# Chapter 9

Ion Implantation

Professor Paul K. Chu

Ion Implantation

- Ion implantation is a low-temperature technique for the introduction of impurities (dopants) into semiconductors and offers more flexibility than diffusion

- Dopant atoms are volatilized, ionized, accelerated, mass selected, and directed at the wafer

- The atoms enter the crystal lattice, collide with the host atoms, lose energy, and finally come to rest at some depth within the solid

- Ion implantation energies range from several hundred to several million electron volts, resulting in ion distributions with average depths from < 10 nm to 10 μm

- Doses range from 10^{11} atoms/cm^2 for threshold adjustment to 10^{18} atoms/cm^2 for buried dielectric formation
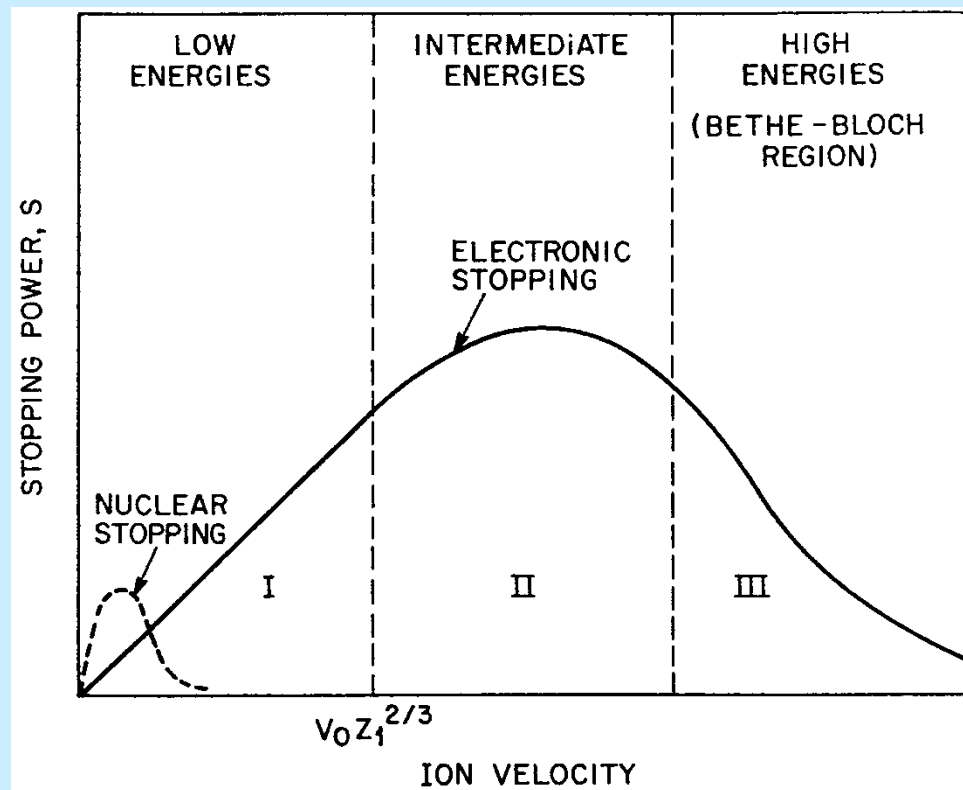
Ion Stopping
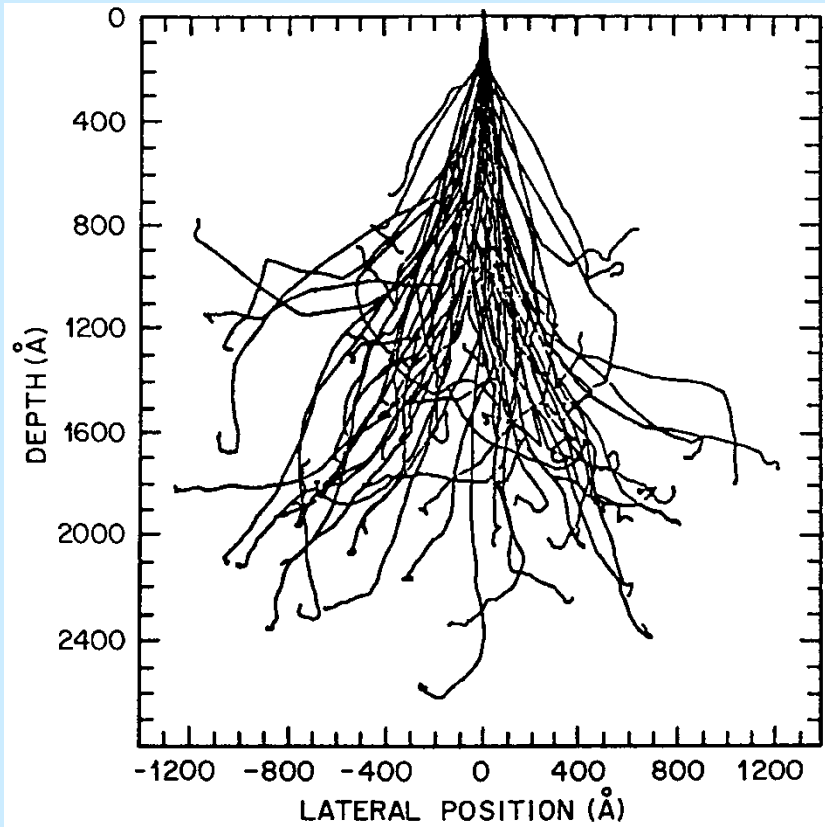
- Implanted ions undergo a series of collisions with the host atoms until they finally stop at some depth
- The initial ion energy, typically several tens of keV, is much higher than lattice binding energies
- The ion scattering process can be simulated based on elastic collisions between pairs of nuclei while ignoring the relatively weak lattice forces
- The second component of scattering comes from inelastic collisions with electrons in the target

The total stopping power **S** of the target, defined by the energy loss (**E**) per unit path length (**x**) of the ion, is

$$S = \left(\frac{dE}{dx}\right)_{nuclear} + \left(\frac{dE}{dx}\right)_{electronic}$$

Nuclear Stopping

- Dominant at low energy

- Caused by a collision between two atoms, and the physics can be described by classical kinematics

- If the atoms were bare nuclei, then at a separation **r**, the coulombic potential between them would be:

$$V_c(r) \;\; = \;\; \frac{q^2 z_1 z_2}{4\pi \varepsilon_o r}$$

where z_1 and z_2 are the atomic number of the implanted and target atoms, ε_o is the permittivity, and q is the electronic charge, respectively

- In reality, electrons screen the nuclear charge and a screening function, $f_s(r)$, must be included such that $V(r) = V_c(r) f_s(r)$

- The equations of motion of atoms can be integrated to yield the scattering angle for any incident ion trajectory

$$T = \frac{4 M_1 M_2}{\left(M_1 + M_2\right)^2} E \; \sin^2\left\{\frac{\theta}{2}\right\}$$

where T is the energy lost by the incoming ion, E is the energy of the ion, θ is the scattering angle, and M_1 and M_2 are the atomic mass numbers of the ion and target atom, respectively

- The rate of energy loss to nuclear collisions per unit path length can be calculated by summing the energy loss multiplied by the probability of that collision occurring

- If the maximum possible energy transfer in a collision is T_{max} and there are N target atoms per unit volume,

$$S_n \;\; = \;\; \left(\frac{dE}{dx} \right)_{nuclear} \;\; = \;\; N \int_0^{T_{max}} T \, d\sigma$$

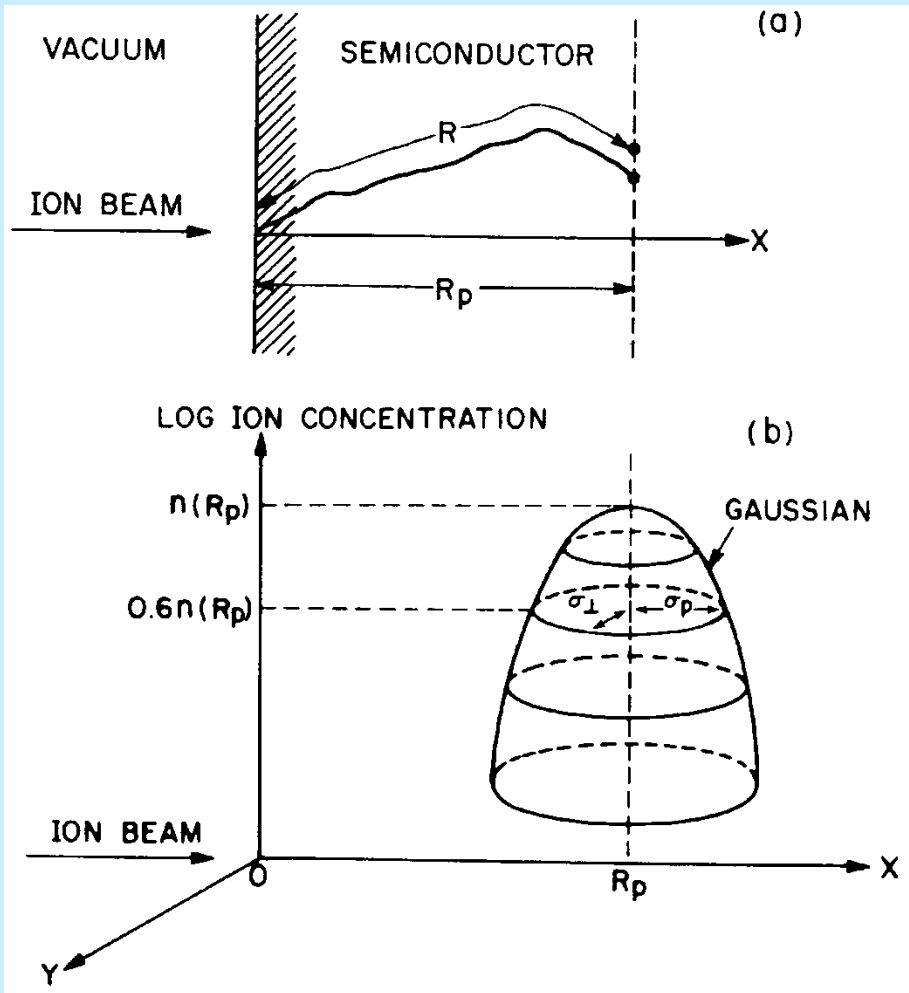where $d\sigma$ is the differential cross section

