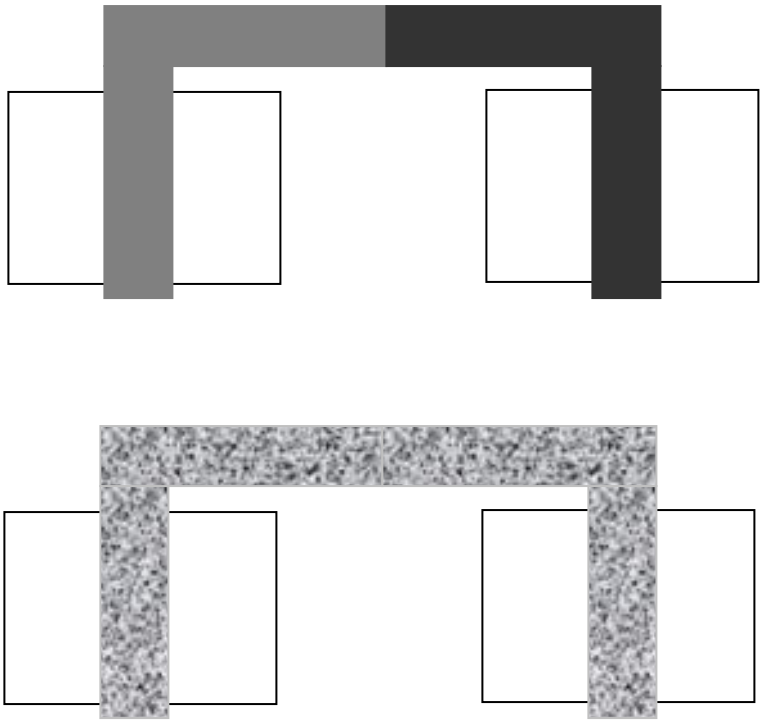
Solutions Summer 2009 – UE 4008

a) With the aid of diagrams describe how certain elements of a MOS transistor have changed in the shrink from processes with minimum dimensions of about 1.0μm to processes with dimensions in the region of 45-60nm. Give a very brief explanation for the technological reasons for these changes. (Hint: concentrate on the "Gate Stack" and substrate.)

Diagram (preferred of the entire gate stack plus substrate showing change areas between older and modern technologies.
In older processes N+ doped polysilicon was used as the gate electrode material for both N and P Channel devices. With small geometry processes this has switched to N doped polysilicon over the N channel devices and P doped polysilicon over the P channel devices. This change adjusts the work function difference between the gate electrode material and the underlying silicon in the channel region, this allows for better control over the threshold voltage. To prevent formation of a diode contact in areas where P and N channel gates are connected and to reduce the polysilicon sheet resistance the polysilicon is combined with a refractory metal such as titanium to create a silicide which effectively shorts out the pn junction that would be formed where the P doped polysilicon meets the n-doped polysilicon.



In processes below the 90nm node nickel silicide is the preferred material for the gate electrode. Using N doped polysilicon over the N-channed transistors and P-doped over the P-Channels also allows reduction of the threshold voltages in line with scaling of the power supply voltages.

Gate dielectric

As Device geometries shrink all of the dimensions of the films are scaled down as well including film thickness and junction depths. It is necessary as well to reduce the power supply voltages. This in turn necessitates the reduction of the Vt. To facilitate this the gate dielectric thickness is made thinner. The traditional material of the gate dielectric has been silicon dioxide because of the exceptional properties of the material for this role. But in processes below 180nm and approaching 100nm the oxide thickness is approaching 1nm this represents a thickness of about 3 atomic layers, difficult to control and the gate leakage current increases. To counteract this materials with higher dielectric constants are being used. The dielectric constant of silicon dioxide is 3.9, materials such as HfO_2 with a dielectric constant of 25 are being used. This enables thicker actual films to be used but with equivalent oxide thickness in the 1nm region. Non-silicon dioxide materials such as HfO_2 are being used from the 45nm node on. In scaling from 90nm to 65nm the gate oxide thickness was not scaled down because of leakage problems

Substrate material

Increasingly strained layers are being used in the substrate material to improve the carrier mobility in the channel region. The strain can be induced in a number of ways; By epitaxially growing a layer of silicon germanium on top of the silicon substrate and then another thin layer of silicon on top of the SiGe. The different size of the germanium atom in the crystal induces strain in the top silicon layer, this in turn improves the carrier mobility in the channel region. This is particularly used to increase the mobility in PMOS transistors. In NMOS transistors the mobility is optimised by thermomechanical design of the sidewall spacers and a strain inducing silicon nitride film.

