Ch. 5: **p-n Junction**

Fabrication of P-N Junction

- Most semiconductor devices contain at least one junction between p-type and n-type material.
- These p-n junctions are fundamental to the performance of functions such as rectification, amplification, switching, and other operations in electronic circuits.
- We have already discussed in Chapter 1 how single-crystal substrates are grown, and how the doping can be performed in the whole manufactured semiconductor wafer.
- doping can be varied laterally across the surface, which is key to making integrated circuits on a wafer →it is necessary to be able to form patterned masks on the wafer corresponding to the p-n junctions circuitry, and introduce the dopants selectively through windows in the mask to make necessary functional device
- few unit process steps can be used in different permutations and combinations to make everything from simple diodes to the most complex microprocessors.

Fabrication of P-N Junction

Main steps in p-n fabrications are:

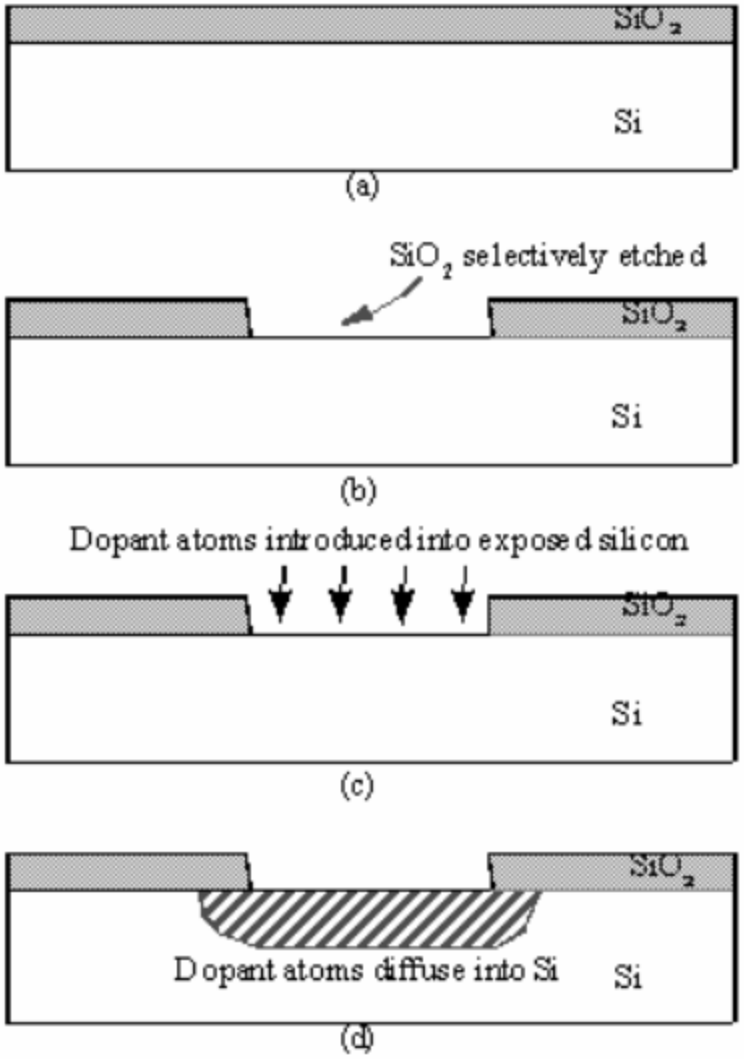
Oxidation

Lithography & Etching

Ion Implantation

Annealing & Diffusion

Thin Film Deposition



Fabrication of P-N Junction

Oxidation: Thermal Oxidation

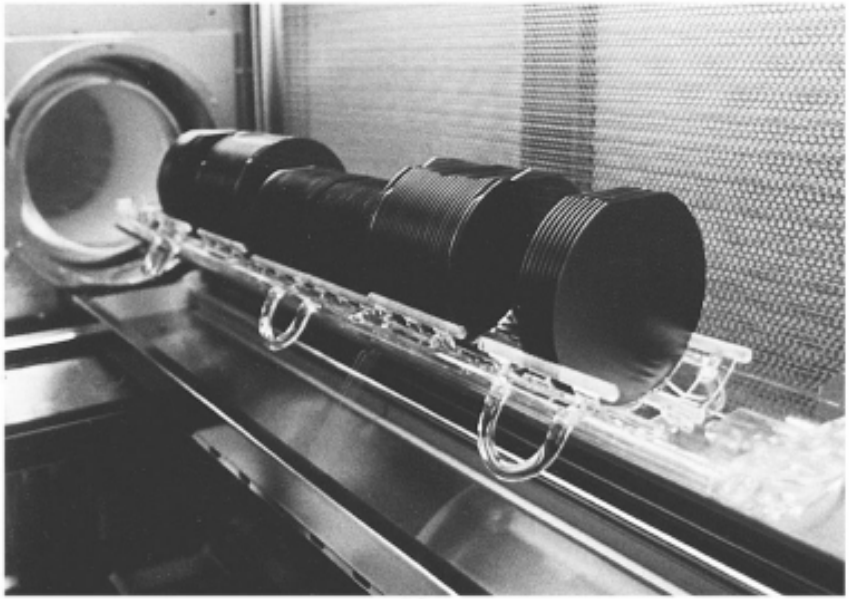
- This represents thermal oxidation of the Si surface to form insulating SiO_2 layer
- This involves placing a batch of wafers in a clean silica (quartz) tube which can be heated to very high temperatures (~800-1000°C) using heating coils in a furnace.
- An oxygen-containing gas such as dry O_2 or H_2O is flowed into the tube at atmospheric pressure to oxidize Si, and flowed out at the other end.
- Horizontal furnaces or vertical furnaces are usually used

Dry Oxidation : $Si + O_2 \rightarrow SiO_2$ Thin oxide

Wet Oxidation : $Si + 2H_2O \rightarrow SiO_2 + 2H_2$ Thick oxide

Fabrication of P-N Junction

Oxidation: Thermal Oxidation



(a)

(b)

(a) Silicon wafers being loaded into a furnace. For 8-inch and larger wafers, this type of horizontal loading is often replaced by a vertical furnace. (b) Vertical furnace for large Si wafers. The silica wafer holder is loaded with 8-inch Si wafers and moved into the furnace above for oxidation, diffusion, or deposition operations. (Photograph courtesy of Tokyo Electron Ltd.)