- **Nuclear stopping is elastic, and so energy lost by the incoming ion is transferred to the target atom that is subsequently recoiled away from its lattice site, thus creating a damage or defect site**

Electronic Stopping

- Dominant at high energy

- Caused by the interaction between the incoming ion and the electrons in the target

- The stopping which is inelastic is similar to a viscous drag force and is proportional to the ion velocity

- The energy loss by incident ions is dissipated through the electron cloud into thermal vibrations of the target

Range Distributions

- The average total path length is called the range R, which is composed of both lateral and vertical motions

- The average depth of the implanted ions is called the projected range R_p

- The distribution of the implanted ions about that depth can be approximated as Gaussian with a standard deviation σ_p (or ΔR_p).

- The lateral motion of the ions leads to a lateral Gaussian distribution with a standard deviation $\sigma_\perp$

VACUUM | SEMICONDUCTOR (a)

ION BEAM

R

R_p

X

LOG ION CONCENTRATION (b)

$n(R_p)$

$0.6n(R_p)$

σ_T σ_p

GAUSSIAN

ION BEAM

R_p X

Y

Far from the mask edge, the lateral motion can be ignored, and **$n(x)$**, the ion concentration at depth **x**, can be written as:

$$n(x) \;=\; n_o \exp\left\{-\frac{\left(x-R_p\right)^2}{2\sigma_p^2}\right\}$$

where **n_o** is the peak concentration, **R_p** is the projected range, and **σ_p** is the standard deviation. This is the Gaussian distribution.

If the total implanted dose is **Φ**, $n_o \;=\; \dfrac{\phi}{\sqrt{2\pi}\,\sigma_p} \;\cong\; \dfrac{0.4\phi}{\sigma_p}$

Pearson IV Fit

- An implant depth profile can be better fitted using a Pearson IV distribution with four moments

- The normalized first moment of the ion distribution is the projected range, $\boldsymbol{R_p}$

- The second moment is the standard deviation, $\boldsymbol{\sigma_p}$

- The third moment is the skewness, a measure of the asymmetry of the distribution - Positive skewness places the peak of the distribution closer to the surface than $\boldsymbol{R_p}$)

- The fourth moment, kurtosis, $\boldsymbol{\beta}$, is an indication of the flatness of the distribution
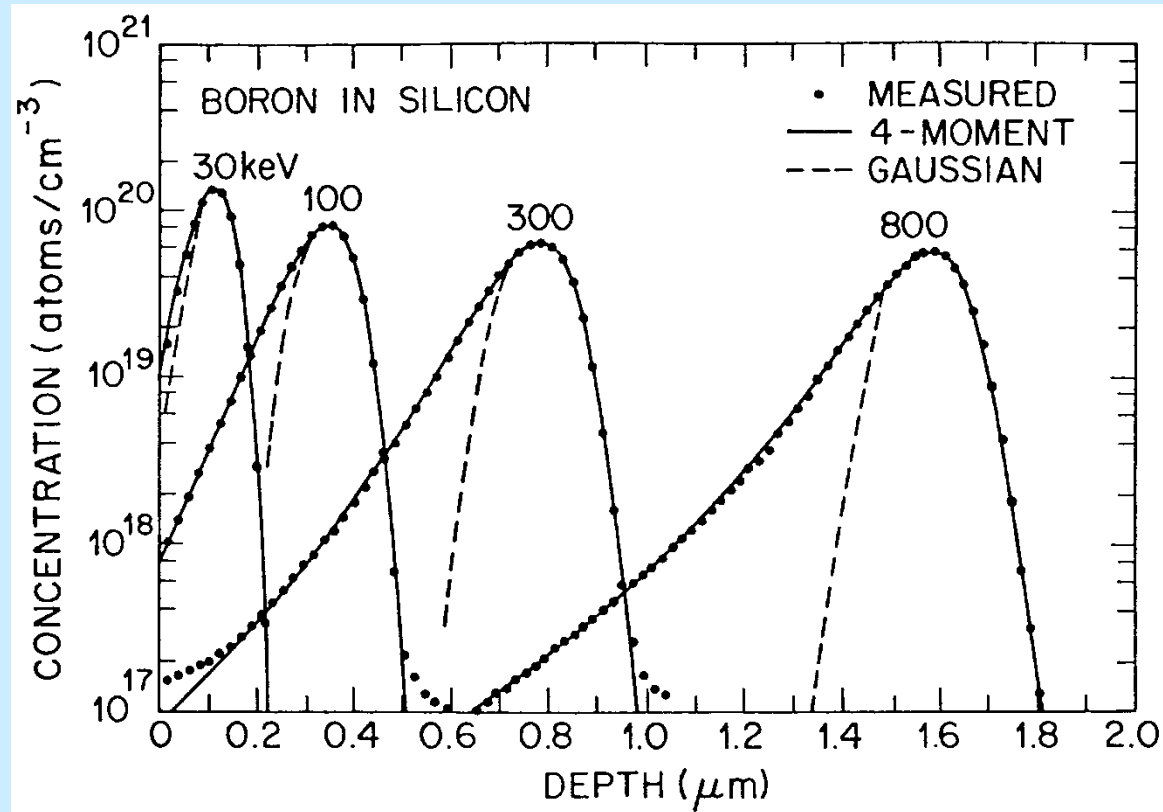
- As the energy is increased, the profiles become more negatively skewed and deviate more from a true Gaussian

- The variation in the skewness can be explained by increased electronic stopping for faster moving ions in the "pre - peak"or "pre - R_p" region

100 keV Boron Implantation

| Materials	Symbol	Density	$R_p(\text{Å})$	$\sigma_p(\text{Å})$
Silicon	Si	2.33	2968	735
Silicon dioxide	SiO_2	2.23	3068	666
Silicon nitride	Si_3N_4	3.45	1883	408
Photoresist	$C_8H_{12}O$	1.37	10569	1202
Titanium	Ti	4.52	2546	951
Titanium silicide	$TiSi_2$	4.04	2154	563
Tungsten	W	19.3	824	618
Tungsten silicide	WSi_2	9.86	1440	555

Mask Edge Ranges
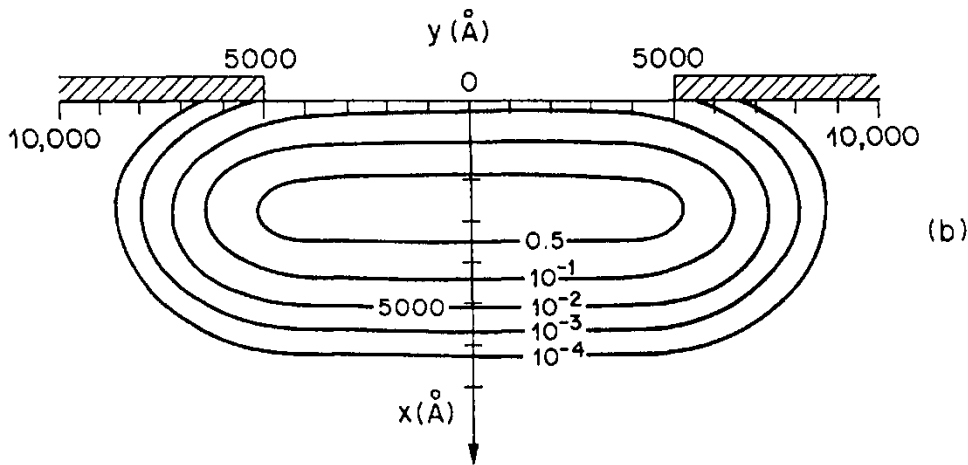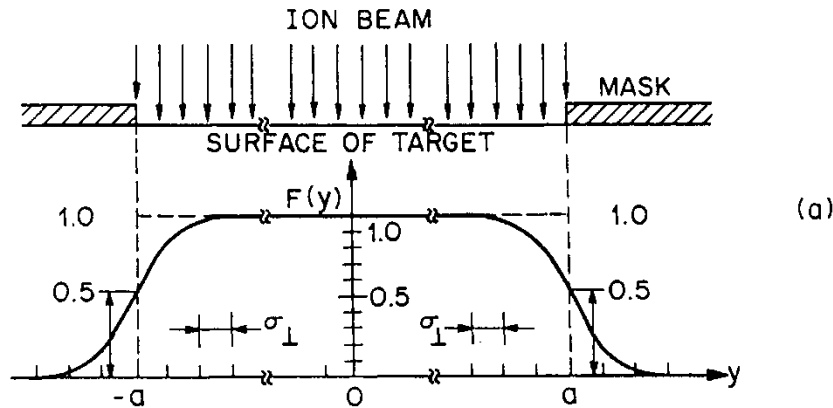
- Use Gaussian to approximate lateral distributions

$$n(x, y) \quad = \quad \frac{n_{vert}(x)}{\sqrt{2\pi}\sigma_\perp} \exp\left(\frac{-y^2}{2\sigma_\perp^2}\right)$$

