta-function approximation is reasonably valid.

The boron implant has a range of $0.15 \mu\text{m}$ and a standard deviation of $0.08 \mu\text{m}$. $\Delta R_p^2 = 6.4 \times 10^{-11} \text{ cm}^2$ and $2Dt = 3 \times 10^{-11} \text{ cm}^2$ are comparable, so the profile does not receive a large amount of diffusion compared to the initial distribution. The simple delta-function approximation is not valid.

(b) The formula for diffusion in an infinite medium is

$$C(x, t) = \frac{Q}{\sqrt{2\pi(\Delta R_p^2 + 2Dt_{\text{eff}})}} \exp\left(-\frac{(x - R_p)^2}{2(\Delta R_p^2 + 2Dt_{\text{eff}})}\right)$$

We can assume an identical imaginary (or virtual) diffusion profile from the ambient side of the interface, contributing to the surface concentration and hence doubling it.

$$C = \frac{5 \times 10^{-13}}{\sqrt{2\pi(6.4 \times 10^{-11} + 3 \times 10^{-11})}} \exp\left(-\frac{(0.15 \times 10^{-4})^2}{2(6.4 \times 10^{-11} + 3 \times 10^{-11})}\right)$$

$$C = 6.2 \times 10^{17} \text{ cm}^{-3}$$

There is an equal contribution from the virtual gaussian on the other side of the interface, giving a surface concentration of

$$C = 1.2 \times 10^{18} \text{ cm}^{-3}$$

- 8.8. A $1 \times 10^{14} \text{ cm}^{-2}$ phosphorus implant through a 200 nm SiO_2 mask layer is performed so the peak concentration is at the silicon/ SiO_2 interface. An anneal is then performed for 30 min at 1000°C . Calculate the location of the junction with the substrate doped at $1 \times 10^{15} \text{ cm}^{-3}$. Assume no diffusion in the masking layer and ignore any segregation effects. Assume the same range statistics for SiO_2 and Si.**

Answer:

The key in this question is to pick the correct formula that applies. The peak of the implant is at the Si/ SiO_2 interface and diffusion occurs only on the silicon side of the interface. It is as if a delta function were introduced at the surface (with dose $Q_{\text{Si}} = 5 \times 10^{13} \text{ cm}^{-2}$) and diffused for some time equivalent to moving the profile a distance ΔR_p , then diffused for an extra 30 minutes at 1000°C .

We assume an imaginary diffusion in SiO_2 as well, then we use Eqn. 8.12 since the peak is at the interface, even though the diffusion is one-sided.

$$C_B = \frac{2 \times Q_{\text{Si}}}{\sqrt{2\pi(\Delta R_p^2 + 2Dt_{\text{eff}})}} \exp\left(-\frac{(x - R_p)^2}{2(\Delta R_p^2 + 2Dt_{\text{eff}})}\right)$$

From the range tables for phosphorus (middle curve)

$$R_p = 2000\text{\AA} = 0.2\mu\text{m} \Rightarrow E = 170\text{keV} \Rightarrow \Delta R_p = 0.065\mu\text{m}$$

The phosphorus diffusivity is

$$D_p = 4.7 \exp\left(-\frac{3.68}{kT}\right) = 1.3 \times 10^{-14} \text{ cm}^2 \text{ sec}^{-1}$$

$$10^{15} = \frac{2 \times 5 \times 10^{13}}{\sqrt{2\pi\left((0.065 \times 10^{-4})^2 + 2 \times 1.3 \times 10^{-14} \times 30 \times 60\right)}} \exp\left(-\frac{(x - 0.2 \times 10^{-4})^2}{2(\Delta R_p^2 + 2Dt_{\text{eff}})}\right)$$

$$x = 0.58 \times 10^{-4} \text{ cm} = 0.58\mu\text{m}$$

8.9. Phosphorus is implanted at 50keV with a dose of $1 \times 10^{14} \text{ cm}^{-2}$. Calculate the junction depth where the phosphorus meets the substrate (p-type, $1 \times 10^{15} \text{ cm}^{-3}$) after (a) a 900°C , 3 hour anneal and (b) a 1200°C , 3 hour anneal.

Answer:

The diffusivity of phosphorus at each of these temperatures, based on values calculated under intrinsic conditions is

$$D_P^{1200} = 1.31 \times 10^{-12} \text{ cm}^2\text{s}^{-1}$$

$$D_P^{900} = 7.7 \times 10^{-16} \text{ cm}^2\text{s}^{-1}$$

Using the range tables, we get for 50 keV phosphorus

$$\begin{aligned} R_P &= 750 \text{ Angstroms} \\ \Delta R_P &= 275 \text{ Angstroms} \end{aligned}$$

The approximate distance the profile will move at each temperature is given by

$$\begin{aligned} \sqrt{Dt}|_{900} &= 288 \text{ Angstroms} \\ \sqrt{Dt}|_{1200} &= 11,899 \text{ Angstroms} \end{aligned}$$

Comparing the \sqrt{Dt} with the ΔR_P is the key step in this problem.

At 900°C it will diffuse like an implanted gaussian profile with some small additional Dt , while at 1200°C it will appear as a delta function of dopant introduced near the surface.

At 900°C

$$C(x, t) = \frac{Q}{\sqrt{2\pi(\Delta R_P^2 + 2Dt)}} \exp\left(-\frac{(x - R_P)^2}{2(\Delta R_P^2 + 2Dt)}\right)$$

$$C(x, t) = 1 \times 10^{15} =$$

$$\frac{1 \times 10^{14}}{\sqrt{2\pi[(275 \times 10^{-8})^2 + 2(288 \times 10^{-8})^2]}} \exp\left(-\frac{(X_J - 750 \times 10^{-8})^2}{2[(275 \times 10^{-8})^2 + 2(288 \times 10^{-8})^2]}\right)$$

$$X_J = 2835 \text{ Angstroms}$$

At 1200°C

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

$$C(x, t) = 1 \times 10^{15} = \frac{1 \times 10^{14}}{\sqrt{\pi(1.41 \times 10^{-8})}} \exp\left(-\frac{X_J^2}{4(1.41 \times 10^{-8})}\right)$$

$$X_J = 5.89 \text{ microns}$$

Note that the answer by simply repeating the formulation from part (a) with the 1200°C numbers gives a numerical answer that is not too much different (5.6 μm). However, it misses the point of the question, which is not simply to calculate the same formula with new numbers, but involves a choice of the most appropriate equation to describe the process. The formulation in part (a) would get completely wrong answers in some cases (e.g. dose or dose penetration beyond a certain distance) if used in part (b).

- 8.10. An implant machine for 300 mm wafers is required to have a throughput of 60 wafers per hour. What beam current is required in order to implant a source/drain region in a CMOS device with a dose of $1 \times 10^{15} \text{ cm}^{-2}$?**

Answer:

The dose Q is given by

$$Q = \frac{\int I \, dt}{q \, A}$$

where the area A is 30 cm x 30 cm (normally a square pattern is used) and $q = 1.6 \times 10^{-19} \text{ C}$. Hence

$$I = 2.4 \text{ mA}$$

- 8.11. In the ion implantation process, positively charged ions impact on the semiconductor surface. Normally these ions are neutralized by capturing an electron from the conducting substrate. However, when the mask is an insulator like SiO_2 , the charge on the ions may not be neutralized as easily. Consider the case where a dose Q is implanted into the surface of an SiO_2 layer (assume all the charge resides at the oxide surface). Further assume that the**

oxide can withstand an electric field of 10^7 Vcm^{-1} before it breaks down. What implant dose Q is required to cause electrical failure of the mask? That is, what dose will cause a field of 10^7 Vcm^{-1} across the oxide.

Answer:

$$E = \frac{V}{d} = \frac{Q/C}{d} = \frac{Q}{\epsilon/d} = \frac{Q}{\epsilon}$$

The charge $Q = nq$ is the product of the number of ions and the charge on each ion, so that the number of ions to cause breakdown is

$$n = \frac{\epsilon E}{q} = \frac{(3.9)(8.854 \times 10^{-14})(10^7)}{1.6 \times 10^{-19}} = 2.16 \times 10^{13} \text{ cm}^{-2}$$

Note that this is independent of the oxide thickness. It is also in the range of typical implants and would be a real problem. To avoid the charging problem, the wafer surface is sometimes flooded with electrons to minimize charging during implantation.

8.12. An engineer investigating solid phase epitaxial regrowth after amorphizing ion implants of various species (P, B, Si, Ge, As, Sb) makes the following observations:

- 1. N-type dopants of very different size or atomic radius (e.g. antimony versus phosphorus) show identical regrowth rates, approximately an order of magnitude faster than the regrowth of silicon implanted and amorphized with silicon ions.**
- 2. P-type and N-type ion implanted regions have much faster regrowth rates than Si or Ge implanted regions, although they are not identical.**
- 3. B-doped regions compensated with an equal dose of an arsenic implant show identical regrowth rates to Ge implanted and amorphized regions.**

Construct a unified physical explanation of all three phenomena, considering possibilities such as size or stress effects, dopant charge or electric field effects, or point-defect based effects (no calculation required).

Answer:

The first observation (large versus small atoms) indicates that size or stress effects play no role in SPE.

The second observation (positive versus negative charge on dopant atoms) shows that dopant charge or E-field effects are not important, if switching the charge from positive to negative has no effect on regrowth rate.

The final observation shows that counter-doped material which is electrically intrinsic behaves just as intrinsic material.

Taking all these observations together, the most likely physical explanation is point defect related. Vacancy concentrations are dramatically increased in p or n-type material above the concentrations in intrinsic material. The Fermi level in counter doped (equal concentrations of p and n type dopant) is identical to that in intrinsic silicon, so the vacancy populations would also be the same. The vacancy concentrations need not be identical in p or n-type material, even if the doping concentrations are identical, because the charge state of the positive vacancy may be different from the charge state of the negative vacancy.

Thus, if SPE is dependent on the point defect concentrations (i.e. vacancies make it easier to move dopant atoms around and allow recrystallization to occur), then the SPE would be dependent on Fermi level and hence on point defect concentrations in the manner noticed in the experiment.

8.13. In two separate experiments, As and then B are implanted through a thin SiO₂ layer into the underlying substrate. As a result of the implantation, some of the oxygen atoms in the SiO₂ layer are knocked into the silicon substrate. Would you expect the As or the B to produce more oxygen knock-ons? Why?

Answer:

We will assume that the As and B are implanted at the same energy. The lighter B atom will slow down mainly by electronic stopping, which will not displace many atoms in the thin oxide layer. On the other hand, the heavy As atom will slow down mainly by nuclear collisions, which will knock on many of the oxygen atoms in the oxide layer. Thus, As produces more oxygen knock-ons than boron.

- 8.14. A boron implant is performed into silicon at 100 keV. The boron beam is aligned with the silicon crystal so that channeling is present. Estimate the range of the channeled boron profile, by considering that electronic stopping is the only mechanism for slowing the boron ions.**

Answer:

If the beam is aligned with the channels, then nuclear stopping will be minimized and electronic stopping will dominate. The range will be

$$R = \frac{1}{N} \int_0^E \frac{dE}{S_e(E)} = \frac{1}{N} \int_0^E \frac{dE}{k\sqrt{E}}$$

The value of k is given by $k = 0.2 \times 10^{15} \text{ eV}^{1/2} \text{ cm}^2$, so that the range is given by

$$R \approx 20\sqrt{E} \text{ Angstroms} \approx 6300 \text{ \AA}$$

This is about twice the range from the range tables, because there is a nuclear stopping component which contributes along with the electronic stopping. However, the calculation does indicate the range of a well-channeled ion.

- 8.15. An amorphizing implant is performed using a high dose of arsenic ($5 \times 10^{15} \text{ cm}^{-2}$, 200 keV) and TEM cross section images indicate complete amorphization to a depth of 50 nm. If solid phase epitaxial (SPE) regrowth instantly removes all of the damage in the amorphous region, calculate the fraction of the “+1” implanted dose that is available for TED.**

Answer:

Only the atoms past the amorphous interface can contribute to the “+1” fraction that remains after SPE regrowth. This is equivalent in form to a masking problem, so we can use the formula

$$Q_p = \frac{Q}{2} \operatorname{erfc} \left(\frac{d - R_p^*}{\sqrt{2\Delta R_p^*}} \right)$$

200keV arsenic has a range of 0.13 μm and a standard deviation of 0.05 μm .

$$Q_p = \frac{5 \times 10^{15}}{2} \operatorname{erfc} \left(\frac{(0.05 - 0.13) \times 10^{-4}}{\sqrt{2 \times 0.05 \times 10^{-4}}} \right)$$

$$Q_p = 2.5 \times 10^{15} \operatorname{erfc}(-1.13) = (1 - \operatorname{erf}(-1.13)) = (1 + \operatorname{erf}(1.13)) = 4.72 \times 10^{15} \text{ cm}^{-2}$$

Thus the fraction of the +1 dose that is available for TED is 94.4%.