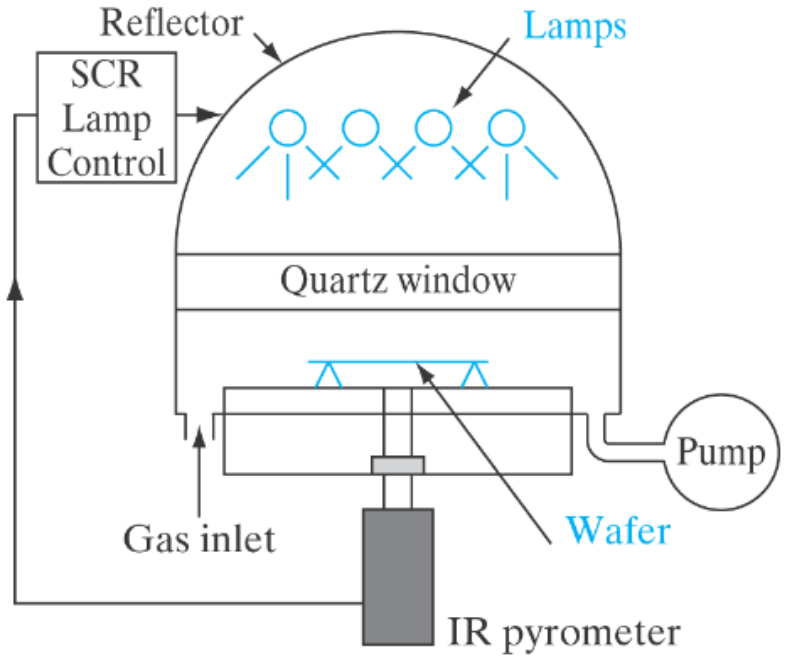
Fabrication of P-N Junction

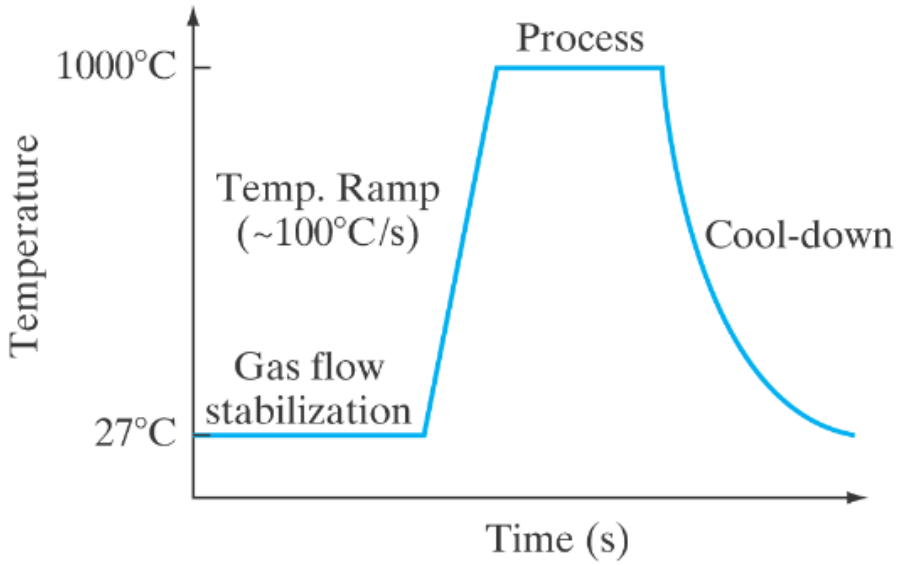
Oxidation: Rapid Thermal Process (RTP)

- Instead of having a large batch of wafers in a conventional furnace where the temperature cannot be changed rapidly, a single wafer is held on quartz pins, surrounded by a bank of high-intensity (tens of kW) tungsten-halogen infrared lamps, with gold-plated reflectors around them.
- By turning on the lamps, the high intensity infrared radiation shines through the quartz chamber and is absorbed by the wafer, causing its temperature to rise *very rapidly (~50-100°C/s).*
- *The processing temperature can be* reached quickly, after the gas flows (O_2 or mixture of gases contains oxygen) have been stabilized in the chamber
- This process takes few second for each wafer instead of hours in the furnace

Fabrication of P-N Junction

Oxidation: Rapid Thermal Process (RTP)





Schematic diagram of a rapid thermal processor and typical time–temperature profile.