Diagram here of the strained region.

Many modern process are now also moving to SOI silicon to improve (reduce) substrate capacitance.. SOI Substrates consist of a thin layer of silicon on top of a buried oxide layer (BOX) which in turn sits on a carrier wafer. SOI offers reduced substrate capacitance, better isolation and improved speed of operation performance.

Additional items

Sidewall spacers, these allow for silicidation of the gate electrode and the contact/S/D areas without shorting out the gates to the source and drain.

The Source drains are generally engineered to have high doping at the surface and lower doping through the full S/D region.

The metallization has moved from aluminium alloys to damascene copper with low k dielectric films between the copper layers.

b) In the shrink from 65nm to 45nm a company changes the gate dielectric from to hafnium oxide with a dielectric constant of 25 if the resultant gate capacitance is 3.0×10^{-6} F/cm^2 what is the actual new dielectric thickness and the silicon dioxide equivalent?

Given:

The permittivity of free space is 8.86 X 10^{-14} F/cm
The dielectric constant of silicon dioxide is 3.9

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

$$\varepsilon_{ox} = k\varepsilon_0$$

$$C_{diel} = \frac{k_{diel} \times \varepsilon_0}{t_{diel}}$$

$$3.0 \times 10^{-6} = \frac{25 \times 8.86 \times 10^{-14}}{t_{diel}}$$

$$t_{Haf} = \frac{25 \times 8.86 \times 10^{-14}}{3.0 \times 10^{-6}}$$

$$t_{Haf} = 7.22 \times 10^{-7} cm$$

$$t_{Haf} = 7.22 nm$$

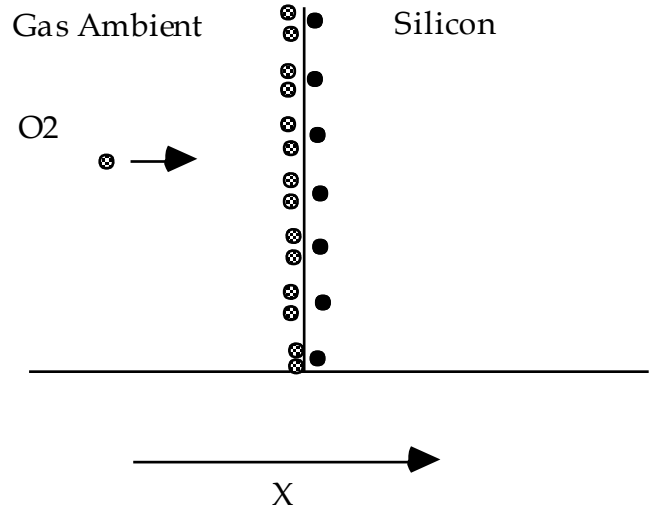
$$t_{ox} = \frac{3.9 \times 8.86 \times 10^{-14}}{3.0 \times 10^{-6}}$$

$$t_{ox} = 1.15 \times 10^{-7} cm$$

$$t_{ox} = 1.15 nm$$

a) Describe a physical model for the thermal oxidation of silicon.
The diagrams below should be explained in text with all of the bullet points below covered.

Initially



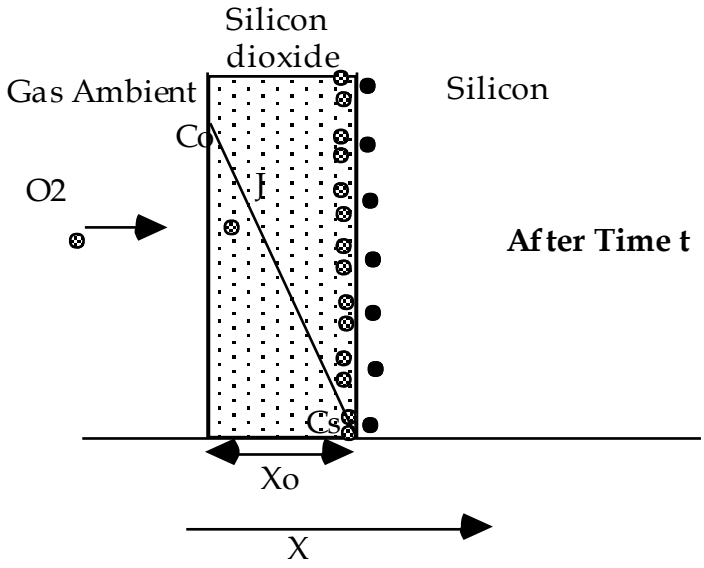


Diagram 1
Reaction controlled
How fast the chemical reaction between the oxidizing species and the silicon can take place determines the oxidation rate