8.16. Calculate the change in junction depth for a 40 keV boron threshold adjust implant of $5 \times 10^{13} \text{ cm}^{-2}$ annealed at 750°C in a furnace or at 1000°C in an RTA for a time just long enough to remove all the damage that causes TED. Assume a uniform well doping of $5 \times 10^{16} \text{ cm}^{-3}$.

Answer:

The primary formula we will use gives the enhancement in diffusivity while TED lasts

$$Dt_{\text{eff}} = D^* \frac{I}{I^*} \tau_{\text{enh}}$$

For 40 keV boron,

$$R_p \approx 0.14 \mu\text{m} \quad \Delta R_p \approx 0.055 \mu\text{m}$$

The initial junction is given by

$$C_B = \frac{Q_{\text{Si}}}{\sqrt{2\pi\Delta R_p^2}} \exp\left(-\frac{(x - R_p)^2}{2\Delta R_p^2}\right)$$

$$x = 0.3 \mu\text{m}$$

$$D_B = 0.76 \exp\left(-\frac{3.46}{kT}\right)$$

At 750°C , the equilibrium boron diffusivity is

$$D_B = 6.9 \times 10^{-18} \text{ cm}^2 \text{ sec}^{-1}$$

and it is enhanced by 7500 (Fig. 8.38) for a time of 400 seconds (Fig. 8.40)

giving

$$Dt_{\text{eff}}^{750^\circ\text{C}} = 2.07 \times 10^{-11} \text{ cm}^2$$

At 1000°C , the equilibrium boron diffusivity is

$$D_B = 1.5 \times 10^{-14} \text{ cm}^2 \text{ sec}^{-1}$$

and it is enhanced by 400 for 0.2 seconds

$$Dt_{\text{eff}}^{1000^\circ\text{C}} = 1.2 \times 10^{-12} \text{ cm}^2$$

Clearly, the effective Dt is higher at 750°C than at 1000°C , and we can calculate the new junction depths from

$$x = R_p + \sqrt{-2(\Delta R_p^2 + 2Dt_{\text{eff}}) \ln \left[\frac{C_B}{Q} \sqrt{2\pi(\Delta R_p^2 + 2Dt_{\text{eff}})} \right]}$$

$$x = 0.14 + 0.23 = 0.37 \mu\text{m} \text{ at } 750^\circ\text{C}$$

$$x = 0.14 + 0.17 = 0.31 \mu\text{m} \text{ at } 1000^\circ\text{C}$$

The 1000°C is deeper by

$$\Delta x_j = 0.06 \mu\text{m}$$

8.17. The diffusion of a buried antimony layer is only 20% of its normal diffusion in an inert ambient after a phosphorus implant is performed at the surface of a silicon wafer. The phosphorus diffusion coefficient is itself enhanced by a factor of 4 (400%).

(a) Suggest a qualitative reason for this observation.

(b) Assuming that the phosphorus diffuses completely by an interstitial mechanism, calculate how much of the antimony diffusion is mediated by interstitials and how much by vacancies.

Answer:

(a) The phosphorus implant creates an equivalent dose of +1 interstitials, which diffuse into the substrate and enhance the diffusion of the phosphorus itself. The interstitials also recombine with vacancies and depress the vacancy concentration, leading to retarded diffusion of the buried antimony layer.

(b) The antimony retardation is given by

$$\left. \frac{D}{D^*} \right|_{\text{sb}} = \Delta R = 0.2$$

If phosphorus diffuses completely by an interstitial mechanism, its diffusion enhancement represents the interstitial supersaturation C_I / C_I^*

$$\left. \frac{D}{D^*} \right|_p = \Delta E = 4.0$$

There are several formulas to use to obtain the fraction of the antimony that diffuses by vacancies:

$$f_V \gg 1 - 0.2 = 0.8$$

A stricter bound uses the fact that the phosphorus diffusion is simultaneously enhanced

$$f_V > 1 - \frac{0.2}{4.0} = 0.95$$

while the expression in Eqn. 7-73 uses only the antimony retardation, but assumes that interstitial and vacancy recombination is at equilibrium

$$f_V > 0.5 + 0.5\sqrt{1 - (0.2)^2} = 0.989$$

8.18. A phosphorus implant is performed into a bare p-type silicon wafer with a background doping of $1 \times 10^{15} \text{ cm}^{-3}$, at an energy of 50 keV and a dose of

$1 \times 10^{14} \text{ cm}^{-2}$. After an anneal at 900°C for 10 min, the junction depth is measured to be $0.25 \text{ }\mu\text{m}$. Calculate the enhancement in the phosphorus diffusion coefficient that was caused by the implantation damage.

Answer:

50 keV Phosphorus has a range of 70 nm and a standard deviation of 27.5 nm.

The final junction depth is large compared to the initial R_p and ΔR_p , so the proper approach is to consider this to be a delta function of dopant that is introduced at the surface. It will then diffuse as a one-sided Gaussian (approximately) to the final junction depth.

If the final junction depth were smaller, we would be better using the formula which takes both the initial ΔR_p and the additional $\sqrt{2Dt}$ into consideration (Eqn.8.12). However, in this particular case, the profile would soon hit the surface of the wafer, complicating the final solution.

The diffusion of phosphorus at 900°C , ignoring concentration dependent diffusion is

$$D = 3.85 \exp\left(-\frac{3.66}{kT}\right) + 4.44 \exp\left(-\frac{4.0}{kT}\right) + 44.2 \exp\left(-\frac{4.37}{kT}\right) = 7.7 \times 10^{-16} \text{ cm}^2 \text{ sec}^{-1}$$

so the Dt at 900°C is

$$Dt = (7.7 \times 10^{-16})(10 \times 60) = 4.62 \times 10^{-13} \text{ cm}^2$$

or in terms of the diffusion distance

$$2\sqrt{Dt} = 2 \times 6.8 \times 10^{-7} \text{ cm} = 136 \text{ Angstroms}$$

i.e. it would normally diffuse very little at 900°C .

The measured junction depth is actually 0.25 microns, so that

$$C(x, t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) = 1 \times 10^{15} \text{ cm}^{-3}$$

giving

$$1 \times 10^{15} = \frac{1 \times 10^{14}}{\sqrt{\pi Dt}} \exp\left(-\frac{(0.25 \times 10^{-4})^2}{4Dt}\right)$$

The apparent Dt that the profile experiences is, by iteration

$$Dt = 1.64 \times 10^{-11} \text{ cm}^2$$

The ratio of the Dt's gives the enhancement

$$\Delta E = \frac{1.64 \times 10^{-11}}{4.62 \times 10^{-13}} = 35.5$$

This enhancement in the phosphorus diffusion is the average enhancement that the dopant experienced over the whole 10 minutes. The enhancement could have been much larger for a shorter time.

- 8.19. An engineer wants to form a shallow boron doped source/drain junction for an advanced technology. The manager wants to know whether the company should buy an inexpensive, batch furnace and achieve the required junction depth using a low thermal budget anneal (1 hour at 800°C) or an expensive, single wafer, RTA (rapid thermal annealer) using a high temperature anneal.**
- (a) Calculate the time required to achieve the target junction depth if the annealing temperature is 1050°C in the RTA.**
 - (b) Make a crude estimate of how far the dopants move during these anneals**
 - (c) Now, consider that the boron is introduced using an implant and that transient enhanced diffusion (TED) due to the implant will be important. Using the charts for the expected enhancement in diffusivity and the time TED lasts, calculate how far the dopants move at each temperature.**
 - (d) Which anneal would you recommend?**

Answer:

To achieve the same shallow junction depth for the low temperature furnace anneal and the high temperature RTA anneal, the Dt for each must be identical.

The Dt of boron for the furnace anneal is

$$Dt = D_B^{800} \times 3600 = 1.55 \times 10^{-13} \text{ cm}^2$$

The required time in the RTA annealer is

$$t = \frac{Dt}{D_B^{1050}} = \frac{1.55 \times 10^{-13}}{5 \times 10^{-14}} = 3.1 \text{ sec}$$

The approximate distance the profile moves is given by

$$2\sqrt{Dt} = 7.9 \text{ nm}$$

Now consider TED for the furnace anneal. From Figure 8-35, the TED causes an enhancement of 4000 in the dopant diffusivity at 800°C and this enhancement lasts for a time of 60 seconds (Figure 8-38). This will cause a diffusivity enhancement for this time after which the diffusivity returns to normal (Figure 8-42).

Thus, the actual Dt seen by the boron at 800°C when TED is considered is

$$Dt = D_B^{800} \times 4000 \times 60 + D_B^{800} \times 3540 = 1.05 \times 10^{-11} \text{ cm}^2$$

At the RTA temperature of 1050°C, the enhancement in diffusivity is 200 and the enhancement lasts for 0.05 seconds, so the Dt seen at the RTA temperature is

$$Dt = D_B^{1050} \times 200 \times 0.05 + D_B^{1050} \times 3.05 = 6.5 \times 10^{-13} \text{ cm}^2$$

The distance the profiles move is now given by

$$\text{Furnace : } 2\sqrt{Dt}|_{800} = 64.7 \text{ nm}$$

$$\text{RTA : } 2\sqrt{Dt}|_{1050} = 16.1 \text{ nm}$$

Thus, the TED has a much bigger effect during the lower temperature anneal even though the Dt 's were identical. This is because the implant damage has a higher supersaturation and stays around longer at lower temperatures. Thus, the RTA machine will enable the shallowest junction to be formed and may be a good investment.

8.20. An NMOS transistor is being built and an ion implantation is done after the gate oxide is grown and before the gate polysilicon deposition, in order to adjust the threshold voltage by +1volt.

(a) Which dopant should be used?

(b) Calculate the dose of the dopant if the oxide thickness is 10 nm.

Answer:

(a) An NMOS transistor has an n-type source and drain and hence needs an n-type channel to connect the source and drain when producing drive current. This requires the p-type substrate, which isolates the source and drain when the device is in the off state, to be inverted during the on state when the MOSFET provides drive current. Hence a positive gate voltage is applied under normal operation when switching on the device, producing an inversion layer. More positive charge in the channel will make the device harder to switch on and shift the threshold from V to $(V+1)$ Volt, so the channel charge should be boron or indium.

(b) If the gate oxide is 10 nm, the oxide capacitance per unit area is

$$C_{\text{ox}} = \frac{\epsilon\epsilon_0}{t_{\text{ox}}}$$

$$C_{\text{ox}} = \frac{3.9 \times 8.854 \times 10^{-14}}{100 \times 10^{-8}} = 345.3 \text{ nF}$$

and the V_t shift is given by

$$\Delta V_T = \frac{qQ}{C_{\text{ox}}}$$

$$1.0 = \frac{1.6 \times 10^{-19} Q}{345.3 \times 10^{-9}}$$

so that

$$Q = 2.16 \times 10^{12} \text{ cm}^{-2}$$

8.21. A deep source/drain implant in a 180 nm technology (see NTRS Table 7.1) using a 200 nm thick polysilicon gate layer requires a dose of $1 \times 10^{15} \text{ cm}^{-2}$ to obtain a low contact resistance. What is the maximum energy that can be used for an arsenic or phosphorus S/D implant if the implant cannot shift the threshold voltage by more than 0.01 volts, due to dose penetrating the gate region?

Answer:

From the NTRS roadmap, the oxide thickness will average 35 nm which gives us a way to calculate the oxide capacitance and hence the amount of dose that would shift the threshold voltage if it penetrates the polysilicon (+gate oxide)

$$Q_p = \frac{C_{\text{ox}}}{q} \Delta V_T$$

$$Q_p = \frac{\epsilon_r \epsilon_0}{t_{\text{ox}}} \frac{\Delta V_T}{q}$$

$$Q_p = \frac{3.9 \times 8.854 \times 10^{-14}}{35 \times 10^{-7}} \frac{0.01}{1.62 \times 10^{-19}} = 6.1 \times 10^{10} \text{ cm}^{-2}$$

This is the dose that penetrates the “mask” composed of the polysilicon and the (negligible) gate oxide thickness,

$$Q_p = \frac{Q}{2} \operatorname{erfc}\left(\frac{d - R_p^*}{\sqrt{2\Delta R_p^*}}\right)$$

$$\frac{2Q_p}{Q} = \operatorname{erfc}\left(\frac{d - R_p^*}{\sqrt{2\Delta R_p^*}}\right)$$

$$\frac{2 \times 6.1 \times 10^{10}}{10^{15}} = \operatorname{erfc}\left(\frac{0.2 \times 10^{-4} - R_p^*}{\sqrt{2\Delta R_p^*}}\right)$$

$$\operatorname{erf}\left(\frac{0.2 \times 10^{-4} - R_p^*}{\sqrt{2\Delta R_p^*}}\right) = 0.9987$$

$$\left(\frac{0.2 \times 10^{-4} - R_p^*}{\sqrt{2\Delta R_p^*}}\right) = 2.29$$

$$R_p^* + 3.2\Delta R_p \leq 0.2\mu\text{m}$$

For arsenic, iterating through several guesses from the range tables

$$R_p \approx 0.09\mu\text{m} \quad \Delta R_p \approx 0.035\mu\text{m} \quad E \approx 135\text{keV}$$

Similarly, if phosphorus was used for the S/D implant

$$E \leq 60\text{keV}$$

8.22. Calculate the change in junction depth for a 40 keV boron threshold adjust implant of $5 \times 10^{13} / \text{cm}^2$ annealed at 750°C in a furnace or at 1000°C in an RTA for a time just long enough to remove all the damage that causes TED. Assume a uniform well doping of $5 \times 10^{16} / \text{cm}^3$.