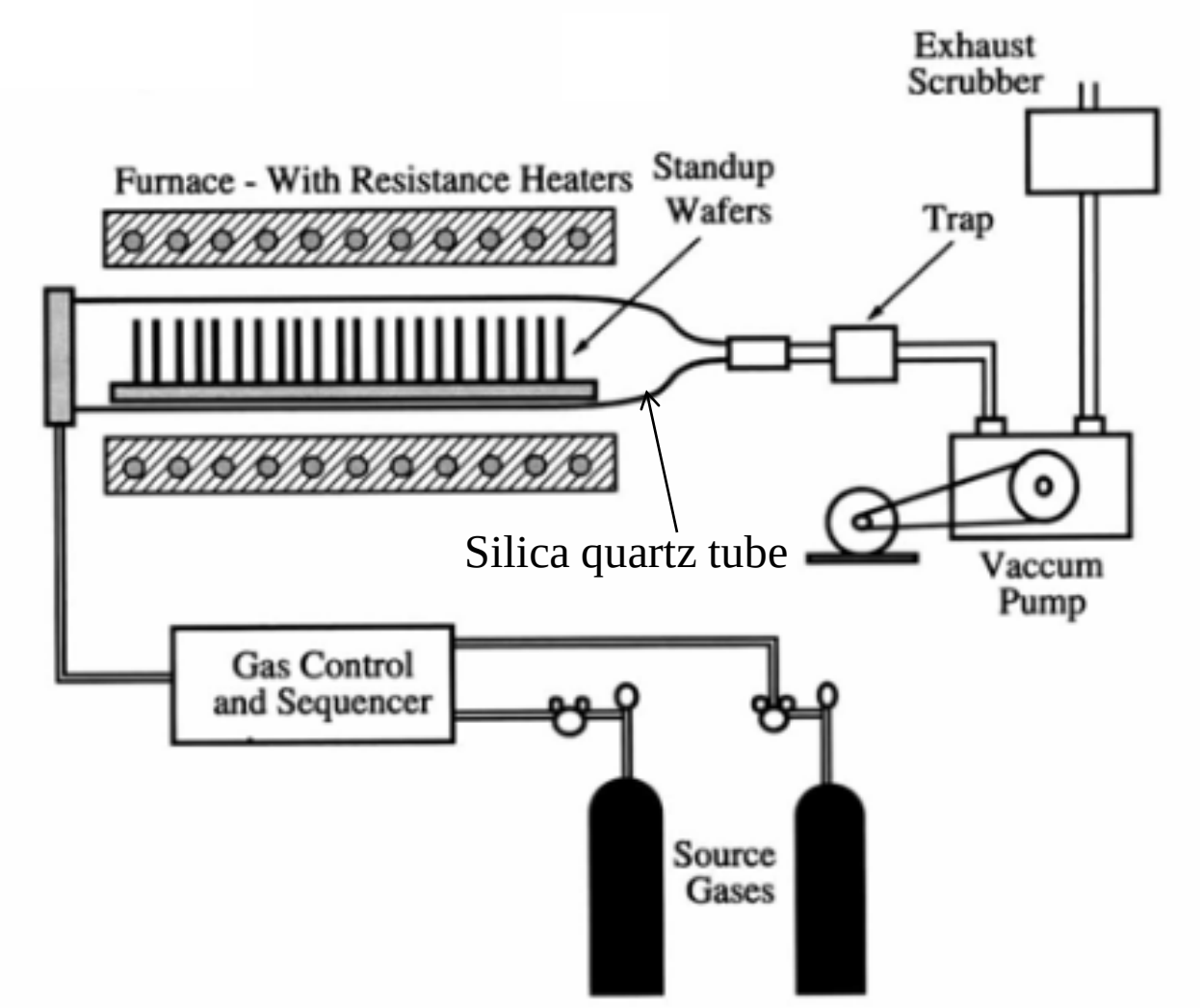
Fabrication of P-N Junction

Oxidation: Chemical Vapor Deposition (CVD)

- Thermal oxidation consumes Si from the substrate, and very high temperatures are required, whereas CVD of SiO_2 does not consume Si from the substrate and can be done at much lower temperatures.
- SiO_2 films can also be formed by *low pressure* (*~100* mTorr) chemical vapor deposition (LPCVD) or plasma-enhanced CVD (PECVD).
- The CVD process reacts a Si-containing gas such as SiH_4 with an oxygen-containing precursor, causing a chemical reaction, leading to the deposition of SiO_2 on the substrate.

Fabrication of P-N Junction

Oxidation: Chemical Vapor Deposition (CVD)