- To minimize lateral scattering, masking layers are often tapered at the edge rather than perfectly abrupt, so that ions are gradually prevented from entering the silicon

Substrate Damage

- As each ion penetrates the target, it undergoes a series of collisions displacing host atoms along the way

- Both the ion and dislodged target atoms can continue and cause further damage

- Eventually, the energy per particle becomes too small and the cascade stops

- After many ions have been implanted, an initially crystalline target will be so perturbed that it will have changed to a highly disordered state

- If the target temperature is sufficiently high, the competing process of self-annealing can occur to repair some or all of the damage as it is generated
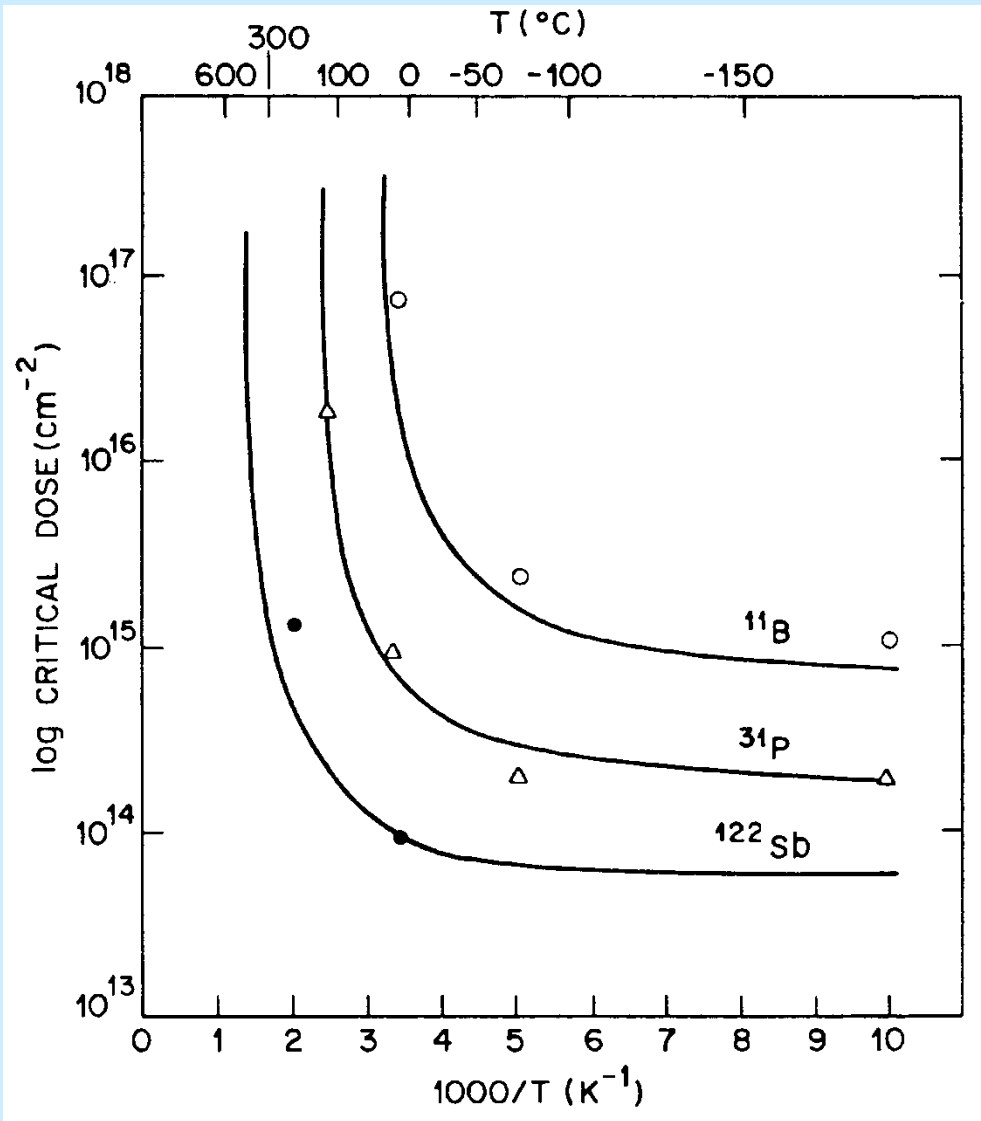
- The critical dose is defined as the minimum necessary to amorphize the target

- Heavy ions displace a greater volume of target atoms per ion, and so a higher temperature is necessary for complete recrystallization

Consequences of Damage

- Point defects influence diffusion in silicon

- Damaged oxide layers etch faster than undisrupted oxide because some of the bonds are already broken

- Dangling oxide bonds can act as electron traps in finished devices

- The organic chains in photoresist films can be broken down into a carbon-rich material which is much harder to remove
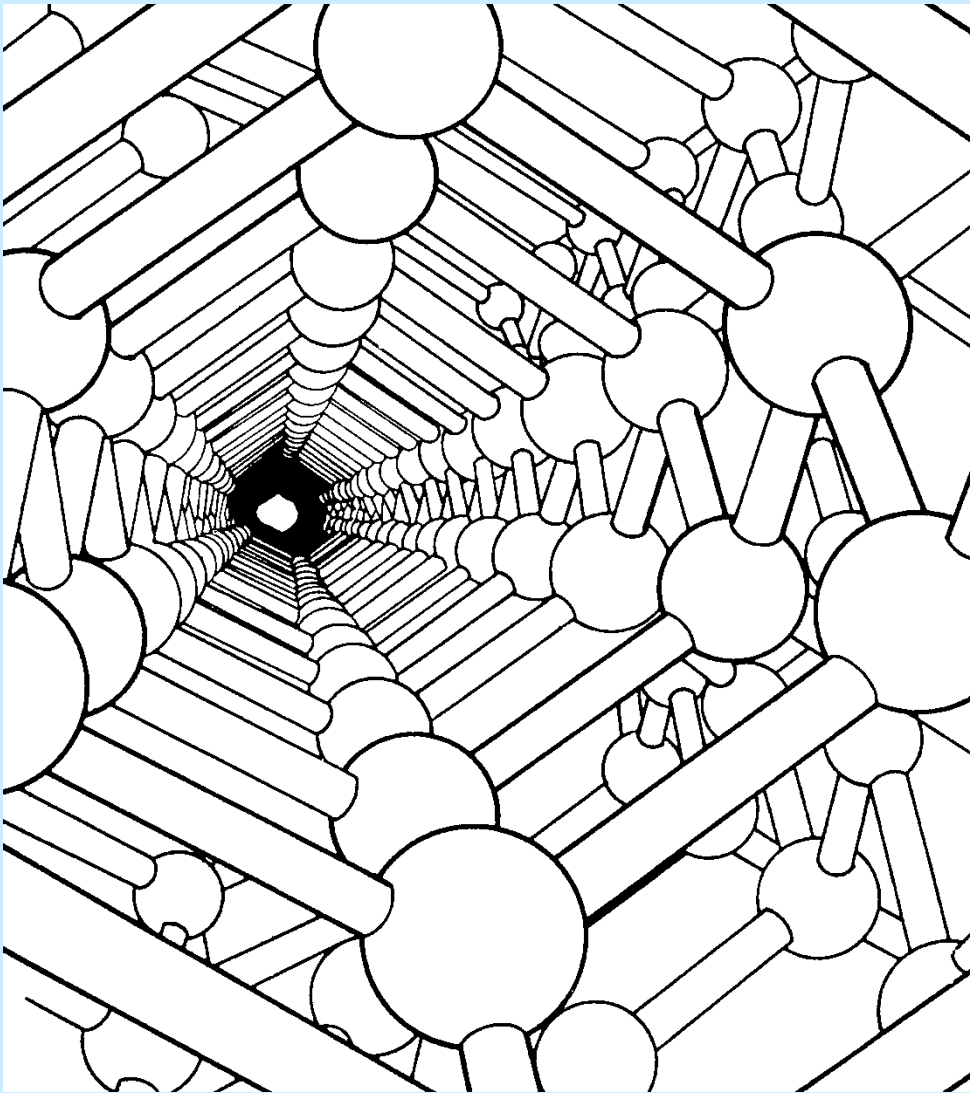
Channeling

For ions moving in certain directions in a crystalline material, there are long-range open spaces through which the ions travel without significant scattering

Ions can be steered down these open channels by glancing collisions with the atom rows or planes extending the final ion distribution deeper into the target


(a)

(b)