Answer:

The primary formula we will use gives the enhancement in diffusivity while TED lasts

$$Dt_{\text{eff}} = D^* \frac{I}{I^*} \tau_{\text{enh}}$$

For 40 keV boron,

$$R_p \approx 0.14 \mu\text{m} \quad \Delta R_p \approx 0.055 \mu\text{m}$$

The initial junction is given by

$$C_B = \frac{Q_{\text{Si}}}{\sqrt{2\pi\Delta R_p^2}} \exp\left(-\frac{(x - R_p)^2}{2\Delta R_p^2}\right)$$

$$x = 0.3 \mu\text{m}$$

$$D_B = 0.76 \exp\left(-\frac{3.46}{kT}\right)$$

At 750°C, the equilibrium boron diffusivity is

$$D_B = 6.9 \times 10^{-18} \text{ cm}^2 \text{ sec}^{-1}$$

and it is enhanced by 7500 (Fig. 8-35) for a time of 400 seconds (Fig. 8-38) giving

$$Dt_{\text{eff}}^{750^\circ\text{C}} = 2.07 \times 10^{-11} \text{ cm}^2$$

At 1000°C, the equilibrium boron diffusivity is

$$D_B = 1.5 \times 10^{-14} \text{ cm}^2 \text{ sec}^{-1}$$

and it is enhanced by 400 for 0.2 seconds

$$Dt_{\text{eff}}^{1000^\circ\text{C}} = 1.2 \times 10^{-12} \text{ cm}^2$$

Clearly, the effective Dt is higher at 750°C than at 1000°C, and we can calculate the new junction depths from

$$x = R_p + \sqrt{-2(\Delta R_p^2 + 2Dt_{\text{eff}}) \ln\left[\frac{C_B}{Q} \sqrt{2\pi(\Delta R_p^2 + 2Dt_{\text{eff}})}\right]}$$

$$x = 0.14 + 0.23 = 0.37 \mu\text{m} \text{ at } 750^\circ\text{C}$$

$$x = 0.14 + 0.17 = 0.31 \mu\text{m} \text{ at } 1000^\circ\text{C}$$

The 1000°C junction is deeper by

$$\Delta x_j = 0.06\mu\text{m}$$

Chapter 9 Problems

- 9.1. What would you expect would happen to the threshold voltage of a MOS transistor if the gate oxide was deposited using $\text{SiH}_4 + 2\text{O}_2 \rightarrow \text{SiO}_2 + 2\text{H}_2\text{O}$ rather than thermally grown?**

Answer:

The threshold voltage is likely to change because of fixed charge at the interface for a deposited interface compared to a thermally grown interface. The change in threshold voltage is given by

$$\Delta V_T = \frac{qQ_f}{C_{\text{OX}}}$$

and shifts by an amount determined by the quantity of fixed charge at the interface.

- 9.2. What are the two commonly observed rate limiting steps in silicon epitaxial growth? Under what conditions do they normally dominate the overall deposition rate?**

Answer:

The two commonly observed rate limiting steps in silicon epitaxial growth are mass transfer (diffusion) through the stagnant boundary layer and surface reaction(s). The former normally dominates at higher temperatures and higher total pressures, and the latter at lower temperatures and lower pressures.

- 9.3. In an epitaxial deposition, under mass transfer limited conditions, is it more important to control the reactor temperature or the source gas composition in the gas stream to obtain reproducible results? Why?**

Answer:

Under mass transfer limited conditions it is more important to control the source gas composition in the gas stream to obtain reproducible results. The mass transfer process is rather temperature insensitive (with a low activation energy), while the source gas composition in the gas stream strongly affects the deposition rate (Eqn. 9-14).

- 9.4. In a reactor used for epitaxial growth, the wafers are normally placed flat on the susceptor, and epi grows on the top side only. If the same reactor were used to oxidize wafers, by introducing O_2 rather than SiH_4 (or another Si gas source), SiO_2 would grow on both sides of the wafer. Explain why SiO_2 grows on both sides and epi grows only on the top side.**

Answer:

The difference in the two processes is a result of what rate limits each process. In the case of epi growth, gas phase transport through the boundary layer is an important process and so no growth will occur on the wafer backside (diffusion distance is too long for the species to get to the backside). In the case of oxidation, gas phase transport is not rate limiting. Diffusion through the SiO₂ and or the Si/SiO₂ interface reaction limit the growth rate. Thus the oxidant species should be able to diffuse under the wafer and oxidize the backside (there is always a small gap between the wafer and the susceptor since neither is perfectly flat).

9.5. For CVD deposition of a film, it is found that the mass transfer coefficient $h_G = 10.0 \text{ cm sec}^{-1}$ and the surface reaction rate coefficient $k_s = 1 \times 10^7 \exp(-1.9 \text{ eV}/kT) \text{ cm sec}^{-1}$. For a deposition at 900°C, which CVD system would you recommend using: (a) a cold-walled, graphite susceptor type: or (b) a hot-walled, stacked wafer type? Explain your answer.

Answer:

At 900°C,

$$k_s = 1 \times 10^7 \exp\left(\frac{-1.9 \text{ eV}}{kT}\right) \text{ cm} \cdot \text{sec}^{-1} = 1 \times 10^7 \exp\left(\frac{-1.9 \text{ eV}}{8.62 \times 10^{-5} \frac{\text{eV}}{\text{K}} 1173 \text{ K}}\right) \text{ cm} \cdot \text{sec}^{-1}$$

$$= 0.069 \text{ cm sec}^{-1}. \quad \text{This is much less than } h_G (10.0 \text{ cm sec}^{-1})$$

Therefore $k_s \ll h_G$. This means that the system will be in the surface reaction controlled regime. The transfer of species through the boundary layer in the gas phase is fast and not important relative to the surface reaction. Therefore you can stack the wafers without causing uniformity problems due to gas transport variations, and achieve maximum throughput. In this regime, the deposition rate is sensitive to temperature, but in a hot-walled system, good temperature control can be achieved. Thus you should recommend using (b) a hot-walled, stacked wafer type system.

Note: We are assuming here that the mass transfer is still very fast relative to the reaction rate when the wafers are in a stacked configuration. However, with the stacked configuration, the reactants must travel from the gas stream to each position on the wafer through a stagnant layer between the wafers, and the value of h_G would decrease the further they must travel. For larger size wafers, h_G thus becomes smaller and the mass transfer of reactants becomes more difficult, even at lower pressure, than for smaller wafers. In that case the spacing between the wafers may need to be increased, or the pressure reduced, to ensure reaction limited conditions and uniform deposition.

- 9.6. Calculate the deposition rate for an LPCVD system with the same parameter values given in the example following Eq. 9.10, but at a reduced total pressure so that h_G is increased by 100 times. Assume that the partial pressure of the incorporating species, P_G , remains the same, and C_T decreases by the same factor as the total pressure.**

Answer:

At this lower pressure:

$$h_G = 100 \text{ cm sec}^{-1}$$

$$k_S = 10 \text{ cm sec}^{-1}$$

By Eq. 9.10:

$$v = \frac{k_S h_G}{k_S + h_G} \frac{C_T}{N} Y = \frac{1}{\left(\frac{1}{h_G} + \frac{1}{k_S} \right)} \frac{C_T}{N} Y$$

Since P_G stays constant and C_T decreases by the same factor as P_T , then $\frac{C_T}{N} Y = \frac{C_T}{N} \frac{P_G}{P_T}$ stays the same as in the previous example.

Therefore:

$$\begin{aligned} v &= \frac{1}{\left(\frac{1}{100 \text{ cm sec}^{-1}} + \frac{1}{10 \text{ cm sec}^{-1}} \right)} \frac{1 \times 10^{19} \text{ cm}^{-3}}{5 \times 10^{22} \text{ cm}^{-3}} \frac{1}{760} \\ &= 2.4 \times 10^{-6} \text{ cm sec}^{-1} \\ &= 4.8 \times 10^{-6} \frac{\text{cm}}{\text{sec}} * \frac{\left(\frac{10^4 \mu\text{m}}{\text{cm}} \right)}{\left(\frac{\text{min}}{60 \text{ sec}} \right)} = 1.4 \mu\text{m min}^{-1} \end{aligned}$$

Since $k_S \ll h_G$, the system is operating in the surface reaction controlled regime.

- 9.7. (a). Plot the deposition rate (on a log scale) versus $1/T$ (Kelvin), for 600-1200°C, for a CVD system with the following parameter values:**

$$h_G = 0.5 \text{ cm sec}^{-1}$$

$$k_S = 4 \times 10^6 \exp(-1.45 \text{ eV}/kT) \text{ cm sec}^{-1}$$

Partial pressure of incorporating species = 1 torr

Total pressure = 1 atm

$$C_T/n = 1/10,000$$

Identify the reaction and mass transfer limited regimes.

(b). Redo the problem when the total pressure is decreased to 1 torr, so that h_G increases by 100 times. Assume that the partial pressure of the incorporating species remains the same, and C_T decreases by the same factor as the total pressure. (Since both the total pressure and the partial pressure of the incorporating species equal 1 torr, this means that the gas is made up of only the incorporating species in this case.)

Answer:

$$\text{By Equation 9.10: } v = \frac{k_S h_G}{k_S + h_G} \frac{C_T}{N} Y = \frac{1}{\left(\frac{1}{h_G} + \frac{1}{k_S} \right)} \frac{C_T}{N} Y,$$

$$h_G = 0.5 \text{ cm sec}^{-1}$$

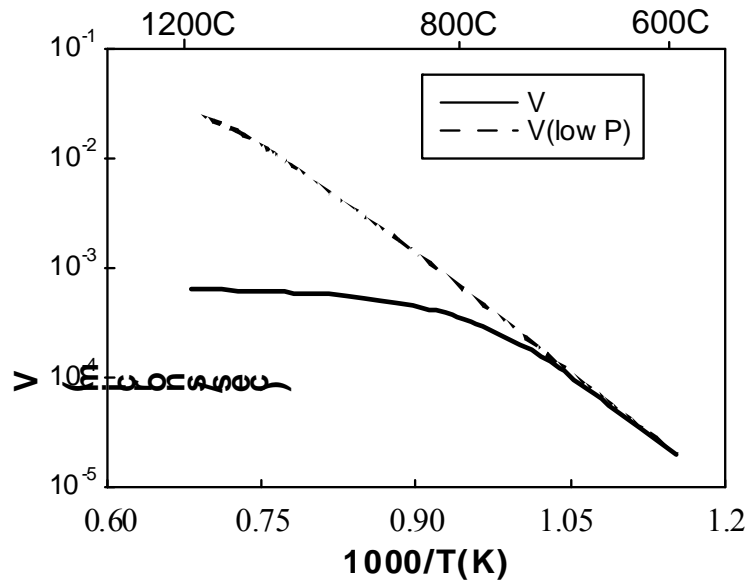
$$k_S = 4 \times 10^6 \exp(-1.45 \text{ eV}/kT) \text{ cm sec}^{-1} \text{ where } T = T^\circ\text{C} + 273$$

$$C_T/N = 1/10,000$$

$$Y = P_g/P_{\text{total}} = 1 \text{ torr} / 760 \text{ torr}$$

For low pressure case, only h_g changes (increased by 100 times). C_T/N times P_g/P_{total} remains constant since n and P_g each remains constant, and the C_T/P_{total} ratio stays constant as stated above.

The plot of v for each case versus $1/T$ below. For the 1 atm total pressure case, curve v , the transition between surface reaction controlled and diffusion controlled is at about 800°C , while for the low pressure case ("low P"), the transition occurs at somewhere above 1200°C .

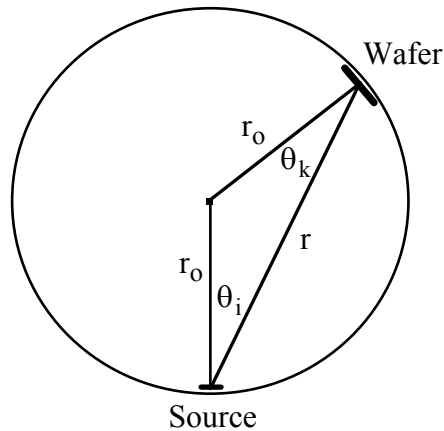


9.8. Show that placing the source in an evaporation system at the inside surface of a sphere facing the center, with the wafers also at the inside surface of the sphere facing the center, leads to uniform deposition wafer-to-wafer assuming a small planar surface source.

Answer:

From Equation 9.23, $v = \frac{R_{\text{evap}}}{\pi N r^2} \cos \theta_i \cdot \cos \theta_k$ for a small area planar source, where

θ_i is the angle between the surface normal of small area i (the source) and the line between areas i and k , while θ_k is the angle between the surface normal of small area k (the wafer) and the line between areas i and k . Therefore, by the figure below and these definitions: for a wafer at any location around the sphere: $\theta_i = \theta_k$ (isosceles triangle) and $\cos \theta_i = \cos \theta_k = r/2r_o$.