Low-pressure chemical vapor deposition (LPCVD) reactor.

Fabrication of P-N Junction

Diffusion

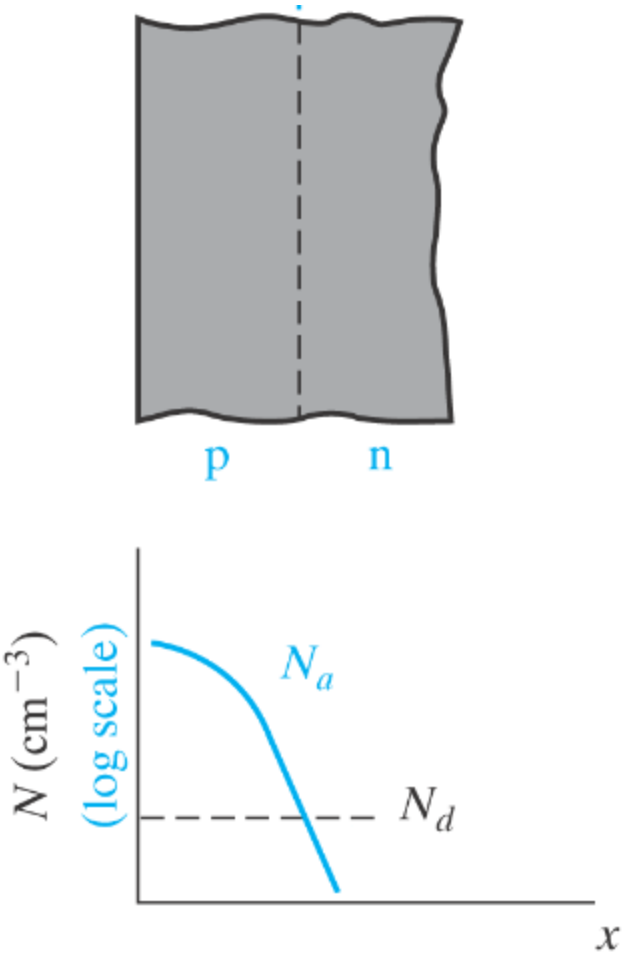
- Thermal in-diffusion of dopants was extensively used in the past in IC fabrication.
- The wafers are first oxidized and windows are opened in the oxide using the photolithography and etching steps (will be soon described).
- Dopants such as B, P, or As are introduced into these patterned wafers in a high temperature (-800-1100°C) diffusion furnace (like above shown furnaces), generally using a gas or vapor source.
- The dopants are gradually transported from the high concentration region near the surface into the substrate through diffusion.