(c)
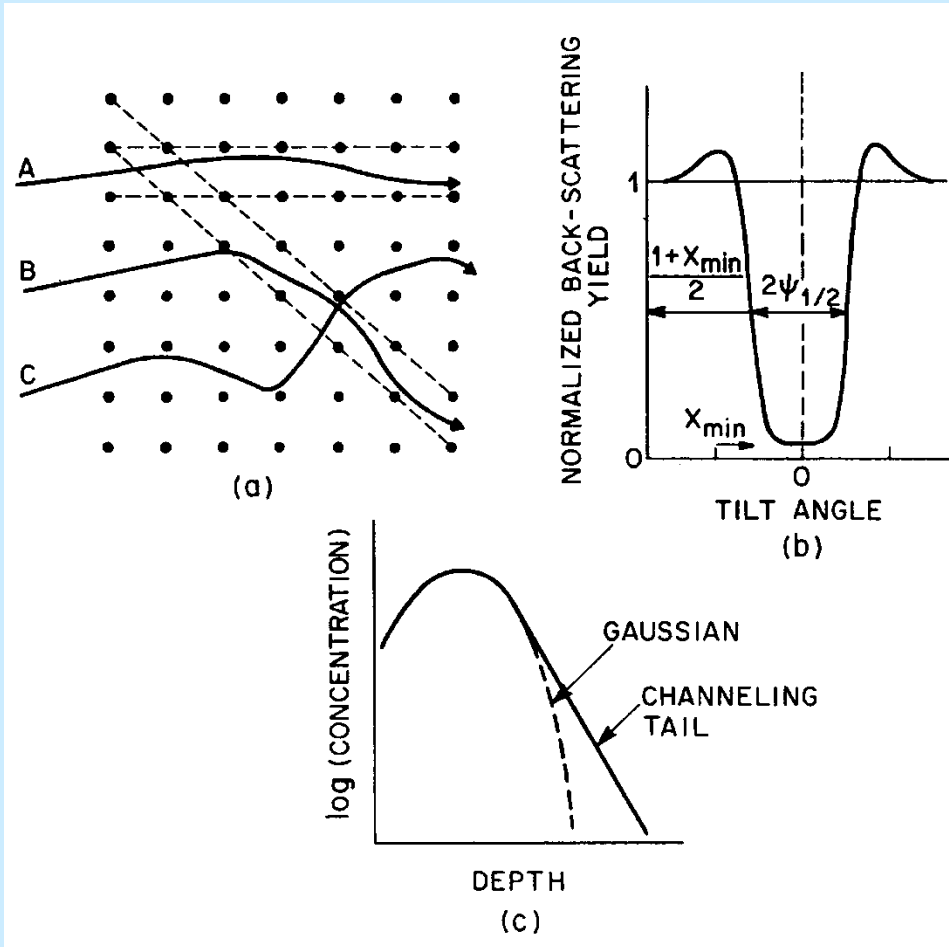
- Ion **A** is well aligned with a channel and so suffers only glancing collisions with the walls as it travels far into the lattice

- Ion **B** is scattered into a channel after a short distance

- Ion **C** is not channeled and undergoes random collisions with lattice atoms

- The angular width of a channel can be determined by measuring the number of ions backscattered from the lattice as a function of angle for a well collimated beam

- Channeling is characterized by a critical angle, Ψ_1, which is the maximum angle between ion and channel for a glancing collision to occur

  $$\Psi_1 = 9.73 \, (Z_1 Z_2 / Ed)^{1/2} \quad \text{(degrees)}$$

 where E is the ion energy (keV) and d is the atomic spacing (A)

- Channeling is more likely for heavier ions (higher Z) and lower energy

- To preclude haphazard channeling, wafers are usually tilted by 7° to avoid the major planes
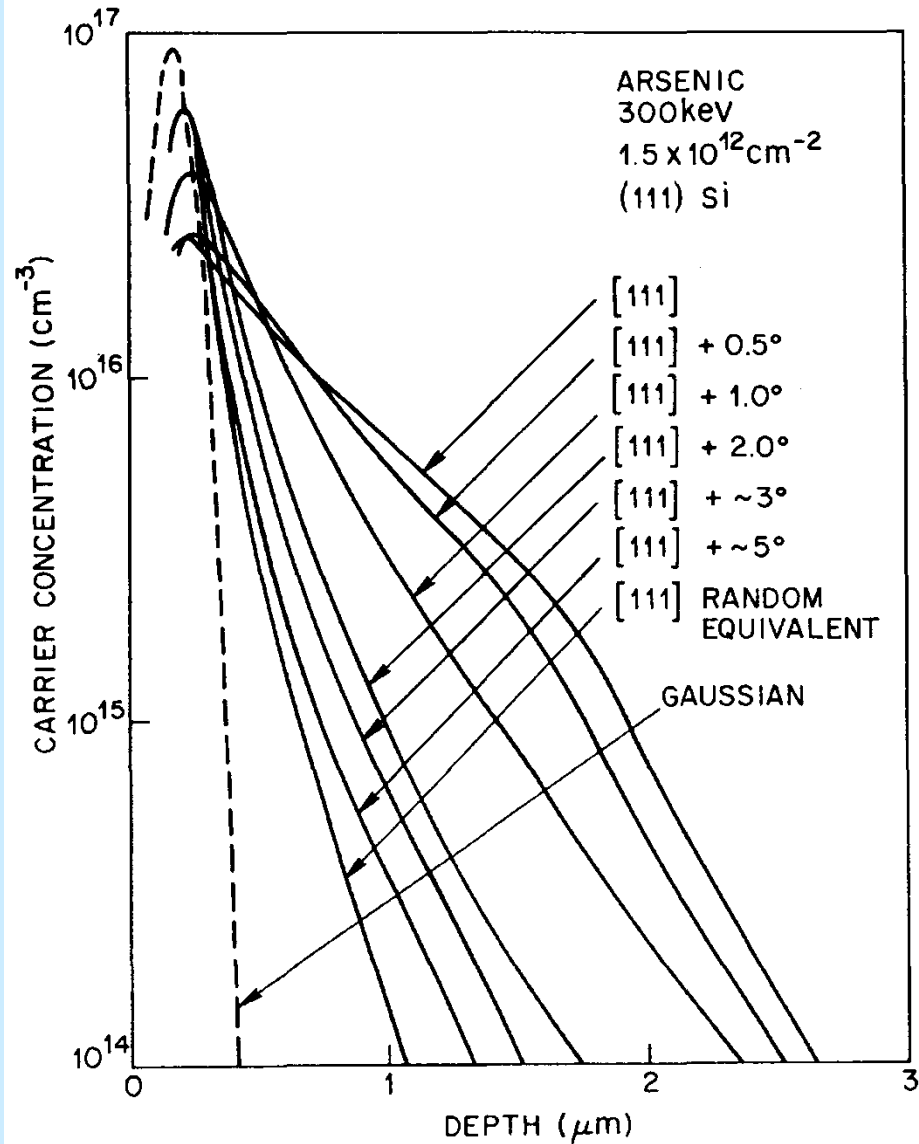
Channeling is not the preferred method to place ions deep inside the target while minimizing lattice damage because:

- the channeled atom profile is sensitive to changes on the order of 1° in wafer tilt and beam divergence
- ions are scattered by amorphous surface films and residual damage from previous processing steps

Recoils
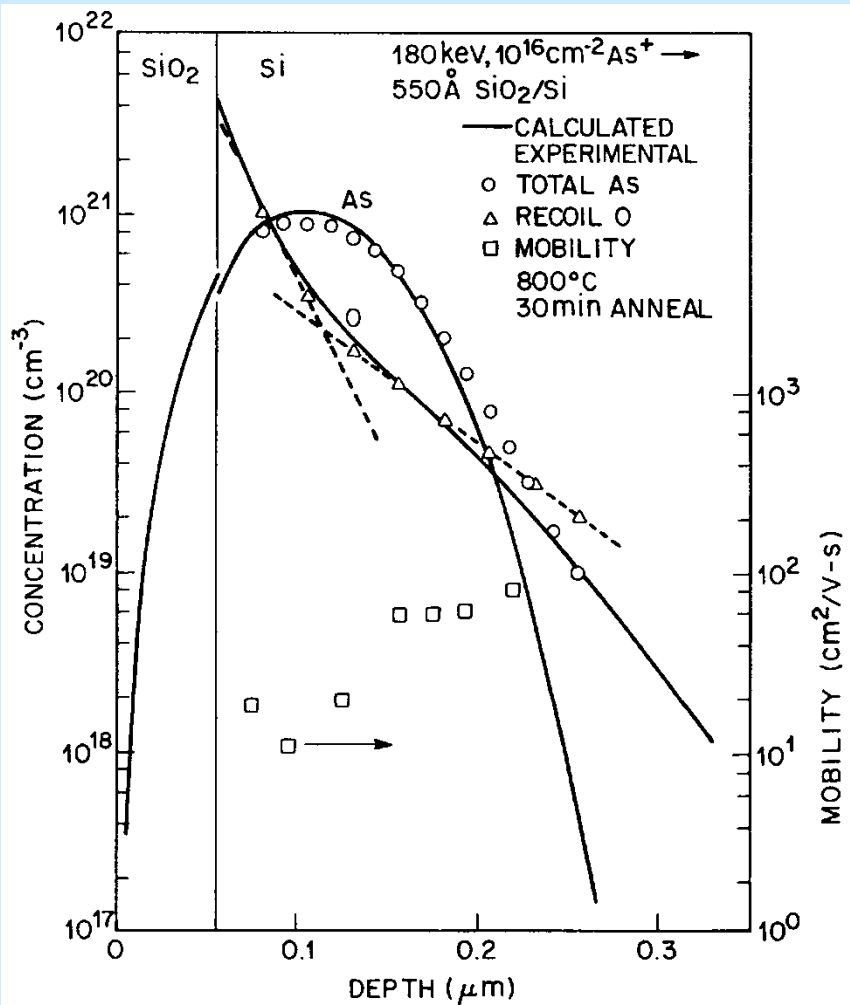
- Atoms are displaced from one layer into the underlying layer during the implantation process

- Oxygen atoms dislodged from the oxide matrix and recoiled into silicon give a profile of two roughly exponential regions

- Close to the interface, the displacement cascades push away very-low-energy oxygen atoms just across the interface.