Growth is linear

Diagram 2
There is now an oxide layer on the surface of the silicon
Oxidizing species must diffuse through the oxide
As the oxide gets thicker this takes longer
Growth rate becomes dominated by the diffusion time through the oxide
Growth rate parabolic

b) Why typically does a thermal oxidation finish with a nitrogen anneal cycle? Relate this practice to the electrical parameters of a MOS device.

In the Vt equation the total charge in the oxide is represented by the Q_{tot} term in the V_{fb}. Part of this charge is the fixed charge this to a large extent is determined by the ambient of the last high temperature step that the gate oxide sees.

$$V_t = V_{fb} + 2\psi_p + \frac{\sqrt{2K_s\varepsilon_o q N_B 2\psi_P}}{C_{ox}}$$

$$V_t = 2\psi_p + \frac{\sqrt{2K_s\varepsilon_o q N_B 2\psi_P}}{C_{ox}}$$

This is known as the ideal Vt equation

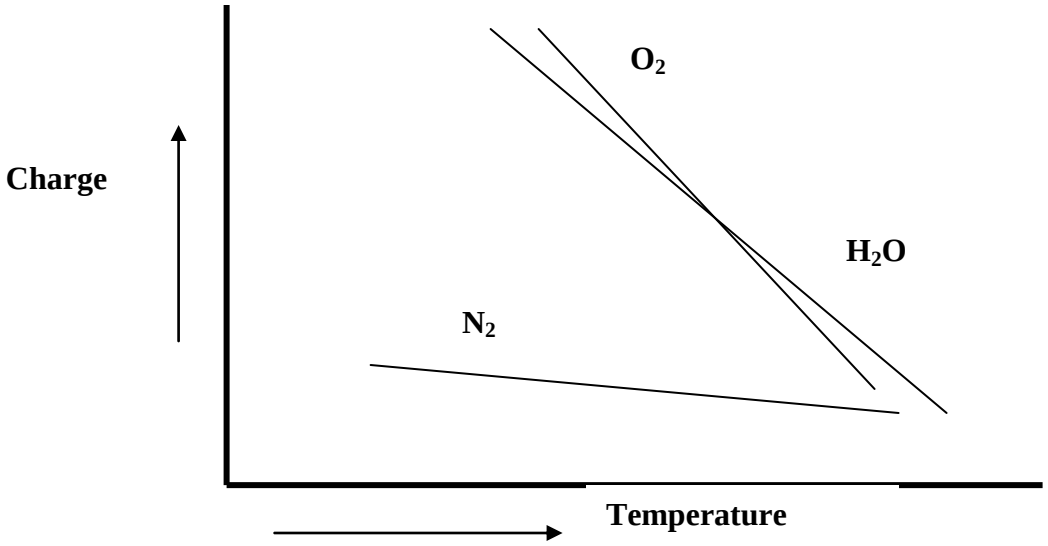
Where

$$V_{fb} = \Phi_{ms} - \frac{Q_{tot}}{C_{ox}}$$

and

$$\psi_p = \frac{kT}{q}\ln\frac{N_B}{n_i}$$

In the Qss triangle plot, the comparison of the fixed charge against anneal temperature and ambient it can be seen that a nitrogen anneal reduces the absolute quantity of charge and is more stable because of a shallower plot slope



For this reason almost all oxidation cycles finish with a nitrogen anneal

c) Describe what happens to boron contained in the silicon during thermal oxidation.