Therefore $v = \frac{R_{\text{evap}}}{\pi N r^2} \cos \theta_i \cdot \cos \theta_k = \frac{R_{\text{evap}}}{\pi N r^2} \frac{r}{2r_o} \cdot \frac{r}{2r_o} = \frac{R_{\text{evap}}}{\pi N 4r_o^2}$. Thus the

deposition rate will depend only on the evaporation rate, R_{evap} , assumed constant, and on the radius of the sphere, another constant, giving constant deposition on each wafer around the sphere. (Or at least constant deposition on the center of each wafer.)

- 9.9. Calculate the deposition rate for a small planar surface evaporation source in which $\theta_i = 30^\circ$, $\theta_k = 45^\circ$, the evaporation rate is $1 \times 10^{-3} \text{ gm sec}^{-1}$, the distance from the source to the wafer is 5 cm, and the density of the material being deposited equal 5 gm cm^{-3} .

Answer:

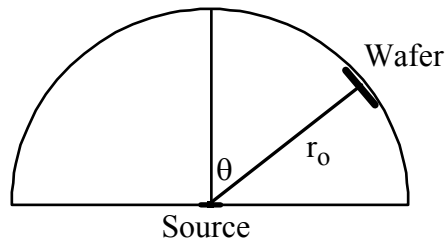
By Eqn. 9.23, the deposition rate, v , is:

$$v = \frac{R_{\text{evap}}}{\pi N r^2} \cos \theta_i \cdot \cos \theta_k$$

Plugging in the numbers into this equation gives:

$$\begin{aligned} v &= \frac{1 \times 10^{-3} \text{ gm sec}^{-1}}{\pi (5 \text{ cm})^2 \cdot 5 \text{ gm cm}^{-3}} \cos 30^\circ \cos 45^\circ \\ &= 2.6 \times 10^{-6} \text{ cm sec}^{-1} * \frac{\sqrt{3}}{2} \frac{1}{\sqrt{2}} \\ &= 1.6 \times 10^{-6} \text{ cm sec}^{-1} = 0.94 \text{ } \mu\text{m min}^{-1} \end{aligned}$$

- 9.10. The new deposition engineer installed the company's new evaporation system. Hoping to get uniform depositions on all the wafers mounted on the inside of the spherical wafer holder, he installed the evaporation source crucible at the center of the sphere. If the evaporation source behaves like an ideal small area planar source, what will be the deposition rate as a function of θ , as defined below? (let the evaporation rate equal $3 \times 10^{-3} \text{ gm sec}^{-1}$, radius r_0 equal 10 cm, and the density of the material being deposited equal 10 gm cm^{-3}). Sketch a plot of d versus θ for θ from -90° to $+90^\circ$, and specify the deposition rate at a point directly facing the planar source (at $\theta = 0^\circ$) and at $\theta = 90^\circ$. Will uniform depositions result?



Answer:

From Eqn. 9.20, but assuming a cosine emission behavior from the small area source, the flux emitted from a small area planar source to an area facing the source is:

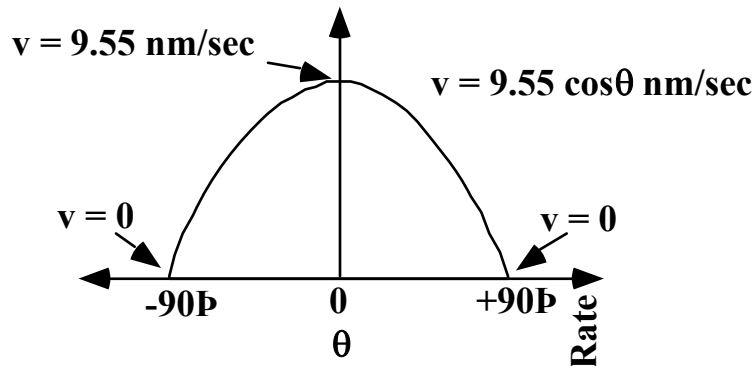
$$F_{ik}^p = \frac{R_{\text{evap}}}{\pi r^2} \cos \theta_i.$$

The deposition rate, v , is just that flux divided by the film density, so that

$$\begin{aligned} v &= \frac{R_{\text{evap}}}{\pi r^2 N} \cos \theta = \frac{3 \times 10^{-3} \text{ gm sec}^{-1}}{\pi (10 \text{ cm})^2 10 \text{ gm cm}^{-3}} \cos \theta \\ &= 9.55 \times 10^{-7} \cos \theta \text{ cm/sec} = 9.55 \cos \theta \text{ nm/sec} \end{aligned}$$

(Or one could just use Eqn. 9.23 with $\cos \theta_k = 1$ to get the same result.)

A plot of v versus θ will be a cosine function, with a value of v at $\theta = 0^\circ$ equal to 9.55 nm/sec, and a value of 0 at $\theta = \text{plus or minus } 90^\circ$. No, uniform deposition will not occur.



- 9.11. Calculate the mean free path of a particle in the gas phase of a deposition system and estimate the number of collisions it experiences in traveling from the source to the substrate in each of the cases below. Assume that in each case the molecular collisional diameter is 0.4 nm, the source-to-substrate distance is 5 cm, and that the number of collisions is approximately equal to the source-to-substrate distance divided by the mean free path.

- An evaporation system in which the pressure is 10^{-5} torr and the temperature is 25°C .
- A sputter deposition system in which the pressure is 3 mtorr and the temperature is 25°C .
- An LPCVD system in which the pressure is 1 torr and the temperature is 600°C .
- An APCVD system in which the pressure is 1 atm and the temperature is 600°C .

Answer:

The mean free path of a gas particle is (Eqn. 9.26) $\lambda = \frac{kT}{\sqrt{2}\pi d^2 P}$ where $k = 1.36 \times 10^{-22} \text{ cm}^3 \text{ atm K}^{-1}$, T is the temperature in K, d is the collision diameter of the molecule in cm (approximately $4 \times 10^{-8} \text{ cm}$ for most molecules of interest), and P is the pressure in atm. The # collisions is approximately equal to the source-to-substrate distance divided by the mean free path in each case. Plugging in the numbers gives:

$$\lambda(\text{in cm}) = \frac{kT}{\sqrt{2}d^2 P} = \frac{1.36 \times 10^{-22} \text{ cm}^3 \cdot \text{atm} \cdot \text{K}^{-1} \cdot T(\text{K})}{\sqrt{2}\pi (4 \times 10^{-8} \text{ cm})^2 \frac{P(\text{torr})}{760 \text{ torr/atm}}} = 1.45 \times 10^{-5} \frac{T(\text{K})}{P(\text{torr})}$$

- a. 433 cm, 1.2×10^{-2} collisions;
- b. 1.44 cm, 3.5 collisions;
- c. 0.013 cm, 392 collisions;
- d. $1.7 \times 10^{-5} \text{ cm}$, 3.0×10^5 collisions

One can see that the relatively high pressures in CVD (both APCVD and LPCVD) systems result in a much shorter mean free path in the gas phase and much larger number of collisions as the reactants travel to the wafer surface. This leads to a more isotropic arrival angle distribution. (Also note that only in the APCVD case is the mean free path less than, or on the order of, typical feature dimensions, and thus only in APCVD systems do significant collisions occur within surface features. For sputtering, evaporation, and LPCVD systems, virtually no gas phase collisions occur within surface features. Re-emission processes or surface diffusion are thus usually required for good coverage in these systems.)

- 9.12.a). Using schematic diagrams, show qualitatively how changing the gas pressure in a standard PVD system (using an appropriate S_c value for this type of system) can affect the deposition profile when depositing a material on a narrow trench structure.**
- b). In a similar fashion, show how changing the sticking coefficient affects the deposition profile when depositing a material on topography with a step.**

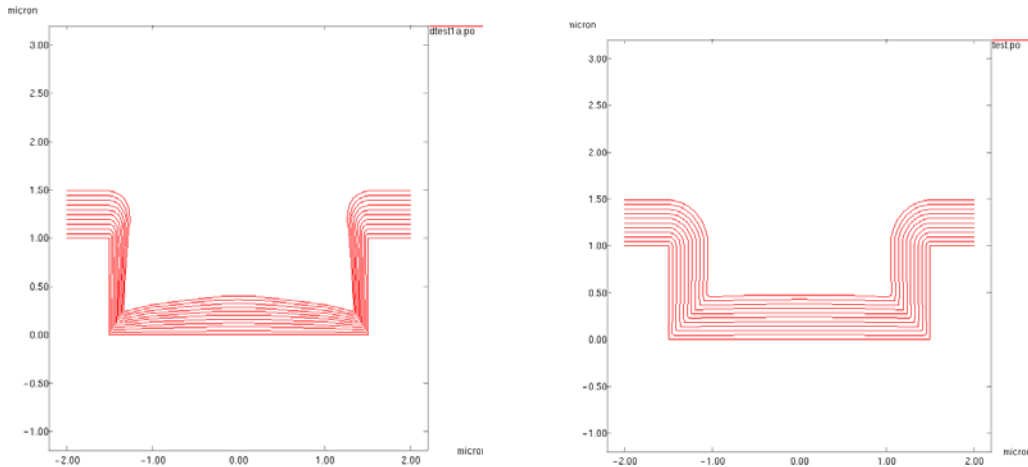
Answer:

a). $S_c=1$ for PVD processes. Increasing pressure means more collisions and a more isotropic arrival angle distribution. Thus n is decreased. Figure 9-47 shows what changing n does to profile when $S_c=1$, with $n=1$ and $n=10$. ($n=1$ corresponds to higher pressure). With higher n (lower pressure), less deposition on sidewalls and less overhang occurs, and more deposition on bottom of trench.

b). Lowering S_c means more re-emission and deposition on the side of the step, giving better step coverage. Shown below are simulations when S_c is decreased from 1 to 0.1, for an n value of 1. Similar trends would occur for other values of n .

$n=1, S_c=1$:

$n=1, S_c=0.1$:



9.13. How does the ability to fill the bottom of a narrow trench using sputter deposition change as the target is moved further away from the wafer? Neglect any gas phase collision effects.

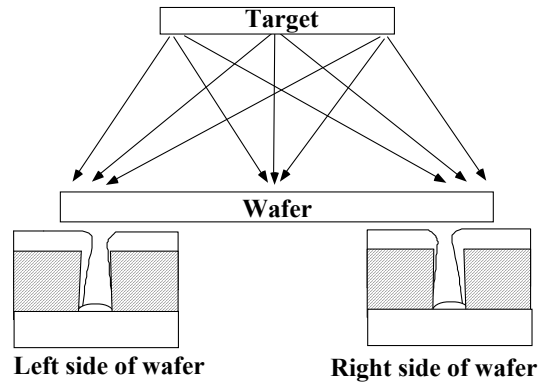
Answer:

The further away the target, the narrower the arrival angle distribution, similar to making the target smaller. So n is greater and better filling of the bottom of a narrow trench is achieved.

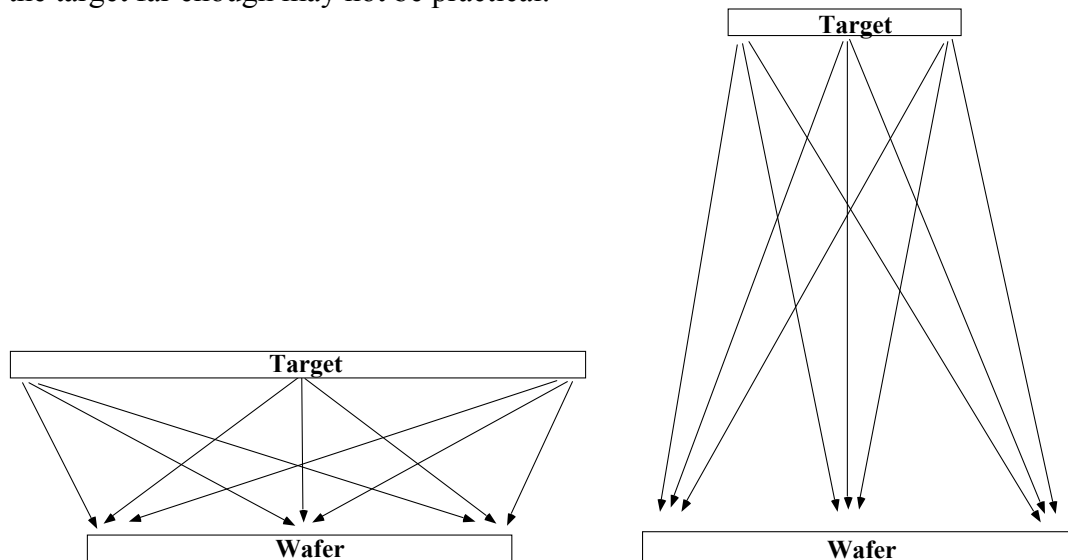
9.14. Explain how asymmetric depositions can occur on a wafer in a sputter deposition system. Asymmetric deposition means that thicker deposition occurs on one side of a feature (a step, for example) than the other. Also suggest a way to decrease the asymmetry. Use schematic diagrams in your explanations.