- Beyond this, primary oxygen recoils give rise to a deeper exponential tail.

- This high interfacial oxygen concentration can degrade carrier mobility and introduce deep level traps

- Recoil mixing can be utilized positively to introduce dopant atoms that cannot be readily made into a source for implantation machines but can be deposited as a thin film on a silicon substrate

- Implanting silicon through the film creates a dopant tail in the silicon

- Self-implantation can break up thin contaminant layers under deposited films prior to further processing, such as the formation of a metal silicide

Ion Implanters
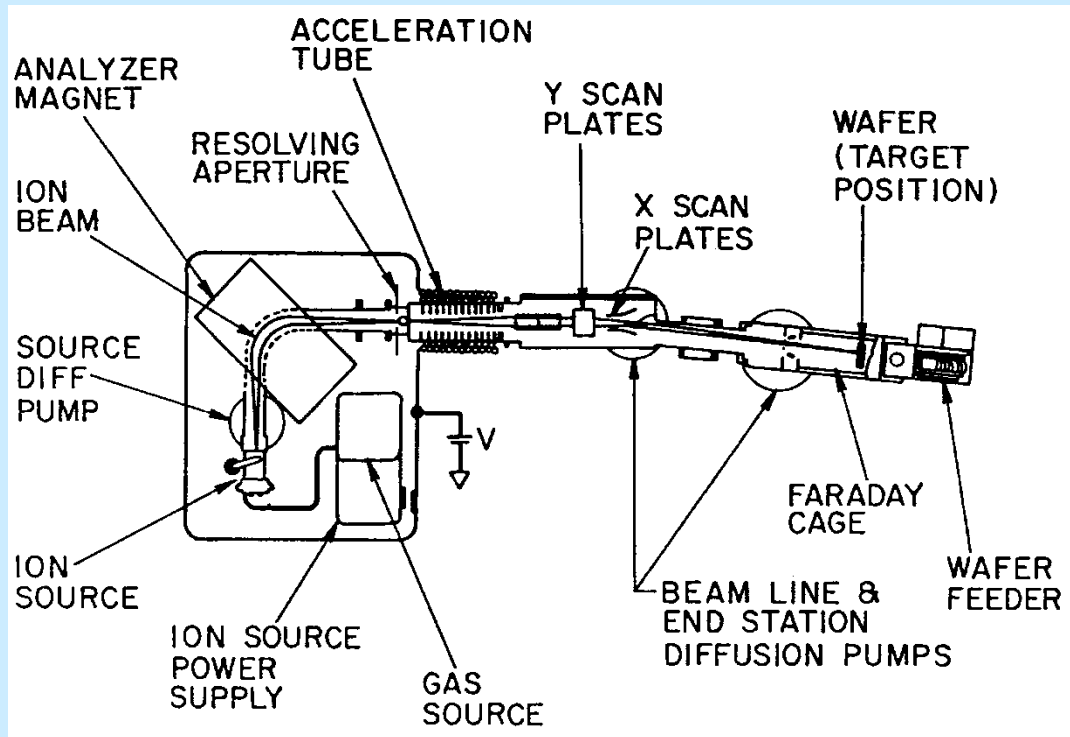
Secondary electrons emitted from the target are usually recaptured by using a small positive bias around the Faraday cup cage

The total number of ions entering the target is called the dose or fluence. If the current in the ion beam is *I*, then for a beam swept over an area *A*, the dose *Φ* is given by $\phi = \dfrac{1}{QA}\int I dt$

where *Q* is the charge on the ion and *t* is the implantation time

Implant Uniformity

Sheet resistance maps of three samples implanted with 50 keV phosphorus as a function of dose: (**a**) 5 x 10^{13} cm^{-2}, (**b**) 2 x 10^{14} cm^{-2}, (**c**) 2 x 10^{15} cm^{-2}

(a)   (b)   (c)
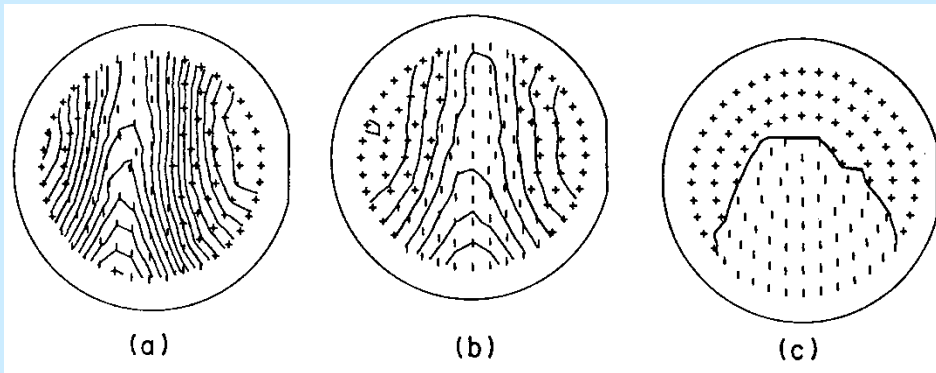
- The contour lines demarcate 1% changes in sheet resistivity

- Channeling is observed in sample (a) as the beam angle varies when being swept across the wafer surface

- At higher doses (b and c), channeling is reduced because of the larger amount of incurred crystal damage

Contamination

Although beam analysis separates contaminant ions from the ion beam before impacting the target, there are other sources of contamination near the end of the beam line:

- adventitious metal atoms knocked from chamber walls
- wafer holders
- clips
- masking apertures
- various hydrocarbons from pump oil
- cross contamination due to sputtering of old source atoms adhering on the walls
- foreign particulates dropping onto the wafer during sample loading, unloading, and transportation

Annealing

- After ion implantation, the wafer is usually so severely damaged that the electrical behavior is dominated by deep-level electron and hole traps that capture carriers and make the resistivity high

- Annealing is required to repair lattice damage and put dopant atoms on substitutional sites

- The success of annealing is often assessed by the fraction of dopant that is electrically active, as found experimentally using the Hall effect technique

- The Hall effect measures an average effective doping level, which is an integral over local doping densities and local mobilities evaluated per unit surface area

$$N_{Hall} = \frac{\left(\int_0^{x_j} \mu n\, dx\right)^2}{\int_0^{x_j} \mu^2 n\, dx}$$

  where μ denotes the mobility, n is the number of carriers, and x_j is the junction depth

- If the mobility is not a strong function of depth, N_{Hall} measures the total number of electrically active dopant atoms

- If annealing activates all of the implanted atoms, this value should be equivalent to the dose, Φ

Furnace Annealing

- The annealing characteristics depend on the dopant type and dose

- There is a clear division between cases where the silicon has been amorphized and where it has been merely partially disordered