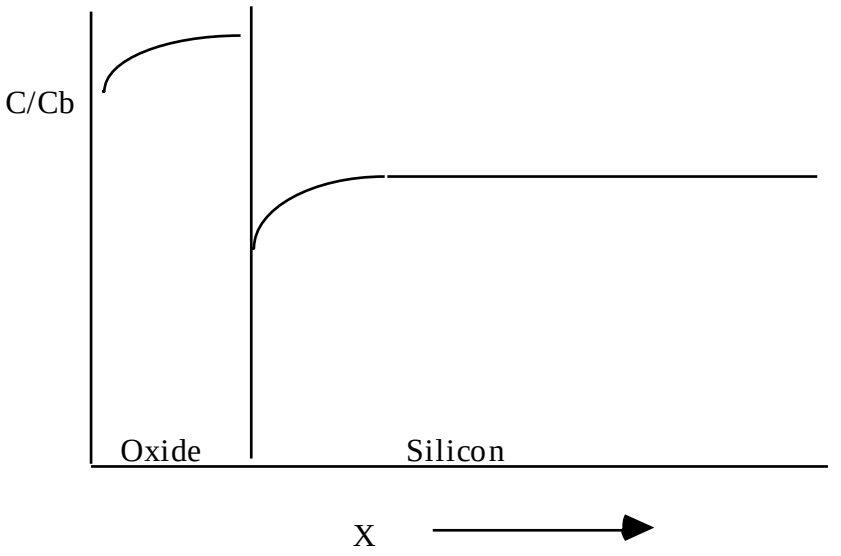
Because most impurities diffuse much slower in silicon dioxide than in silicon we expect that an oxide grown on the surface will effectively seal the impurities in the silicon.

The situation is much more complex!! With any two phases in contact, any impurity will be redistributed between the two phases, until equilibrium is reached. In equilibrium the ratio of the concentrations will be constant.

The ratio of the equilibrium concentrations in the silicon and silicon dioxide is denoted by the term segregation coefficient and is defined by :-

$$M = \frac{\text{Equilibrium conc. of impurity in silicon}}{\text{Equilibrium conc. of impurity in silicon dioxide}}$$

When oxide takes up the impurity M<1



For M>1
Boron Segregation Coefficient 0.3

What happens therefore is that boron close to the interface is “sucked out” into the oxide causing dopant depletion at the surface. This is critically important in the field areas where it is important to keep the doping concentration high to prevent turning on parasitic field devices.

d) A silicon <100> wafer, which had been patterned for ion implant with 90nm of oxide in the windows to be implanted and 0.5μm in the other areas to protect against the implant, is put through a thermal oxide process at 1000°C in pyrogenic steam for 1 hours 40minutes, what is the final oxide thickness in:
   i) The areas which had 90nm on the surface prior to oxidation?
   ii) The areas which had 0.5μm on the surface prior to oxidation?

   i) Calculate from the supplied graph the time it would have taken to grow 90nm on the silicon at 1000°C in steam

   Answer: 10 Minutes

   This time must then be added to the time the wafers are to be in the furnace, 1hour 40 minutes, 100 minutes becomes 110 minutes read off the new thickness from the same graph the final answer is 0.55μm

   ii) Calculate from the supplied graph the time it would have taken to grow 0.5μm on the silicon at 1000°C in steam

   Answer:100 minutes

   Add this time to the new furnace time, also 100 minutes total equivalent time is 200 minutes final oxide thickness in these areas is 0.8μm

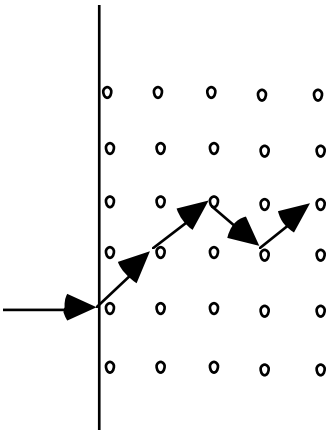
Question 3

a) In relation to a high energy beam of ions striking the surface of crystalline silicon explain the following terms
   (i) Range
   (ii) Projected range
   (iii) Straggle

Range

Ions with a particular energy collide with the atoms of the crystal lattice. These collisions and interactions mean that the incident ions do not have a straight path to their ultimate resting place in the target material.

The ions do not come to rest until they have lost all of their energy through the collisions, there are 2 types of collision Nuclear and Electronic.