Fabrication of P-N Junction

Diffusion

• The diffusivity of dopants in solids, *D, has a strong Arrhenius de*pendence on temperature, *T. It is given by* $D = D_0 \exp(-E_a/kT)$, *where* D_0 is a constant depending on the material and the dopant, and E_a is the activation energy.

• The average distance the dopants diffuse is related to the diffusion length $L = \sqrt{Dt}$, where t is the processing time



Impurity concentration profile for fabricating a p-n junction by diffusion.

Fabrication of P-N Junction

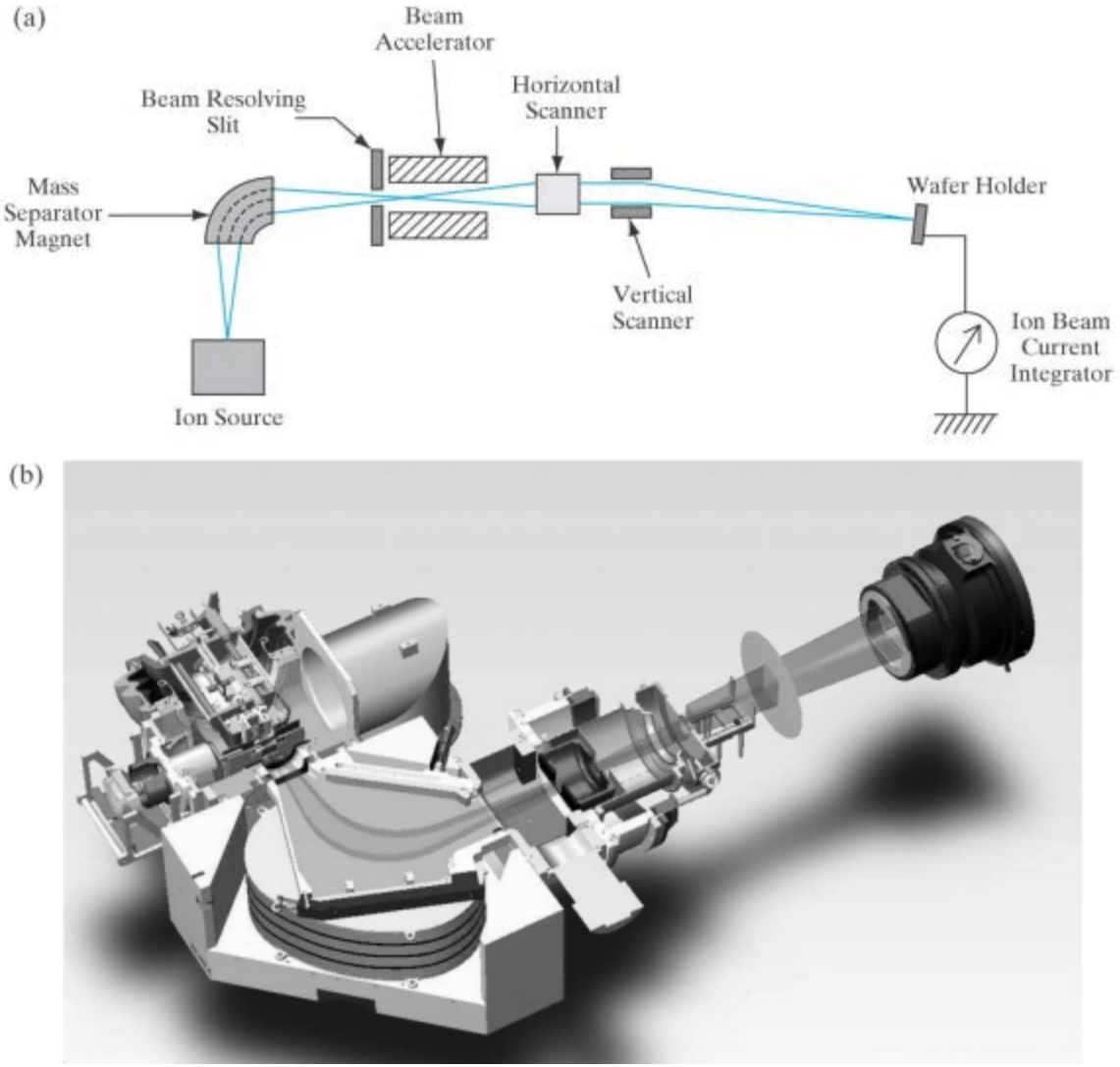
Ion Implantation

- A useful alternative to high-temperature diffusion is the direct implantation of energetic ions into the semiconductor.
- In this process a beam of impurity ions is accelerated to kinetic energies ranging from several keV to several MeV and is directed onto the surface of the semiconductor.