- For amorphized silicon, regrowth proceeds via solid phase epitaxy (SPE), in which the amorphous / crystalline interface migrates toward the surface at a fixed velocity that depends on temperature, doping, and crystal orientation
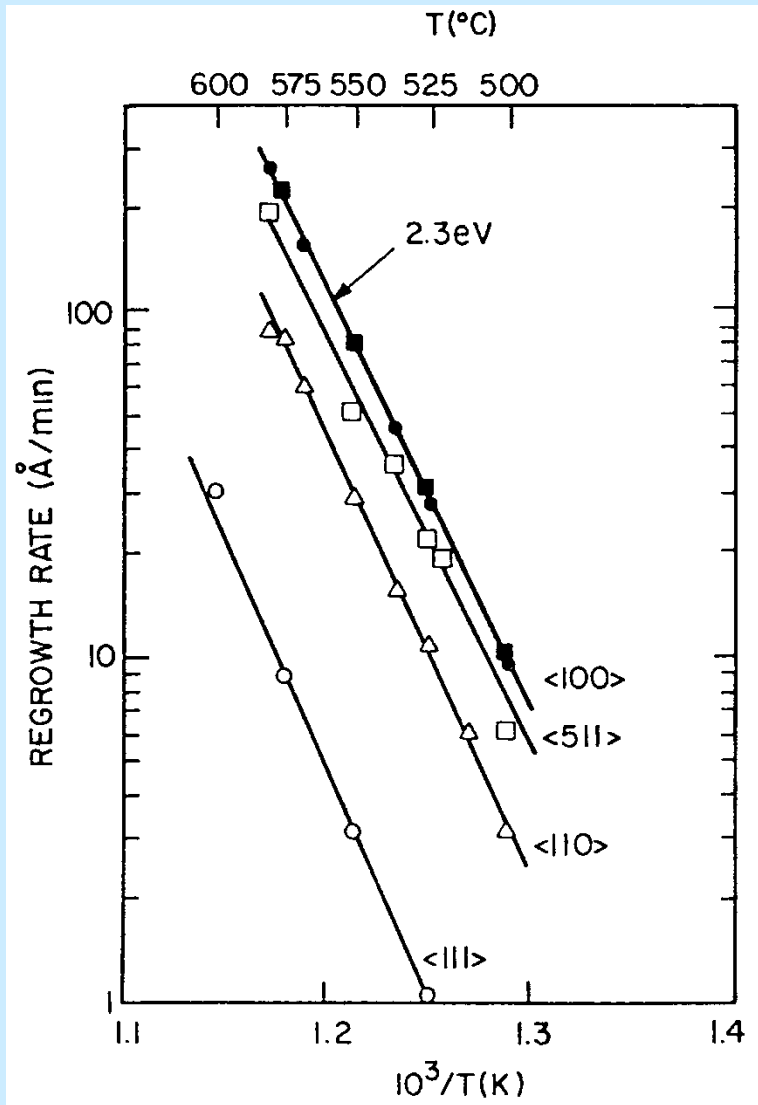
- The activation energy for SPE is 2.3 eV, implying that the process involves bond breaking at the interface
- The presence of impurities such as O, C, N, and Ar impedes the regrowth process, as these impurities bind to broken silicon bonds
- Dopants such as B, P, and As increase the regrowth rate (by a factor of 10 for concentrations in the regime of 10^{20} atoms/cm^3), because substitutional impurities weaken bonds and increase the likelihood of broken bonds
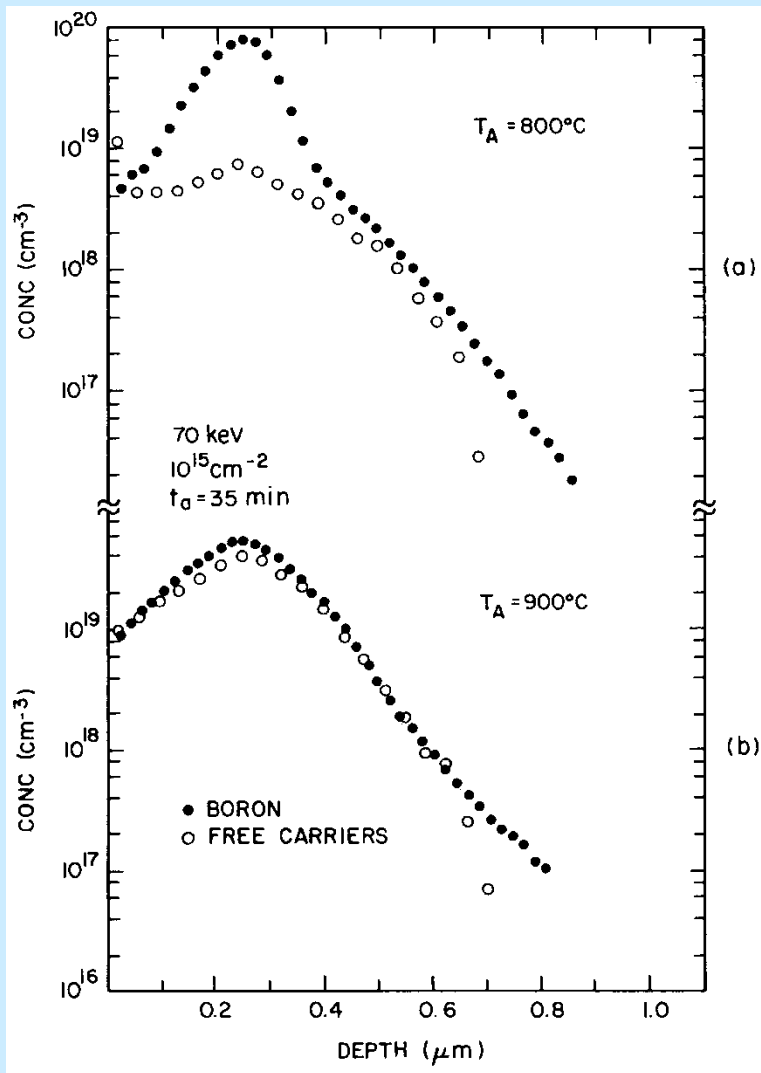
- If the implantation conditions are not sufficient to create an amorphous layer, lattice repair occurs by the generation and diffusion of point defects

- This process has an activation energy of about 5 eV and requires temperatures on the order of 900°C to remove all the defects

- It is thus easier in many cases to repair a fully amorphized layer than a partially damaged one

- The result of incomplete annealing is a reduction in the fraction of active dopant

Rapid Thermal Annealing

- Lattice damage repair has an activation energy of 5 eV compared to dopant diffusion with an activation energy of 3 to 4 eV

- At sufficiently high temperature, repair is faster than diffusion

- Furnace annealing is capable of supplying high temperature but the practical steps required to insert and remove wafers without stressing them lead to a minimum furnace annealing time of about 15 minutes. This is much longer than required to repair damages giving rise to unnecessary dopant diffusion

- Rapid thermal annealing (RTA), which allows repair of damage with minimal diffusion, is a generic term incorporating several methods of heating wafers for periods of time from 100 seconds down to nanoseconds
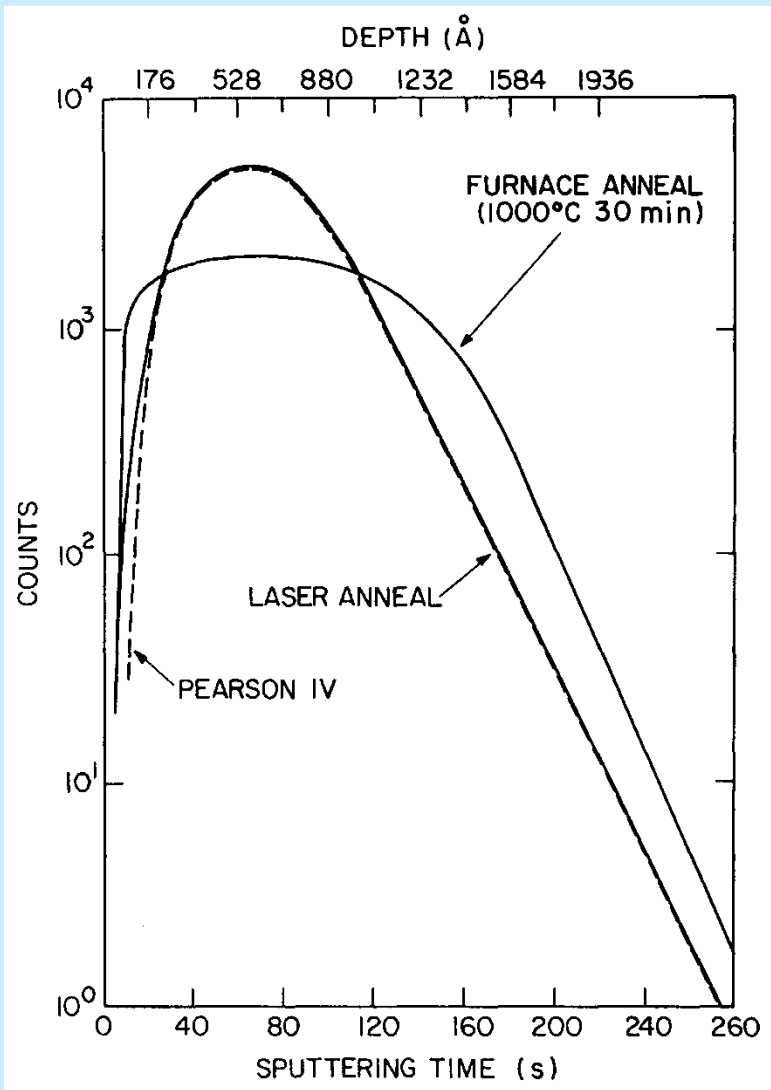
Isothermal RTA System
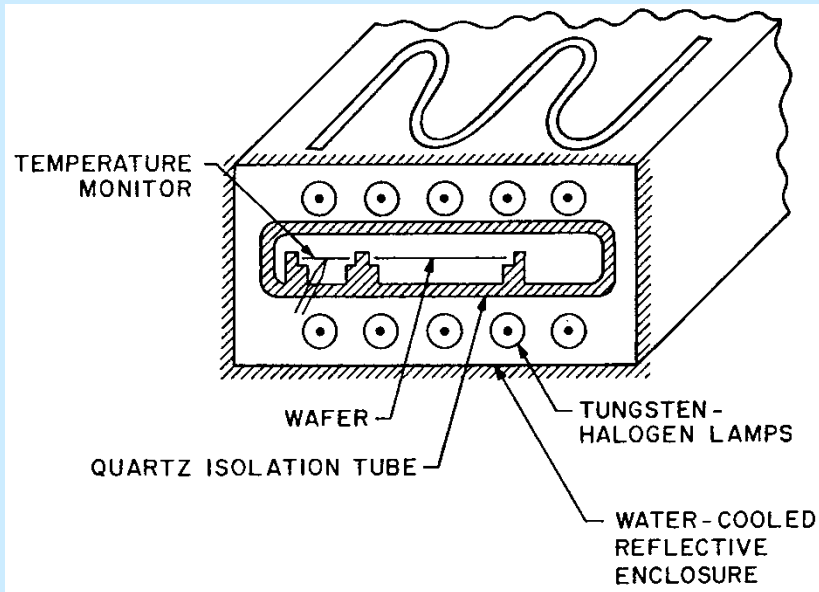
Tungsten-halogen lamps heating the wafer from one or both sides