Answer:

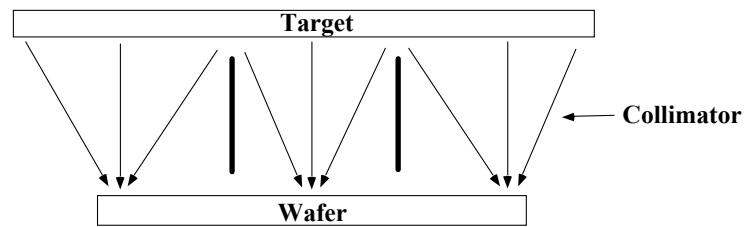
Asymmetry can occur as a result of the limited size of the target, the finite distance between target and wafer, near line-of-sight transport from the target to the wafer, and little or no re-emission from the wafer surface. Assuming a wide range of emitted angles from each point of the target (diffuse emission, for example), the flux striking the wafer can arrive at different angles at each point on the wafer, as shown below for a relatively small target. (Few or no collisions are assumed to occur between the target and wafer here.) In this case, more flux will be coming from the right on the left side of the wafer, and more flux will be coming from the left on the right side of the wafer, resulting in asymmetric depositions.



A variety of things can be done which can decrease the asymmetry. The target size can be widened, as shown below on the left. As one can see, the arrival angle distribution and deposition are less asymmetric across the wafer. For an infinitely wide target, perfectly symmetric deposition would occur. Moving a small target further from the wafer, as shown below on the right, may also help. The arrival distribution becomes more perpendicular all across the wafer as the target is moved away, leading to more uniform and symmetric deposition, although moving the target far enough may not be practical.



Moving the wafers relative to the target during the deposition, which is similar to widening the target size, would also help. Another way to reduce any asymmetry is to increase the pressure and increase the collisions – ultimately isotropic arrival at all locations will occur and no asymmetry will result. However, increasing the pressure in PVD systems can result in worse film properties. Using collimated and ionized sputter deposition systems also result in more symmetric depositions, while narrowing the arrival angle distributions. Placing collimator cells between the target and wafer, as shown below, filters out the species arriving at wide angles and decreases the asymmetry of the deposition at the edges of the wafer. In an ionized sputtering system, the ionized species arrive in a near-perpendicular direction all across the wafer.



- 9.15. What value of n in the arrival angle distribution is desired for good step coverage over a step in topography? For good bottom filling of a via? Explain with schematic diagrams. What value of S_c , the sticking coefficient, do you want in each case? Explain with diagrams.**

Answer:

For better step coverage, usually want n low ($=1$). See Figure 9-46.
For good bottom filling, want n high. See Figure 9-47.

For both good step coverage and good bottom filling, want S_c low. See Figure 9-48.

- 9.16. We have seen how conformal coverage in low pressure systems results from a low sticking coefficient, which causes multiple reemission and redeposition processes within the feature topography. Describe how in atmospheric CVD systems, gas phase collisions by the reactants within the topographical features can lead to good coverage and filling.**

Answer:

Here is the short answer:

Good coverage occurs in low pressure systems when atoms are re-emitted or “bounce off” the surfaces that they strike (unshadowed regions) and then strike other surfaces, including shadowed regions. In atmospheric systems, good coverage can occur when reactant atoms or molecules collide with and “bounce off” other species in the gas phase (i.e. diffusional flow) within the topographical features (as well as bounce off or be re-emitted from surfaces), and then strike the shadowed regions.

Here is the more complete answer:

And the collision or bouncing rate (i.e, the gas phase diffusion rate) should be high compared to the surface reaction rate to ensure that all surfaces get equal deposition, resulting in conformal coverage. Otherwise, some surfaces within the topography could get more deposition and could even lead to pinching off of holes and trenches.

And now all of this in more detail:

In the APCVD systems, the pressure is high enough, and the mean free path is short enough, that many gas phase collisions occur near the surface and between the features. The dominant gas phase transport mechanism near the surface and within the topographical features of surface (such as steps, trenches, and vias) is thus gas-phase diffusion instead of the free-molecular flow that occurs in the low pressure systems (including LPCVD, PECVD, HDPCVD, PVD and evaporation). A significant number of collisions now occur in the space within the surface features,

which randomizes the direction of the species within the features and can result in deposition on surfaces that would normally be shadowed in low pressure systems where line-of-sight deposition occurs. To achieve very conformal deposition (the same deposition rate at every position over or within the feature), the deposition must be limited by the surface reaction, and not the gas phase diffusion.

The transport of reactant species to each surface position occurs by isotropic diffusion through the gas phase, down the concentration gradient to the surface. If the diffusion process is the rate limiting one, then the surface concentration at each position can vary depending on its location, since it will take different times for the species to reach the different positions depending on the topography. For example it will take longer for a molecule to reach the bottom of the trench than the top, resulting in uneven deposition in the trench. And as the deposition on top corners builds up (with the top corners having a view angle of 270 degrees versus 180 in low pressure systems), this can pinch off the top of the trench and hinder the species from reaching the bottom.

But by increasing the gas phase diffusion relative to the surface reaction, the transport of species to each surface location is faster. When the gas diffusion is much greater than the surface reaction, then the surface reaction controls the deposition. Since the surface reactions are slow compared to the arrival of new species at the surface, there is usually an abundant supply of species at each point along the surface waiting to react, regardless of its position within the topography. Therefore all surfaces get uniform deposition, and hence conformal deposition.

This is the same idea as we discussed for obtaining uniform deposition across a wafer and from wafer-to-wafer on a more macroscopic scale. If you operate in the reaction limited regime, the surface concentration becomes constant everywhere, giving constant deposition. For APCVD systems, this extends to the feature scale where diffusional flow still dominates the gas transport mechanism.

One can also increase the gas diffusion relative to the surface reaction by decreasing the temperature, which decreases the reaction rate much faster than decreasing the gas diffusion. But for most systems, decreasing the reaction rate decreases the deposition rate too much or degrades the film properties.

One can choose a chemistry (such as TEOS/O₃ for oxide deposition) with a very low surface reaction rate relative to the gas diffusion rate, which does have adequate deposition rate and film quality and operates in the reaction limited regime. (But even if one is then operating in the reaction limited regime, and obtaining conformal coverage, one may still not want to stack the wafers as in LPCVD systems. As discussed in the answer to problem 9.5, when stacking the wafers the distance the gas must travel can increase and vary a lot depending on position, resulting in uneven deposition. Only in low pressure systems are wafers usually stacked.) Chemistries that have a low sticking coefficient in low pressure systems would generally have a low surface reaction rate, and vice versa. This is

because having a relatively low reaction rate allows for more re-emission to occur. Likewise, having a low sticking coefficient means that the net reaction rate is slow. (Remember that the sticking coefficient is defined as the “reacting” flux divided by the incident flux.)

Alternately, one can increase the gas phase diffusivity by decreasing the pressure and making the process more reaction limited. But if the pressure is decreased too much so that the transport is more line-of-sight molecular flow, rather than by diffusional transport, the LPCVD models apply instead. In that case reemission and line-of-sight redeposition (with a low sticking coefficient) are necessary to obtain conformal coverage – there are no gas phase molecules within the features to bounce off of.

Chapter 10 Problems

10.1. Why might pure chemical etching, such as in wet etching, be adequate for patterning the silicon nitride layer used to define the field oxidation area in the CMOS process flow given in Chapter 2?

Answer:

For a purely chemical process, the etching is usually assumed to be isotropic with the degree of anisotropy A_f equal to 0. By Eqn 10.7,

$$A_f = 1 - \frac{b}{x_f}$$

this would give an etch bias equal to the film thickness (overetching may add 5-10% to this.) But the nitride layer for the field oxidation process is only about 80 nm, or 0.08 μm . Thus the undercutting is only about 0.08 μm . And the field oxidation delineation is not that critical. So a 0.08 μm undercutting may be adequate for this step.

10.2. If the etch anisotropy is 0, as in wet etching, what is the undercut or etch bias when etching a 0.5 μm thick film? What is the undercut when the anisotropy is 0.75? Assume no overetch in each case.

Answer:

The anisotropy in this case is given by Eqn. 10.7. Rearranging gives:

$$b = x_f(1 - A_f)$$

For A_f equal to 0, the undercutting or etch bias, b , equals the thickness of the film being etched, or 0.5 μm .

For A_f equal to 0.75:

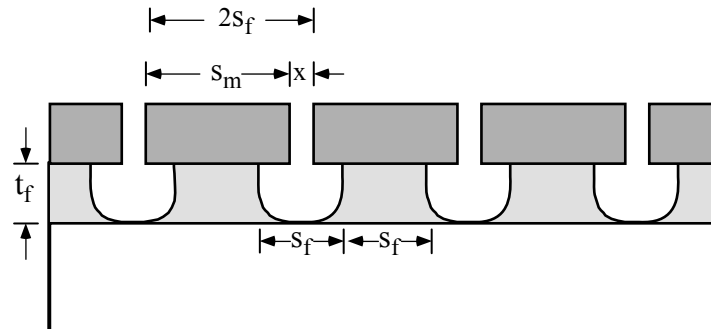
$$b = 0.5(1 - 0.75) = 0.125 \mu\text{m}$$

10.3. In a certain process, it is desired that the pitch of metal lines be equal to or less than 1.0 μm (the pitch equals one metal line width plus one spacing between metal lines, measured at top of features). Assume that the metal line width and spacing are equal (that is, 0.5 μm each). The height of such structures is also 0.5 μm , and the minimum lithographic dimension is 0.25 μm .

a. What minimum degree of anisotropy is needed in an etch process in order to produce such a structure?

b. What minimum pitch could be obtained for such a structure with *wet* etching? (Again with minimum lithograph dimension of $0.25\ \mu\text{m}$, thickness of $0.5\ \mu\text{m}$, and equal metal width and spacing.)

Answer:



a. Metal pitch = $S_f + S_f = 1\ \mu\text{m}$, or $S_f = 0.5\ \mu\text{m}$.

$x = 0.25\ \mu\text{m}$

$t_f = 0.5\ \mu\text{m}$

From Eqn. 10.8, $S_f = x + 2t_f(1 - A_f)$,

$$0.5 = 0.25 + 2(0.5)(1 - A_f) \quad \text{- all in } \mu\text{m}$$

which leads to A_f , the degree of anisotropy, equal to 0.75.

(this could not be achieved by wet etching.)

b. If wet etching is used, we assume the degree of anisotropy is equal to 0.

$$\text{Again, } S_f = x + 2t_f(1 - A_f)$$

$$\begin{aligned} \text{so that } S_f &= 0.25 + 2(0.5)(1 - 0) \\ &= 1.25\ \mu\text{m} \end{aligned} \quad \text{- all in } \mu\text{m}$$

The pitch equals $2 * S_f = 2.5\ \mu\text{m}$

10.4. What are the advantages and disadvantages of reactive ion etching versus sputter etching? Cite a hypothetical example of when you might want to use sputter etching rather than RIE?

Answer:

RIE gives more selectivity; sputtering etches almost everything, at about the same rate. You might want to use sputter etching when etching multilayer stack of different materials on the substrate, when etching substrate a little doesn't matter. (You might also use sputter etching in a dep/etch/dep process to remove overhang, exploiting angle-dependent sputtering behavior. Selectivity is not important here so there is no need to develop an RIE process for this.)

10.5. Explain how loading effects can affect endpoint detection.

Answer:

The signal to be detected from either the reactants or products will depend on how much material is being etched. Therefore the critical point, or threshold, for the endpoint could depend on the etching area, or circuit layout. Also, as the etching nears completion, the etch rate can increase dramatically.

10.6. It is found that a certain plasma etch chemistry in a certain RIE etch system produces vertical sidewalls with zero etch bias when etching a particular film. Adding chemical A to the etch chemistry results in non-vertical sidewalls, and an etch bias. Adding chemical B to the original etch chemistry results in non-vertical sidewalls, but with zero etch bias. Explain what may be going on.

Answer:

Adding A results in a higher active species to inhibitor ratio (either by increasing free radical production or decreasing inhibitor/polymer formation), reducing the sidewall inhibitor formation relative to the etch rate, and leading to more isotropic etching. Adding B results in a lower active species to inhibitor ratio, increasing the the sidewall inhibitor formation relative to the etch rate. This increases the inhibitor deposition to etch ratio, enough so that excess inhibitor forms. This leads to sloped sidewalls, as illustrated in Figure 10-14 or 10-24c, with no undercutting, or etch bias.

- 10.7. a. In a particular etch process, if selectivity is the biggest concern, which type(s) of etch equipment should be used?**
b. If the biggest concern is ion bombardment damage, which type(s) of etch equipment should be used?
c. If the biggest concern is obtaining vertical sidewalls, which type(s) of etch equipment should be used?
d. If the biggest concerns are selectivity *and* vertical sidewalls, which type(s) of etch equipment should be used?
e. What about selectivity *and* vertical sidewalls *and* damage, while maintaining reasonable etch rate?