- As the impurity atoms enter the crystal, they give up their energy to the lattice in collisions and finally come to rest at some average penetration depth, called the *projected range.*
- *Depending on the impurity and its implantation energy, the* range in a given semiconductor may vary from a few hundred angstroms to about 1 μm

Fabrication of P-N Junction

Ion Implantation

- One of the major advantages of implantation is the precise control of doping concentration it provides. Since the ion beam current can be measured accurately during implantation, a precise quantity of impurity can be introduced



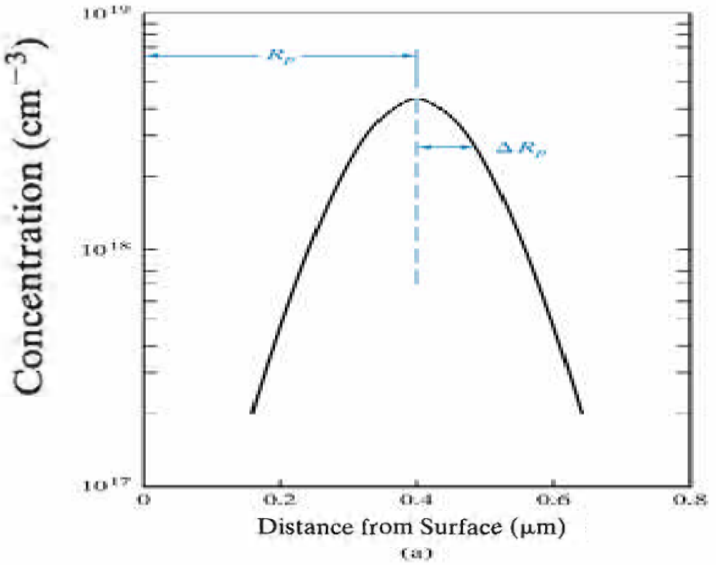
(a) Schematic diagram of an ion implantation system; (b) schematic of ion beam path through mass analyzer magnet. (Courtesy of Applied Materials.)

Fabrication of P-N Junction

Ion Implantation

$$N(x) = \frac{\phi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\frac{1}{2}\left(\frac{x - R_p}{\Delta R_p}\right)^2\right]$$

Φ is implementation dose (ions/cm^2)