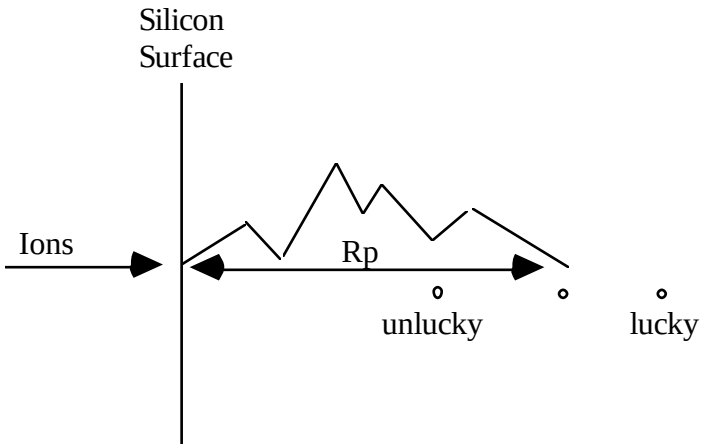


The total Distance Travelled in the Silicon is Called the Range

The total distance the ion travels in the target material is the range generally called R.

Projected range

The projection of the distance along the x-axis is the projected range, Rp, and is the effective distance into the material.



Some ions will be 'lucky' and travel further than the mean distance. Some will be unlucky and travel less far. The distribution is is statistical about a mean point, the peak concentration occurs at the statistical mean, Rp.

Straggle

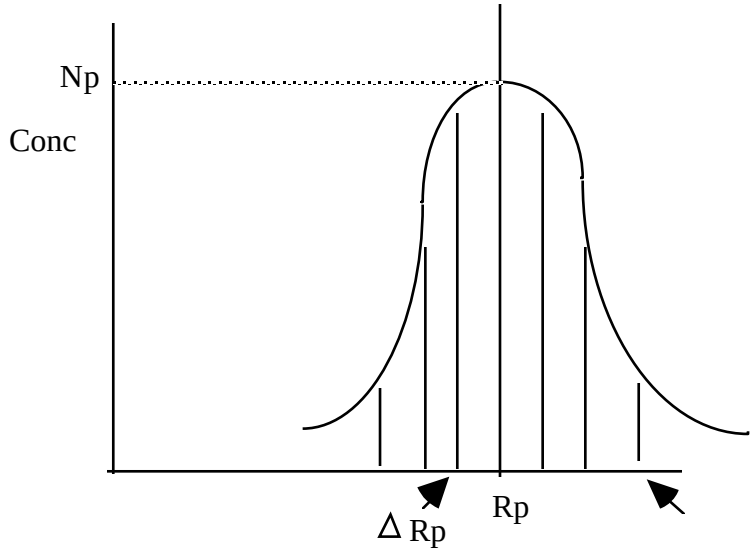
The distribution is characterised by the std. deviation. This is descriptive of the spread of the implanted ions about the mean this standard deviation for ion implant is called the straggle and is assigned the notation ΔRp. The ion distribution is Gaussian with mean Rp and std. deviation ΔRp.

The expression which describes this distribution is :

$$C_{(x,t)} = C_p \exp\left[\frac{-(x-R_p)^2}{2\Delta R_p^2}\right]$$

or

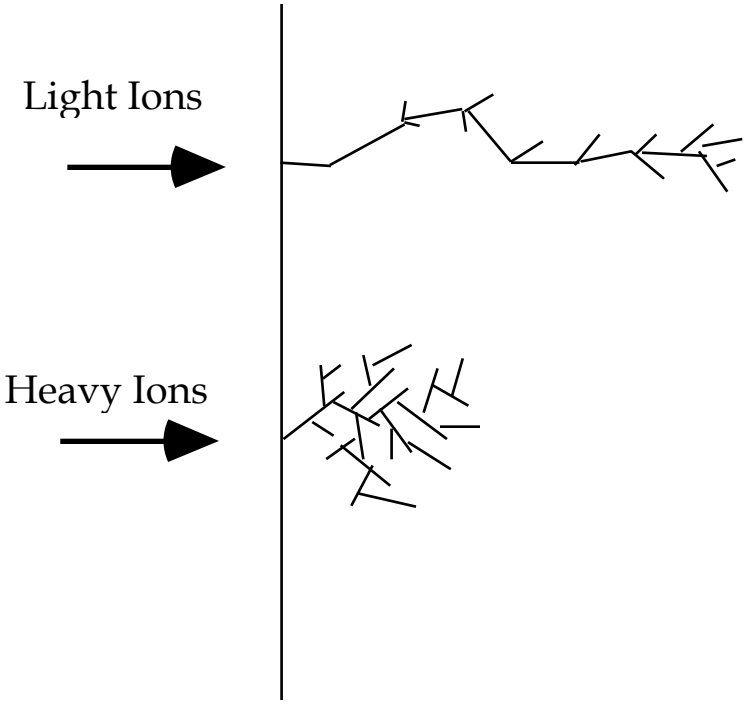
$$N(x) = Np \exp[-(x-Rp)^2/2\Delta Rp^2]$$