Answer:

- a. If biggest concern is selectivity? barrel or remote plasma or wet etching
- b. If biggest concern is damage? barrel or remote or wet
- c. If biggest concern is vertical sidewalls? sputtering or RIE/HDP
- d. What about selectivity AND vertical sidewalls? RIE/HDP
- e. What about selectivity AND vertical sidewalls AND damage, while maintaining reasonable etch rate? HDP

10.8. It is observed that the sidewall slope in an etch process becomes more sloped as the temperature is reduced. Why?

Answer:

Plasma etch processes are rather independent of temperature, but inhibitor deposition usually goes up when the temperature decreases (less desorption as temperature decreases). The etch/deposition ratio thus decreases, and the slope is increased.

10.9. If the anisotropy of an etch process is 0.45, sketch the etch profile. What percentage of the etch rate in the vertical direction is due to the chemical component and what percentage is ionic/physical, assuming a linear etch mechanism? State all assumptions.

Answer:

By Eq. 10.5, $A = 1 - r_{\text{lat}}/r_{\text{vert}}$. $0.45 = 1 - r_{\text{lat}}/r_{\text{vert}}$. $r_{\text{lat}}/r_{\text{vert}} = 0.55$ (= amount of bias for each μm deep). The profile would look like Figure 10-3b.

For a linear etch mechanism we assume that the etch rate just equals the sum of the physical and chemical etching at each point following Eqn. 10.15, with no coupling between the physical and chemical components. The physical etching occurs only in the vertical direction and the (spontaneous) chemical etching occurs equally in both lateral and vertical directions (assuming that the physical etching is completely anisotropic, the chemical etching is perfectly isotropic, and the etch rate of each component is constant with time).

$$r_{\text{lat}}/r_{\text{vert}} = 0.55 = \text{chem}(\text{lat}) / [\text{chem}(\text{vert}) + \text{phys}(\text{vert})].$$

But $\text{chem}(\text{lat}) = \text{chem}(\text{vert})$, so

$$0.55 = \text{chem (vert)} / (\text{chem(vert)} + \text{phys(vert)})$$

$$0.55 = \text{chem (vert)} / \text{total(vert)}$$

So the chemical etch component in the vertical direction = 0.55
and the physical etch component in the vertical direction = 0.45

10.10. For a particular plasma etch process in which the linear etch model is applicable, a degree of anisotropy of 0.8 or better is desired. If the unobstructed ionic flux on a flat surface is 3×10^{16} atoms $\text{cm}^{-2} \text{sec}^{-1}$ (with K_i equal to 1), what unobstructed chemical flux would result in an anisotropy of 0.8. For this process S_c is 0.01 and K_f is 0.1.

Answer:

The degree of anisotropy is given as: $A_f = 1 - \frac{r_{\text{lat}}}{r_{\text{ver}}}$. For an A_f value of 0.8,

$\frac{r_{\text{lat}}}{r_{\text{ver}}} = 0.2$. Following the example given in the text, for the linear etch model the

etch rate in the lateral direction is given by Eqn 10.15 with only the chemical flux term. The etch rate in the vertical direction is given by Eqn 10.15 with both the chemical flux and ionic flux terms. Assuming no shadowing by the mask of the ion species in the center of the etch window (usually a reasonable assumption for the very directed ion species), we can let F_i equal the unobstructed flux of the ion species for the vertical etch rate. We also let F_c equal the unobstructed flux of the chemical species for both the vertical and lateral etch rates. This assumes completely isotropic behavior of the chemical species giving equal flux everywhere on the surface, even for narrow etch features, due to a low S_c .

Thus:

$$\frac{r_{\text{lat}}}{r_{\text{ver}}} = \frac{\left(\frac{S_c K_f F_c}{N} \right)}{\left(\frac{S_c K_f F_c + K_i F_i}{N} \right)} = \frac{S_c K_f F_c}{S_c K_f F_c + K_i F_i} = 0.2$$

Rearranging gives:

$$\begin{aligned} \frac{S_c K_f F_c}{S_c K_f F_c + K_i F_i} &= 0.2 \Rightarrow S_c K_f F_c = 0.2(S_c K_f F_c + K_i F_i) \\ \Rightarrow 0.8(S_c K_f F_c) &= 0.2(K_i F_i) \Rightarrow S_c K_f F_c = 0.25(K_i F_i) \\ \Rightarrow F_c &= \frac{0.25(K_i F_i)}{S_c K_f} \end{aligned}$$

Plugging in the numbers gives:

$$\begin{aligned}
 F_c &= \frac{0.25(K_i F_i)}{S_c K_f} \\
 &= \frac{0.25(1 * 1 \times 10^{16} \text{ cm}^2 \text{ sec}^{-1})}{0.01 * 0.1} \\
 &= 2.5 \times 10^{18} \text{ cm}^2 \text{ sec}^{-1}
 \end{aligned}$$

Any higher chemical flux would give a lower anisotropy.

10.11. We want to see how the etch rate in the vertical direction might depend on pressure assuming that the etch follows the saturation/adsorption model. Assume that for a particular etch system that the chemical flux is directly proportional to the pressure, while the ion flux is inversely proportional to the pressure. That is $F_c = F_c' * P$ and $F_i = F_i' / P$. (P is normalized to 1 atm and unitless.) Also assume that density = 1 atom/nm³, and that $K_i F_i' = S_c F_c' = 1$ atom/nm²/sec.

a. Plot the vertical etch rate versus pressure, P , from $P = 0$ to 10.

b. Repeat with $K_i F_i' = 40$ atoms nm⁻² sec⁻¹ and $S_c F_c' = 1$ atom nm⁻² sec⁻¹.

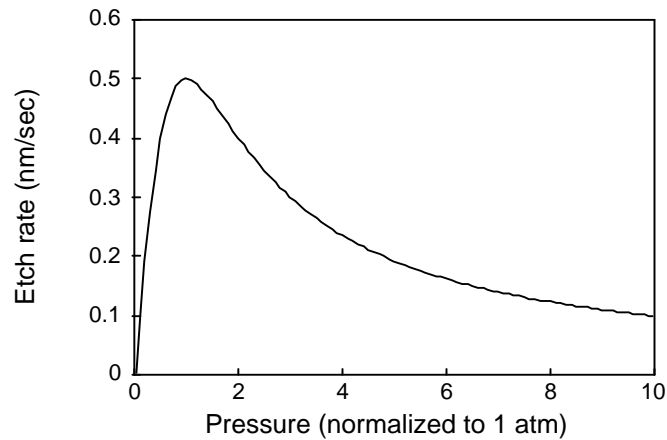
Answer:

In the vertical direction, the etch rate will have contributions from both the chemical and ionic etch components, but acting in a synergistic fashion. For the saturation/adsorption etch model (ion enhanced etching):

$$\text{Etch Rate} = \frac{1}{\text{density}} \frac{1}{\left(\frac{1}{K_i F_i} + \frac{1}{S_c F_c} \right)}$$

Plugging in $F_i = F_i' / P$ and $F_c = F_c' * P$, the density = 1 atom/nm³, and that $K_i F_i' = S_c F_c' = 1$ atom/nm²/sec gives:

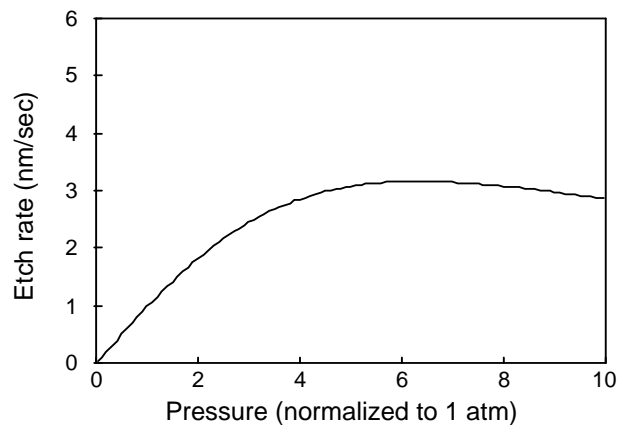
$$\begin{aligned}
 \text{Etch rate} &= \frac{1}{\text{density}} \frac{1}{\left(\frac{1}{K_i F_i' / P} + \frac{1}{S_c F_c' * P} \right)} \\
 &= \frac{1}{1 \text{ atom/nm}^3} \frac{1 \text{ atom/nm}^2 / \text{sec}}{\left(\frac{1}{1/P} + \frac{1}{1 * P} \right)} \\
 &= \frac{1}{\left(P + \frac{1}{P} \right)} \text{ nm/sec}
 \end{aligned}$$



b. For $K_i F_i' = 40 \text{ atoms/nm}^2/\text{sec}$ and $S_c F_c' = 1 \text{ atom/nm}^2/\text{sec}$:

Etch rate

$$\begin{aligned}
 &= \frac{1}{\text{density}} \frac{1}{\left(\frac{1}{K_i F_i' / P} + \frac{1}{S_c F_c' * P} \right)} \\
 &= \frac{1}{1 \text{ atom/nm}^3} \frac{1 \text{ atom/nm}^2/\text{sec}}{\left(\frac{1}{40/P} + \frac{1}{1 * P} \right)} \\
 &= \frac{1}{\left(\frac{P}{40} + \frac{1}{P} \right)} \text{ nm/sec}
 \end{aligned}$$



10.12. In an etch process, there is a finite amount of purely chemical etching without any ion bombardment (i.e. spontaneous chemical etching). In addition, ion bombardment greatly increases the etch rate by facilitating the breaking up of

SILICON VLSI TECHNOLOGY 142 © 2000 by Prentice Hall
Fundamentals, Practice and Modeling Upper Saddle River, NJ.
By Plummer, Deal and Griffin

the etch precursor. At high ion flux the etch rate saturates. No etching occurs when there is only ion bombardment with no chemical component.

- Write a generalized etch rate equation that can describe this behavior.
- Sketch an etch rate versus ion flux curve for this process for some non-zero chemical flux.
- Sketch what the etch profile might look like for this process (i.e. etching through a window in a mask as in Figure 10-3).

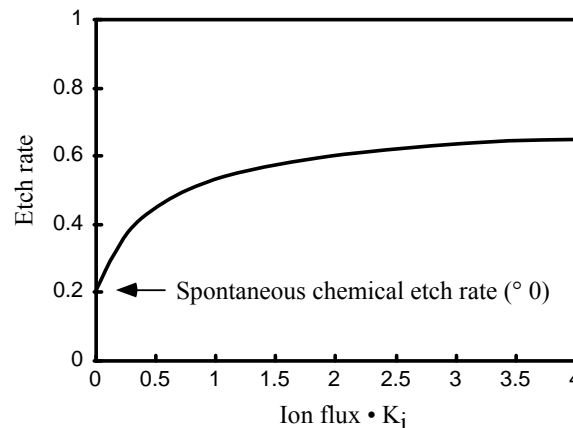
Answer:

- This etch process is a combination of purely chemical etching and ion-enhanced etching (with no purely ionic or physical etching). Purely chemical etching can be given by the first term in the linear etch rate expression, Eq. 10.15, while ion enhanced etching can be written by the series expression given in the text, Eq. 10.20, regardless of the exact physical mechanism. The two parts can be added together to give:

$$\text{Etch Rate} = K_f S_c F_c + \frac{1}{\text{density}} \frac{1}{\left(\frac{1}{K_i F_i} + \frac{1}{S_c F_c} \right)}$$

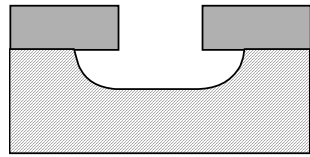
K_f can include any ratio'ing factor between the spontaneous chemical etching and the ion-enhanced chemical component since they will probably have different relative rates. More complex expressions can be derived, such as including any coupling between the spontaneous etching and ion bombardment or surface coverage.