Furnace RTA System
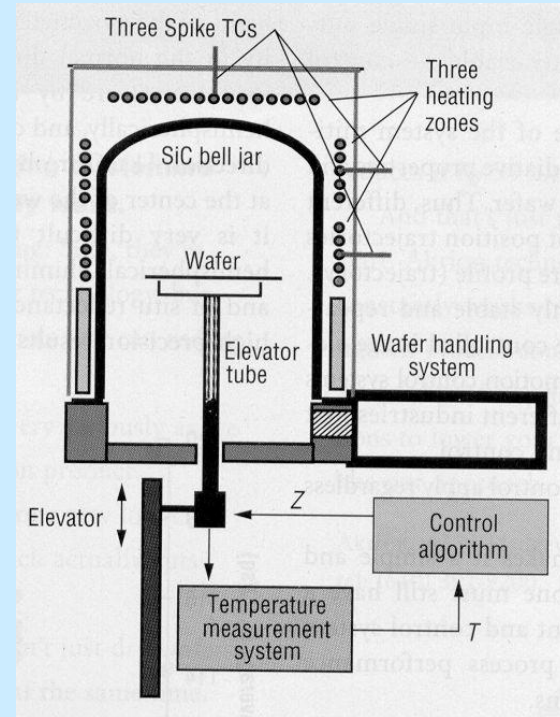
A thermal gradient is established by adjusting the power supplied to different zones of the bell jar, with the hottest zone on top, and the sample is moved into and out of the heated zone to achieve RTA

- Both RTA schemes offer significant advantages for VLSI processing because good activation can be achieved with insignificant dopant diffusion

- The left figure depicts an arsenic implant after RTA and almost complete electrical activation is accomplished without much diffusion

- In order to produce ultra-shallow junctions, spike RTA that exposes the wafer to a high temperature for a fraction of a second is employed

City University of Hong Kong

Shallow Junction Formation

- For ULSI CMOS devices, shallow N+ and P+ layers are needed for the source and drain regions

- In scaling horizontally to sub-micrometer and deep sub-micrometer dimensions, dopant profiles must also be scaled vertically

- Junction depths on the order of tens of nanometers or less are required for deep sub-micrometer devices

Low Energy Implantation

- Arsenic is heavy enough to form shallow N+ layers with implantation energies permitted by commercial low energy ion implantation machines

- Shallower boron can be achieved by implanting the BF_2^+ molecular ion. The effective implantation energy is reduced by a factor of 11/49 because upon impact, BF_2^+ dissociates into atomic boron and fluorine

- The extra fluorine atoms increase the lattice damage, thus minimizing channeling and facilitating annealing

- New advances have enabled modern low energy ion implanters to deliver reasonably high ion current at energy below 1 keV

- For ultra-shallow junction formation, laser doping and plasma doping are alternative techniques, but low energy beam-line ion implantation continues to be the main stream technology

- A profile can be moved closer to the surface by implanting through a surface film such as silicon dioxide

- This shifts the profile by roughly the oxide thickness, but recoiled oxygen atoms can be problematic

- This recoil effect can be utilized by knocking dopant out of a deposited surface film using silicon or inert gas implantation, a process termed ion mixing

Tilted Ion Beam
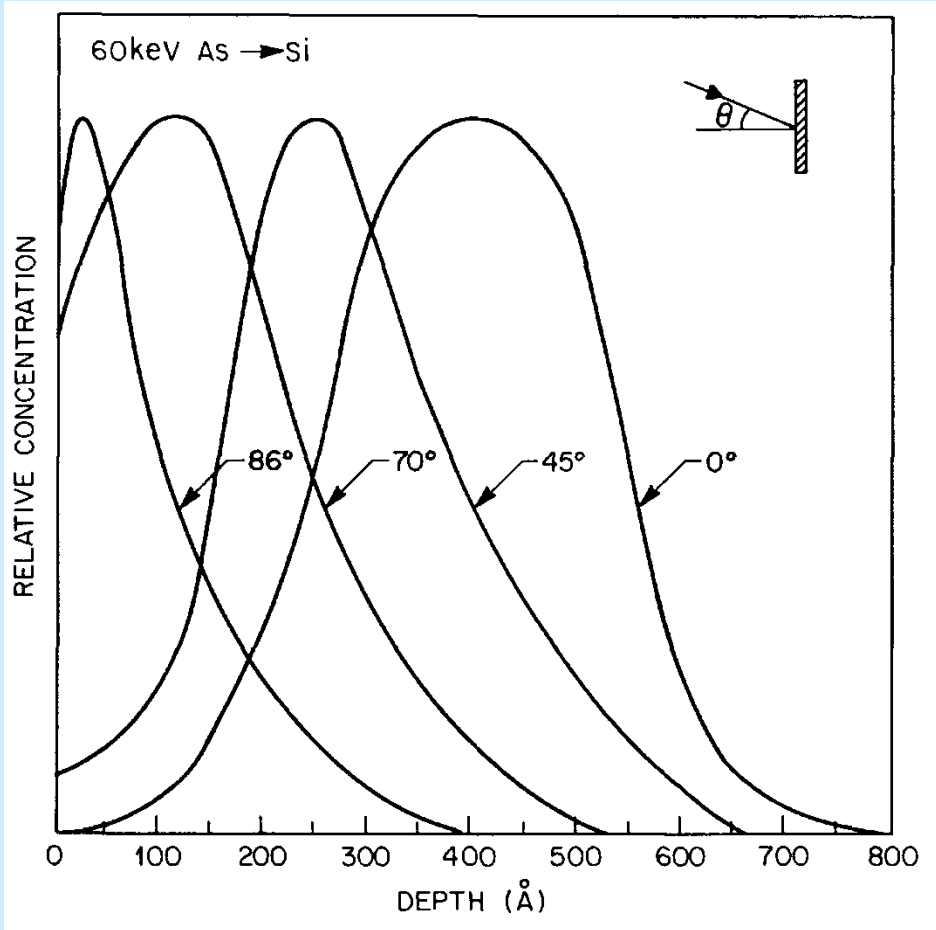
- If the wafer is tilted at a large angle relative to the ion beam, the vertical projected range can be decreased

- However, for large tilt angles, a significant fraction of the implanted ions is scattered out of the surface, so the effective dose is reduced

- This technique is only useful when the wafer surface is not patterned because large tilt angles cause long shadows and asymmetries at mask edges

Implanted Silicides / Polysilicon

- The problem of forming a shallow layer can be circumvented if a surface layer is deposited and the dopant is subsequently diffused into the substrate from the surface layer

- This is most often done when the surface film is to be used as an ohmic contact to the substrate

- Carefully controlled diffusion can result in steep dopant profiles without damaging the silicon lattice
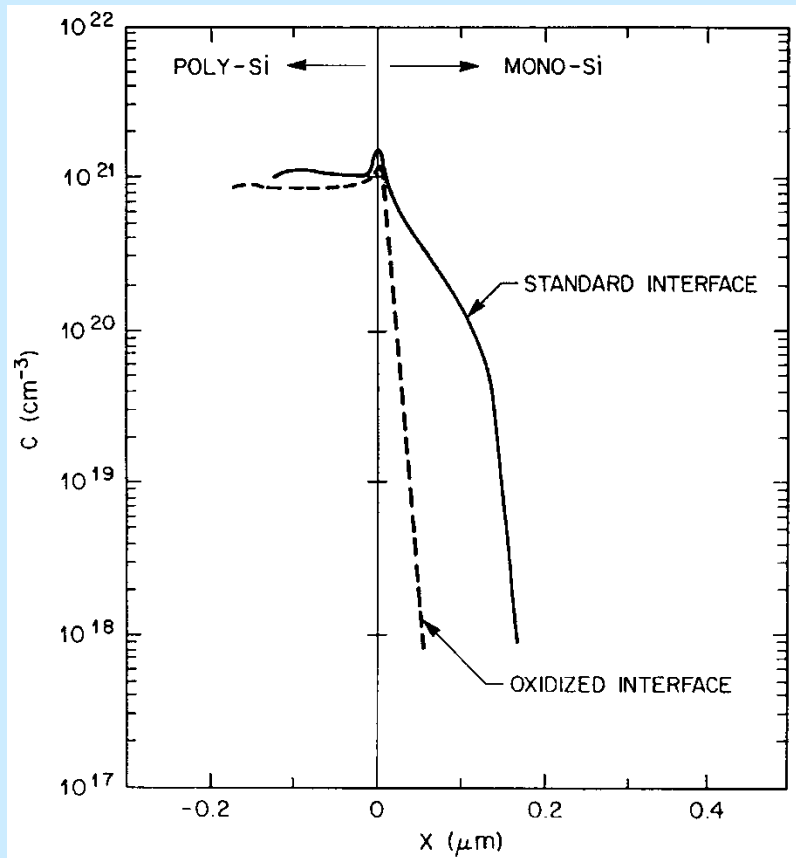
The presence of a 2.5 nm thick oxide layer between the polysilicon and the substrate is sufficient to block most of the diffusion

- Dopant diffusion in polysilicon (or silicides) is generally much faster than in single-crystal silicon, and so the implanted atoms soon become uniformly distributed in the surface thin film