Graphic representation of Rp and ΔRp, Np is the peak concentration.

b) Explain the function(s) of the high temperature heat treatment given to wafers following ion implantation

The heat treatment given to wafers following implant serve two functions;
One is to activate the dopant following the implant. Immediately after implant the implanted material is simply sitting in the silicon the dopant atoms have not taken the place of the silicon in the crystal lattice. The heat treatment causes this to happen.
The second thing is that the heat repairs any damage done to the crystal during the implant. Different atoms cause different amounts of damage to the crystal structure. Light ions like boron cause relatively little damage but travel deeper for a given implant energy. Heavier ions/atoms can severely disrupt the crystal structure even to the extent of rendering it amorphous.(Diagram of tree of disorder would be useful)



Counterintuitively above a certain dose when the crystal becomes amorphous less heat (temperature is needed to anneal the damage because of the phenomenon of “Solid Phase Epitaxial Growth” (Plot/Diagram would be useful

c) What energy is needed to place the peak of a phosphorus implant at the silicon/oxide interface if the oxide thickness is 0.07μm, and what thickness of oxide is needed to protect other parts of the silicon from this implant? The implant dose is 1X 10^{16}/cm^{2}and the background concentration is 1 10^{15}/cm^3.

Given:

$C_{(x)} = C_p.\exp[-(x-R_p)^2/2\Delta R_p^2]$

Projected Range and Projected Standard Deviation Graphs attached.

Answer

To place the peak of the implant at the interface, simply a read off of from the projected range graph from 0.07μm on the X axis to the Phosphorus - Oxide line, giving an implant energy of 60KeV

For the thickness of oxide needed to block the implant use the equation

$$Xo = Rp + \Delta Rp\sqrt{2\ln 10 Np / Nb}$$

This assumes that any implant “leaking” into the protected area is less than 1/10 of the background concentration.

From the question R_p is 0.07μm this is a 60KeV implant, ΔR_p is 0.0325μm

N_p is calculated from the equation

$$Q = \sqrt{2\pi} Np \Delta Rp$$

Where Q is the implant dose
From the above figures this gives a value of 1.228 x 10^{21} /cm^3 for N_p

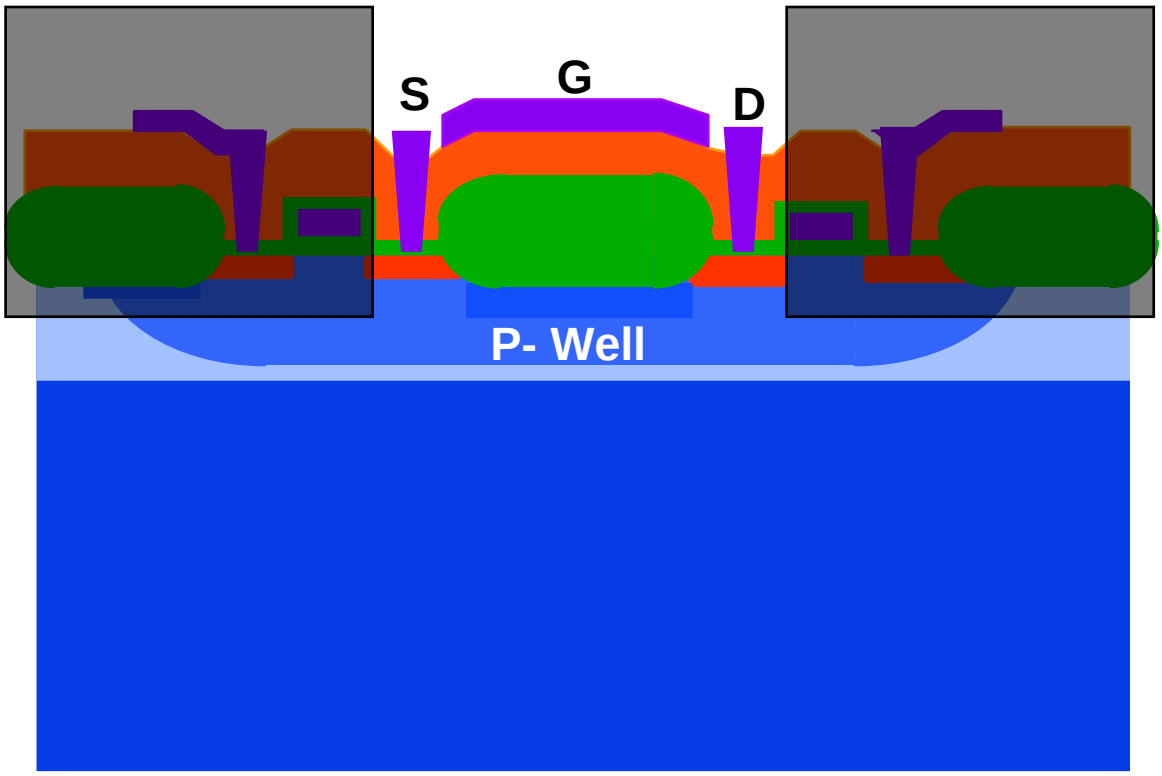
$$Xo = Rp + \Delta Rp \sqrt{2 \ln 10 Np / Nb}$$