- The etch rate versus ion flux curve would be like that for ion enhanced etching, plus a spontaneous, purely chemical component, which would shift the ion enhanced curve up by that amount of spontaneous chemical etching:

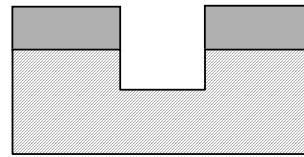


- Again, the profile would reflect a linear combination of isotropic purely chemical etching and very directional ion-enhanced etching, giving an anisotropic

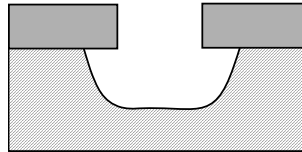
etch profile with some lateral etching but with a vertical etch depth greater than the lateral etch distance:



PLUS



GIVING:



Chapter 11 Problems

11.1. Replot the delay time versus chip area curve (Figure 11-4) for $F_{\min} = 0.25 \mu\text{m}$, for the following cases (all on the same graph), assuming no driver or load elements:

- a. With Al as the interconnect metal (assume the resistivity of Al is $3.0 \mu\Omega \text{ cm}$) and SiO_2 as the dielectric (assume $K_{\text{ox}} = 3.9$).**
- b. With Al as the interconnect metal but with a new dielectric material, whose dielectric constant, K , is equal to 2.2.**
- c. With Cu as the interconnect metal (resistivity is $1.7 \mu\Omega \text{ cm}$) and a low K dielectric material (K is equal to 2.2).**

Answer:

From Equation 11.6:
$$\tau_L = 0.89 \cdot K_{\text{ox}} \epsilon_0 \rho \frac{A}{(F_{\min})^2}$$

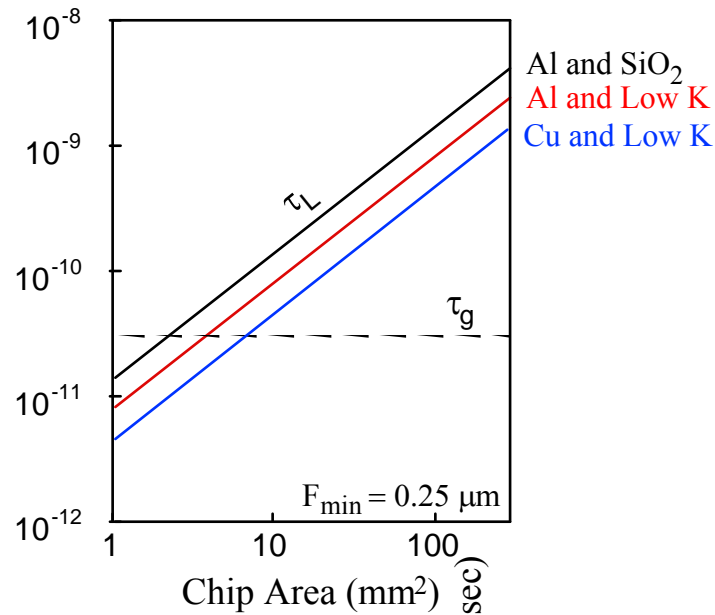
Plugging in:

$$\begin{aligned} K_{\text{ox}} &= 3.9 \text{ for SiO}_2 \\ \epsilon_0 &= 8.86 \times 10^{-14} \text{ F/cm} \\ \rho &= 3.0 \times 10^{-6} \Omega \text{ cm for Al} \\ A &= 1 \text{ mm}^2 = 0.01 \text{ cm}^2 \\ F_{\min} &= 0.25 \mu\text{m} = 0.25 \times 10^{-4} \text{ cm} \end{aligned}$$

Al and SiO_2 , for area = 1 mm^2 : $\tau_1 = 1.5 \times 10^{-11} \text{ sec}$
(This is just the bottom curve on the far right plot in Figure 11.3.)

Al and low K : $\tau_2 = \tau_1 \times 2.2/3.9 = 8.4 \times 10^{-11} \text{ sec}$

Cu and low K : $\tau_3 = \tau_2 \times 1.7/3.0 = 4.8 \times 10^{-11} \text{ sec}$



11.2. Calculate the percentage increase in the interconnect RC delay according to Eq. 11.3 if the thicknesses H and x_{ox} , remain constant while the lateral dimensions W and L_S scale with (and equal) F_{min} . Assume F_{min} is decreased from 0.5 to 0.35 μm and H and x_{ox} equal 0.5 μm . Also assume that the interconnect length L remains constant.

Answer:

From Equation 11.3, the interconnect delay is:

$$\tau_L = 0.89 \cdot K_I K_{ox} \epsilon_o \rho L^2 \left(\frac{1}{H \cdot x_{ox}} + \frac{1}{W \cdot L_S} \right)$$

Assuming that H and x_{ox} equal 0.5 μm and W and L_S equal F_{min} , then the equation becomes:

$$\begin{aligned} \tau_L &= 0.89 \cdot K_I K_{ox} \epsilon_o \rho L^2 \left(\frac{1}{0.5\mu\text{m} \cdot 0.5\mu\text{m}} + \frac{1}{F_{min} \cdot F_{min}} \right) \\ &= 0.89 \cdot K_I K_{ox} \epsilon_o \rho L^2 \left(\frac{1}{0.25\mu\text{m}^2} + \frac{1}{(F_{min})^2} \right) \\ &= \text{constant} \cdot \left(\frac{1}{0.25\mu\text{m}^2} + \frac{1}{(F_{min})^2} \right) \end{aligned}$$

For F_{min} equal to 0.5 μm , the interconnect delay equals

$$\begin{aligned}
 \tau_L &= \text{constant} \cdot \left(\frac{1}{0.25\mu\text{m}^2} + \frac{1}{(0.50\mu\text{m})^2} \right) \\
 &= \text{constant} \cdot (4.0 + 4.0) \\
 &= \text{constant} \cdot 8.0
 \end{aligned}$$

For F_{\min} equal to $0.35\mu\text{m}$, the interconnect delay equals

$$\begin{aligned}
 \tau_L &= \text{constant} \left(\frac{1}{0.25\mu\text{m}^2} + \frac{1}{(0.35\mu\text{m})^2} \right) \\
 &= \text{constant} \cdot (4.0 + 8.2) \\
 &= \text{constant} \cdot 12.2
 \end{aligned}$$

Thus the interconnect delay increases by $12.2 / 8.0 = 1.52$ times, or about 50%, as the feature size decreases - less than if the vertical dimensions were also scaled (about 100% increase), but still significant considering that the gate delay decreases as F_{\min} is decreased.

11.3. Al spiking of Si would not be a problem if the Si diffusion in Al were small and little Si diffused into the Al, leaving voids in the Si for the Al to fill. Just how much does the Si diffuse into the Al? If the diffusivity of Si in Al at 450°C is about $40\mu\text{m}^2/\text{min}$, calculate how far the Si will diffuse into the Al after 60 minutes at 450°C ? (Calculate the distance in which the Si concentration falls to 50% of the surface concentration.)

Answer:

We assume the Si substrate acts as an infinite source of Si. Using Eq. 7.26 from the diffusion chapter or Eqn. 9.17 from the deposition chapter, where $C_s = C(\text{substrate})/2$, gives

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

For $C(x, t)/C_s = 0.5$:

$$0.5 = \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$

$$0.5 = 1 - \operatorname{erf}\left(\frac{x}{2\sqrt{Dt}}\right)$$

$$0.5 = \operatorname{erf}\left(\frac{x}{2\sqrt{Dt}}\right)$$

From erf Table :

$$0.48 = \frac{x}{2\sqrt{Dt}} \quad \rightarrow \quad x = 2 * 0.48 * \sqrt{Dt}$$

$$x = 2 * 0.48 * \sqrt{Dt}$$

$$x = 0.96 \sqrt{40 \mu\text{m}^2 \text{ min}^{-1} * 60 \text{ min}} = 47 \mu\text{m}$$

Yep, that's pretty far, allowing a lot of Al to backfill into the Si.

- 11.4. For 0.35 μm technology, the junction depth (before silicidation) is about 100 nm. If you want to leave 50 nm of Si after silicidation to ensure low leakage current, how much TiSi_2 is formed and how much Ti is needed if all is consumed?**

Answer:

You need to leave 50 nm of Si in the junction, so you want to consume 100-50 nm = 50 nm Si.

$$2.51 \times 50 / 2.27 = 55 \text{ nm TiSi}_2 \text{ formed}$$

$$55 / 2.51 = 22 \text{ nm Ti consumed}$$

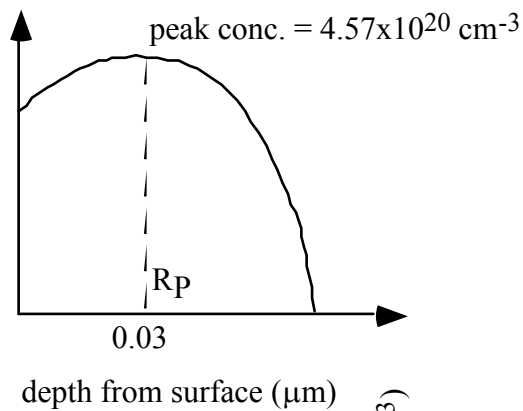
- 11.5. A source/drain implant is done with arsenic with a dose of $1 \times 10^{15} \text{ cm}^{-2}$ and an energy of 40 keV. A titanium silicide layer is then formed on top of the source/drain regions to reduce the sheet resistance of those regions. This is done by depositing Ti on the surface and annealing. 55 nm of silicide is formed, which consumes the top surface of the Si in the source/drain regions and as a result reduces the amount of As dopant in the Si. What is the peak concentration of As in the Si in the source/drain regions after the silicidation process? (Assume that the implant is done directly into the Si with no oxide and that no dopant diffusion or segregation occurs during the silicidation.)**

Answer:

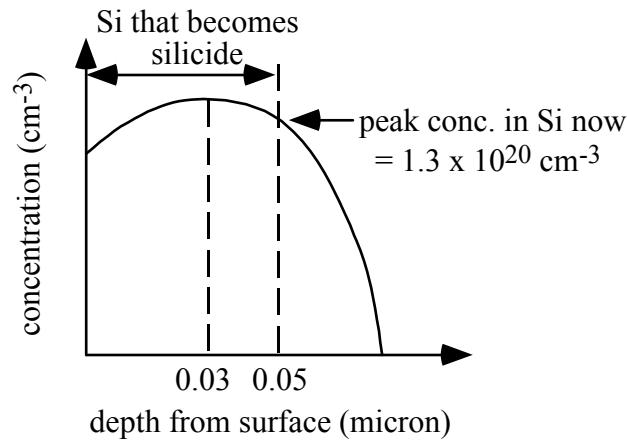
From Figure 8-3 in Chapter 8, As implanted at an energy of 40 keV gives $R_p =$ about $0.03 \mu\text{m}$ and $\Delta R_p = 0.013 \mu\text{m}$. Therefore the equation for the implanted profile in the Si is (assuming Gaussian implanted profile):

$$\begin{aligned} n(x) &= \frac{Q}{\sqrt{2\pi}\Delta R_p} \exp\left(\frac{-(x - R_p)^2}{2(\Delta R_p)^2}\right) \\ &= \frac{1 \times 10^{15} \text{ cm}^{-2}}{\sqrt{2\pi} \cdot 0.013 \times 10^{-4} \text{ cm}} \exp\left(\frac{-(x - 0.03 \times 10^{-4} \text{ cm})^2}{2(0.013 \times 10^{-4} \text{ cm})^2}\right) \\ &= 4.57 \times 10^{20} \text{ cm}^{-3} \exp\left(\frac{-(x - 0.03 \times 10^{-4} \text{ cm})^2}{3.38 \times 10^{-12} \text{ cm}^2}\right) \end{aligned}$$

This is plotted below:



2.27 nm of Si are consumed for every 2.51 nm of silicide formed. The silicidation process thus consumes $55 \text{ nm} * (2.27 \text{ nm} / 2.51 \text{ nm}) = 50 \text{ nm} = 0.050 \mu\text{m}$ of the Si. This is shown below:



The peak concentration of dopant in the Si is now just the surface concentration, which can be determined by letting $x = 0.05$ in the expression above:

$$n(x = 0.05 \text{ micron}) = 4.57 \times 10^{20} \text{ cm}^{-3} \exp \left(\frac{-\left(0.05 \times 10^{-4} \text{ cm} - 0.03 \times 10^{-4} \text{ cm}\right)^2}{3.38 \times 10^{-12} \text{ cm}^2} \right)$$

$$= 4.57 \times 10^{20} \text{ cm}^{-3} \cdot 0.306 = 1.33 \times 10^{20} \text{ cm}^{-3}$$

Therefore the peak concentration of As in the Si in the source/drain regions after the silicidation process is: $1.3 \times 10^{20} \text{ cm}^{-3}$.

11.6. What are two reasons why the damascene process (single damascene version) might be used instead of the normal masked plasma etch process?

Answer:

1 - The damascene process, by filling a hole or trench and etching back, provides for planarization, such as in the W plug process. In the normal masked plasma etch process, a non-planar topography often results.

2 - The damascene process might be used when a good plasma etch process can not be developed for a film, such is for Cu (due to the low volatility of the etch products primarily.) In a damascene process, CMP can be utilized for the etchback, which can etch back almost anything, and a masked plasma etch of that material is not required.