- Some of the dopant atoms diffuse into the substrate, consequently yielding a fairly abrupt profile

- The small peak at the interface may be due to grain boundary segregation or impurities trapped at the interface

High Energy Implantation

Implantation at MeV energy is frequently used in VLSI fabrication. The most common application is to form deep isolation regions among individual devices. It is also used to form tubs in CMOS structures
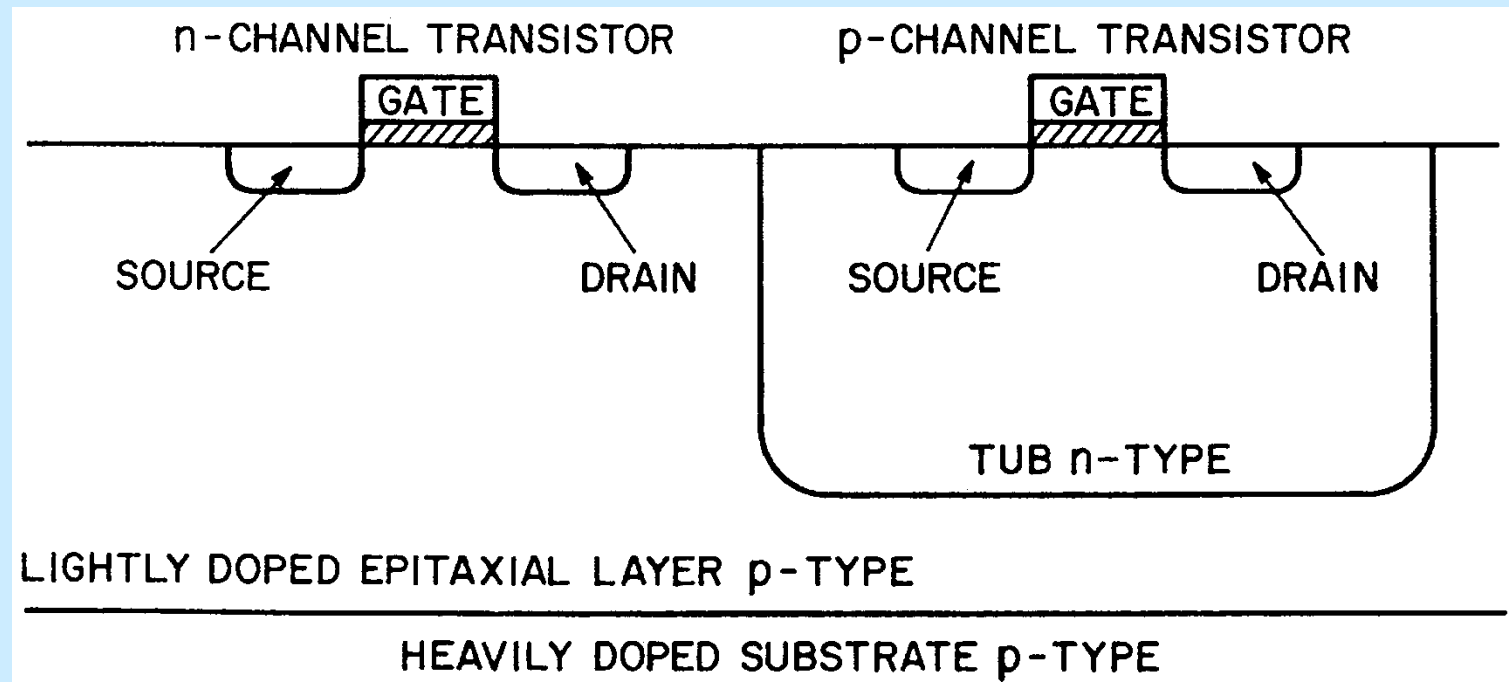
- High-energy ion implantation offers three ways in which the traditional epitaxial CMOS process can be improved

- The tubs can be implanted rather than diffused from the surface. In order to achieve a roughly uniform doped layer, a series of implants at different energies are required, accompanied by a short annealing step

- The structure can be improved further by retrograde doping of the tub, i.e., varying the implant doses such that the tub surface is less doped than the tub bottom. This method has the same advantages for the tub transistors as the use of epi-substrates has for the transistors outside the tub

- The epi-substrate can be dispensed altogether by using a blanket high-energy implant in the first processing step. This allows the formation of a buried, heavily-doped layer serving the same function as a heavily-doped substrate

Buried Insulator

- Devices can be fabricated in a thin silicon layer on an insulating substrate using two types of materials: SOS (silicon-on-sapphire) and SOI (silicon-on-insulator)

- Both technologies have the advantage of increasing radiation hardness due to the reduced collection volume for charges generated by ionizing radiation

- They offer a compact way to isolate devices from each other to reduce parasitic capacitance and to eliminate latchup for CMOS circuits

SOS

- SOS technology is more mature

- High quality silicon epitaxial films can be grown on sapphire wafers

- Sapphire substrates are very expensive, thereby limiting the use of SOS to demanding applications, such as military devices

SIMOX

- The **SIMOX (Separation by Implantation of Oxygen)** process introduces a blanket buried oxide to isolate the device active region from the bulk wafer.

- The principle behind the formation of a buried oxide layer is quite simple. If oxygen is implanted at a dose on the order of 10^{18} atoms/cm^2, there will be twice as many oxygen atoms as silicon atoms at the vicinity of the peak. Upon annealing, buried silicon dioxide will form

- The details of high dose oxygen implantation differ from traditional implantation steps in several ways
- The goal is to maintain a surface layer of high-quality single-crystal silicon for device fabrication
- The substrate is kept near 600ºC during implantation so that self-annealing maintains the crystal integrity
- Each incident oxygen ion sputters on the order of 0.1 silicon atom, and so the large number of implanted oxygen ions erode many layers of silicon atoms
- This is more than offset by the expansion in volume during the formation of oxide (44%), resulting in slight swelling of the silicon surface

High Dose Oxygen Implantation
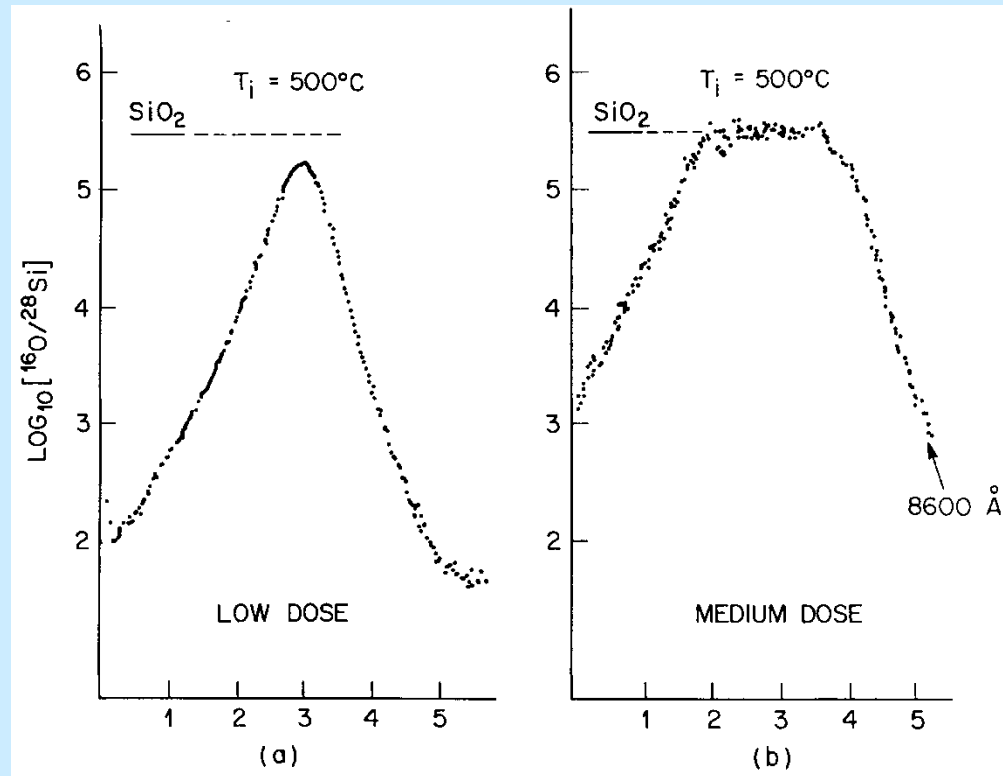
At high dose, the implant profile changes from a Gaussian-like distribution to a flat-topped distribution due to diffusion of oxygen to the silicon-oxide boundaries after oxygen saturates the substrate around the peak region

- The surface layer contains a substantial amount of oxygen and much damage, albeit still single-crystal

- Annealing is performed at a high temperature above 1300°C to cause a strong segregation of oxygen into the buried layer from both sides, consequently depleting the surface of almost all the implanted atoms, including impurities

- This leaves a high-quality surface film containing very little oxygen and less than 10^9 dislocations per cm^2

Wafer Bonding / Back Etching

- Using wafer bonding technology, two silicon wafers (one or both having a surface oxide layer) are pasted together resulting in a Si / SiO$_2$ / Si structure

- One side of the structure is thinned by polishing to yield the required silicon layer thickness

- The advantage is that the thickness of the oxide and silicon layers can be adjusted independently

- Two wafers are required to make one SOI wafer thereby raising the cost