Solving this equation for X_o

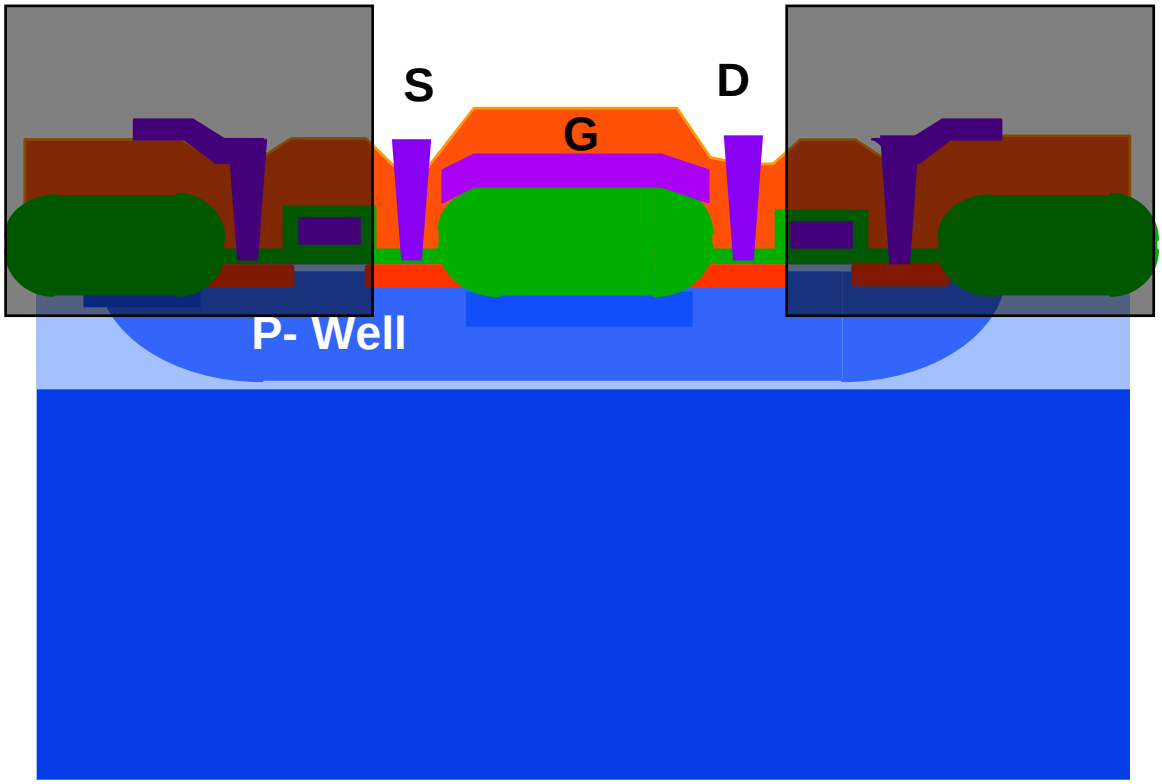
$$Xo = Rp + \Delta Rp \sqrt{32.65}$$
$$Xo = Rp + \Delta Rp (5.71)$$
$$Xo = 0.07 + 0.0325(5.71)$$
$$Xo = 0.07 + 0.186$$
$$Xo = 0.256 \mu m$$

Question 4

a) Describe how parasitic or field devices can be formed in a CMOS process; what measures can be taken to prevent the turn on of these devices, use diagrams to illustrate the answer.