- 11.7. An oxide layer is deposited over 0.5- μm -thick metal lines, as in Figure 11-17. The resulting step height in the oxide layer is 0.4 μm . What is the degree of planarization?**

Answer:

By Eq. 11.17, the degree of planarization achieved is:

$$\text{DOP} = 1 - \frac{x_{\text{step}}^f}{x_{\text{step}}^i} = 1 - \frac{0.4 \mu\text{m}}{0.5 \mu\text{m}} = 0.2$$

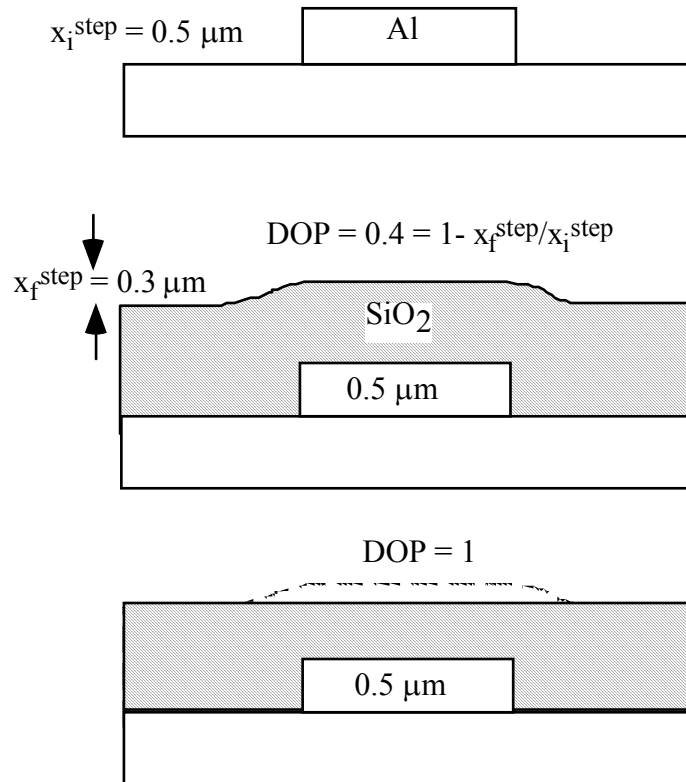
- 11.8. 0.5- μm -thick Al lines are formed on a flat surface. A layer of SiO_2 (at least 1 μm thick) is then deposited over this. A degree of planarization (DOP) of 1 is desired but the actual DOP obtained was 0.4. A CMP etchback of the oxide is then done to obtain a net degree of planarization of 1. If the CMP etch rate is 0.12 μm per minute, what is the minimum CMP etch time needed to obtain this net DOP of 1? Assume no variations in thicknesses or CMP etch rates.**

Answer:

By Equation 11.17:
$$\text{DOP} = 1 - \frac{x_{\text{step}}^f}{x_{\text{step}}^i}$$

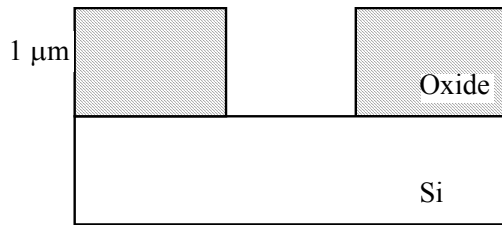
DOP = 0.4 means $x_{\text{step}}^f = (1.0 - 0.4) \cdot 0.5 = 0.3 \mu\text{m}$. To get DOP = 1, need to etch 0.3 μm oxide.

$$\text{Time} = \text{thickness/etchrate} = 0.30/0.12 = 2.5 \text{ min}$$

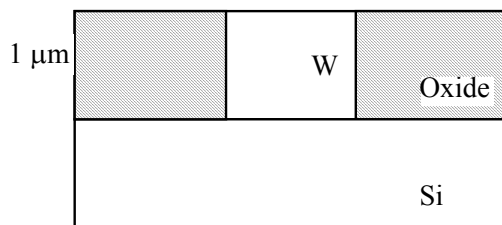
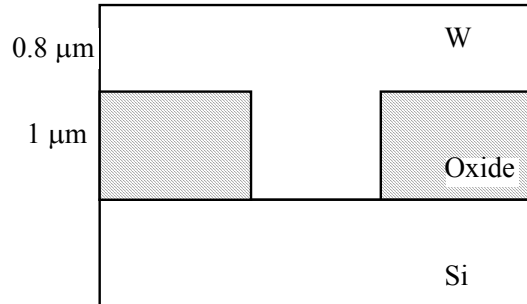


- 11.9. A damascene process is used to fabricate a tungsten via through an SiO_2 dielectric layer. First the SiO_2 dielectric layer is deposited, with a thickness of 1 μm. A via hole is etched in the SiO_2 , and W is blanket deposited by CVD. Right after the W deposition, the degree of planarization (DOP) equals 1.0, and the thickness of the W directly above the dielectric layer is equal to 0.8 μm. A plasma etchback of the W layer is now done to remove the W that covers the top of the dielectric layer, and leaving the W only in the via hole. If the etch rate of the W in the etchback process is 5.0 nm sec⁻¹, and the etch selectivity of W with respect to SiO_2 is 4:1, what is the profile of the structure after 180 sec of etching? Specify the heights of the W and SiO_2 layers in nm. (Neglect any variations in thicknesses or etch rate.)

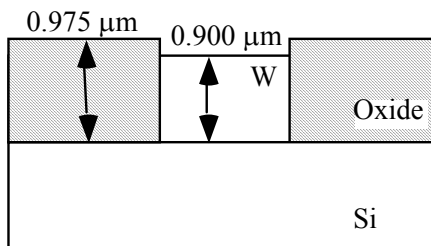
Answer:



DOP = 1 (flat)



After 800 nm / 50 nm per sec
= 160 sec, have etched down
to top of W.



During remaining 20 sec, will etch:
W: 5.0 nm/sec x 20 sec = 100 nm
Oxide: 5.0nm/sec x 1/4 x 20 sec
= 25 nm
Remaining W: 1,000nm - 100nm
= 900 nm
Remaining oxide: 1,000nm- 25nm
= 975nm

Thus: 900 nm for W, 975 nm for SiO₂.

11.10. An SiO₂ layer is deposited by LPCVD over a single 1-μm-wide, 0.5-μm-high metal line. The deposition flux and time are such that 0.25 μm of SiO₂ would be deposited on a flat surface. The sticking coefficient for this process is 0.01 and the arrival angle distribution parameter n is equal to 1. Very conformal deposition is obtained, but the DOP is 0. Would changing n or S_c improve (increase) the DOP? Explain.

Answer:

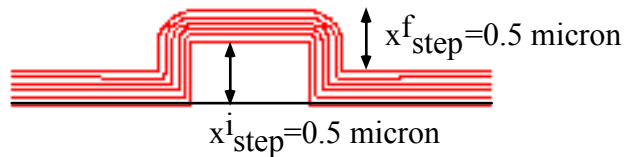
No changing either or both would not help the DOP. The degree of planarization (DOP) is defined as

$$\text{DOP} = 1 - \frac{x_{\text{step}}^f}{x_{\text{step}}^i}$$

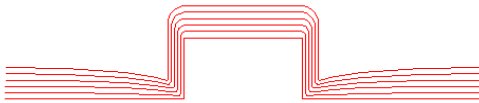
where x_{step}^i is the initial step height and x_{step}^f is the final step height.

Changing either n or S_C will only affect the sidewall coverage, as shown below, not increase the DOP. (In fact, it can actually get a little worse if you measure the step height right at the cusp point, in which case for the two bottom examples, the DOP is negative.)

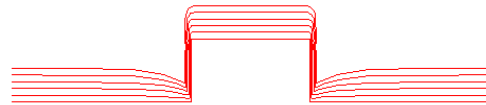
$S_C = 0.01, n = 1$ (~ same if $n = \text{anything}$) DOP = 1



$S_C = 1, n = 1$



$S_C = 1, n = 5$



11.11. SiO₂ is deposited on a Si wafer at 800°C. How much stress is induced in the film due to the cooling of the wafer from 800°C to 25°C? Is the stress tensile or compressive? Assume no intrinsic stress in the film and no stress relaxation processes.

Answer:

The stress induced is due to the difference in thermal expansion coefficients (or thermal contraction in this case). Eqn. 11.21 is used to calculate the stress due to this:

$$\sigma_f = (\alpha_s - \alpha_f) \Delta T \frac{Y_f}{1 - \nu_f}$$

From Table 11-4, α_s is $2.6 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$, α_f is 0.55×10^{-6} , and $Y/(1-\nu)$ is 0.83×10^5 MPa.

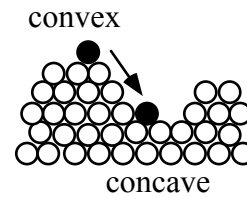
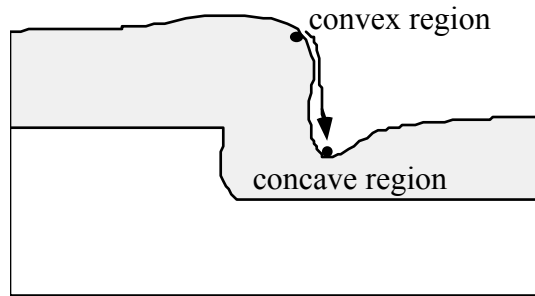
$$\begin{aligned} \sigma_f &= ((2.6 - 0.55) \times 10^{-6} \text{ }^\circ\text{C}^{-1}) (25 - 800 \text{ }^\circ\text{C}) (0.83 \times 10^5 \text{ MPa}) \\ &= -132 \text{ MPa} = -1.32 \times 10^9 \text{ dynes cm}^{-2} \end{aligned}$$

Since the SiO₂ film contracts less than the Si upon cooling, the film will be in compressive stress, as indicated by the negative sign for the stress

11.12. Consider a surface that has a region that is concave and as well as a region that is convex. During reflow, will atoms move from a concave region to a convex region, or from a convex region to a concave region? Explain both in terms of the force equations for reflow as well as in terms of the atomistic picture.

Answer:

A convex region has a positive curvature, while a concave region has a negative curvature according to our conventions. Thus by Eq. 11.31, the force acting on the atom will be positive from a region of positive curvature to a region of negative curvature, similar to diffusion from a region of high concentration to low concentration. Thus atoms will move from a convex region to a concave region during reflow. On an atomistic level, a surface atom in a convex region will be surrounded by fewer atoms and will have less bonds to other atoms (or have distorted bonds at non-equilibrium spacings and angles) compared to a surface atom in a concave region, as illustrated below. The driving force will therefore be from a convex region to a concave region in order to achieve more bonding at equilibrium spacings and lower the energy of the system.



11.13. If the grain boundary mobility, M_{gb} , is $1 \times 10^{-15} \text{ cm}^4 \text{ sec}^{-1} \text{ erg}^{-1}$, the grain boundary energy is 300 ergs cm^{-2} , and the curvature, K , is $20 \text{ } \mu\text{m}^{-1}$, calculate the average grain radius after 60 minutes of grain growth. Assume the initial grain size is zero.

Answer:

By Equation 11.45, the average grain size is

$$\begin{aligned}
 \bar{r} &= \left(r_0^2 + 2M_{gb} \gamma_{gb} t \right)^{1/2} \\
 &= \left(0^2 + 2 \cdot 1 \times 10^{-15} \text{ cm}^4 \text{ sec}^{-1} \text{ erg}^{-1} \cdot 300 \text{ ergs cm}^{-2} \cdot 3600 \text{ sec} \right)^{1/2} \\
 &= 4.6 \times 10^{-5} \text{ cm} = 0.46 \text{ } \mu\text{m}
 \end{aligned}$$

- 11.14.** An Al interconnect line has polygranular clusters that are all equal or less than 5 μm long, and a current density of 0.5 MA cm^{-2} is passed through it. If the critical stress for failure is $0.4 \times 10^9 \text{ dyne cm}^{-2}$ (and let Z^* equal -6), should this line fail due to electromigration? What is the critical current density for which electromigration failure will not occur below for this line?

Answer:

One can use either Eq. 11.63 or 11.64 can be used. For a current density of 0.5 MA cm^{-2} and a critical stress of $0.4 \times 10^9 \text{ dyne cm}^{-2}$, $(L_p)_{\text{crit}}$ is calculated using Eq. 11.64:

$$\begin{aligned} (L_p)_{\text{crit}} &= \left| \frac{2\sigma_{\text{crit}}\Omega}{Z^*q\rho J} \right| \\ &= \left| \frac{2 * 0.4 \times 10^9 \text{ dyne cm}^{-2} * 16 \times 10^{-24} \text{ cm}^3}{-6 * 1.6 \times 10^{-19} \text{ C} * 3 \times 10^{-6} \Omega \text{ cm} * 0.5 \times 10^6 \text{ A cm}^{-2}} * (10^{-7} \text{ dyne cm J}^{-1}) \right| \\ &= 9.0 \times 10^{-4} \text{ cm} = 9.0 \mu\text{m} \end{aligned}$$

Since all the polygranular clusters have lengths less than or equal to 5 μm , which is less than the critical length, the line should not fail. The critical current density for this line can be calculated by Eq. 11.63, using 5 μm for L_p :

$$\begin{aligned} |JL_p|_{\text{crit}} &= \left| \frac{\sigma_{\text{crit}} 2\Omega}{Z^*q\rho} \right| \Rightarrow |J_{\text{crit}}| = \left| \frac{\sigma_{\text{crit}} 2\Omega}{Z^*q\rho L_p} \right| \\ |J_{\text{crit}}| &= \left| \frac{\sigma_{\text{crit}} 2\Omega}{Z^*q\rho L_p} \right| \\ &= \left| \frac{0.4 \times 10^9 \text{ dyne cm}^{-2} * 2 * 16 \times 10^{-24} \text{ cm}^3}{-6 * 1.6 \times 10^{-19} \text{ C} * 3 \times 10^{-6} \Omega \text{ cm} * 5 \times 10^{-4} \text{ cm}} * \left(\frac{J}{10^7 \text{ dyne cm}} \right) \right| \\ &= 9.0 \times 10^5 \text{ A cm}^{-2} \\ &= 0.90 \text{ MA cm}^{-2} \end{aligned}$$

Current densities below this will not cause electromigration damage.