The diagram above shows how a metal gate parasitic MOS transistor can be formed in a CMOS process



The diagram above shows how a polysilicon gate parasitic MOS transistor can be formed in a CMOS process.

Both diagrams must be fully labelled to show the active devices and how the parasitics are formed between adjacent devices.

The measures to be taken to prevent turn-on of these transistors include field implants under the FOX areas.
Important to mention that implants into the boron regions are more important that those into N type areas because of the difference in segregation coefficients and the fact that boron "sucks out" and phos "piles up". These implants must be carried out prior to the FOX growth process.
Also the thickness of the field oxide (FOX) and the BPSG or other dielectric in the case of the metal gate device and the FOX alone in the case polygate combined with the doping in the "channel region of the field device is sufficient to push the field threshold above the power supply voltage. Typically aiming for a value of 10-20% above the PS value.
This could be illustrated by referring to the Vt equation.

b) What can an analogue designer do to increase or decrease the current flow in an NMOS transistor in a mature CMOS process?

In any mature process all of the process parameters have been defined and cannot be changes to suit particular designs. If one looks at the Ids current equation for a device in saturation

$$I_{DS} = \mu_n C_{ox} {}^{W}\!/_{2L} \left(V_G - V_t\right)^2$$

μ_n is the mobility of the carriers in the silicon. This is largely determined by the doping concentration of the substrates. All processes use a particular substrate that is not changed during the life of that process.
Cox is the oxide capacitance, this is determined by the material in, and the thickness of the gate dielectric, again this remains constant within process limitations (thickness uniformity, usually better than 3%)
Vg, the maximum voltage that can be applied to the gate oxide, this is determined by the design power supply voltage. All process are designed to work at a particular power supply voltage. Circuits (chips) are designed to systems to operate at this voltage again this cannot be changed.
Vt, the threshold voltage; one of the key measurement criteria for a MOS process is the threshold voltage for the transistors. This is controlled through setting process parameters such as gate ox thickness, substrate or well doping, metals used in the gate electrodes. God process will control the final Vt to better than 10% of the nominal value, again this cannot be changed to get a particular current value.
The only parameter determining Ids that can be changed is the W/L ratio. In other words by designing the transistor to a particular size a designer can get the current value required.

In an NMOS process with a minimum allowed drawn gate length of 2.0μm, a threshold voltage of 0.8V and a power supply voltage of 10V. What is the smallest device that can be used to deliver a current of 10mA between the source and drain?
Given:
$\mu_n = 1000 cm^2/V.s$
$C_{ox} = 5 \times 10^{-8}\ F/cm^2$

Assume that the maximum current will be delivered with the device in saturation, the current equation that determines this current is:

$$I_{DS} = \mu_n C_{ox} {}^{W}\!/_{2L} \left(V_G - V_t\right)^2$$

To find the minimum sized device that will deliver a 10mA current solve the above equation for W/L, which is the device dimensions.

$$I_{DS} = \mu_n C_{ox} {}^{W}\!/_{2L} \left(V_G - V_t\right)^2$$

$${}^{W}\!/_{L} = \frac{I_{DS}}{\mu_n \times C_{ox} \times 0.5 \times \left(V_G - V_t\right)^2}$$

$${}^{W}\!/_{L} = \frac{10 \times 10^{-3}}{1000 \times 5 \times 10^{-8} \times 0.5 \times \left(10 - 0.8\right)^2}$$

$${}^{W}\!/_{L} = 4.72$$

$$\text{Minimum allowed L is } 2\mu\text{m}$$

$${}^{W}\!/_{2} = 4.72$$

$$W = 9.44$$

$$\text{Smallest transistor to deliver 10mA}$$

$$2 \times 9.44\mu\text{m}$$

$$\text{Or more realistically } 2 \times 9.5 \text{ or } 2 \times 10\mu\text{m}$$