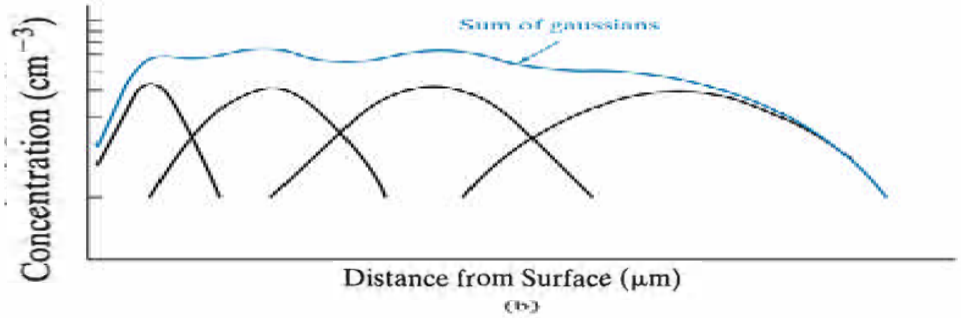


Figure 5—4
Distributions of implanted impurities: (a) gaussian distribution of boron atoms about a projected range R_p (in this example, a dose of 10^{14} B atoms/cm^2 implanted at 140 keV); (b) a relatively flat distribution obtained by summing four gaussians implanted at selected energies and doses.

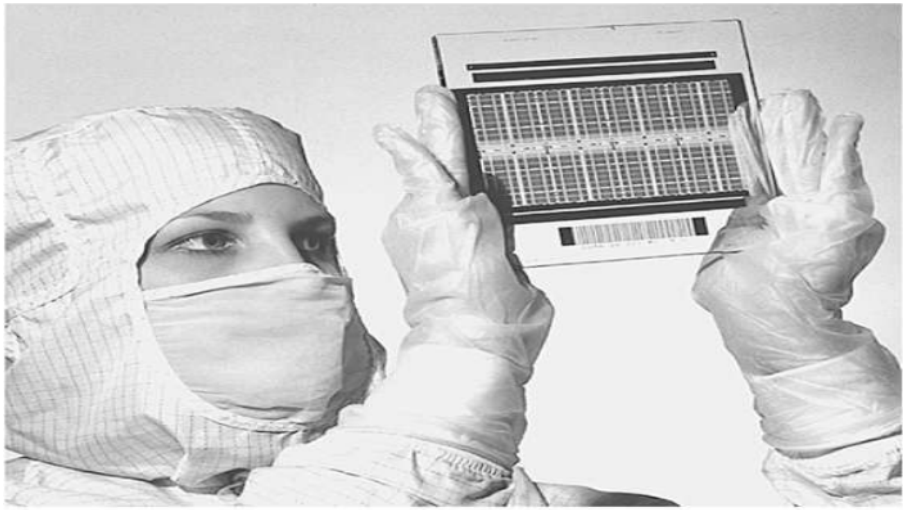
Fabrication of P-N Junction

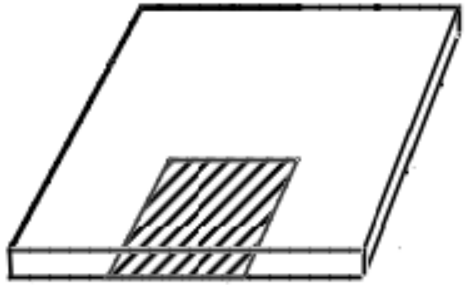
Photolythography: mask

• Patterns corresponding to complex circuitry are formed on a wafer using *photolithography.*

•*This involves first generating a reticle (or mask for the entire wafer) which is a transparent* silica (quartz) plate containing the pattern (Fig. 5-7a). Opaque regions on the mask are made up of an UV light-absorbing layer, such as iron oxide.

• Masks are usually done by computer controlled electron beam





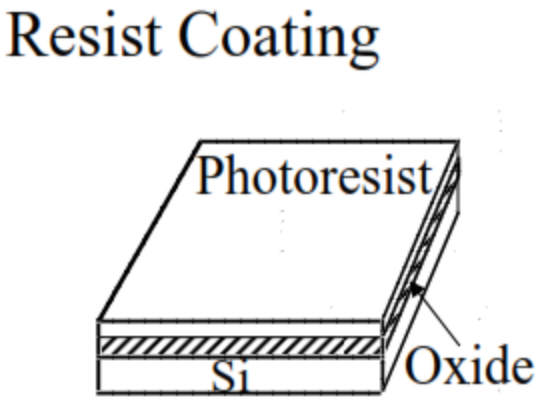
Simple mask

Figure 5—7a
A **photolithographic reticle** used for one step in the processing of a **16 Mb dynamic random access memory (DRAM).** In a "stepper" projection exposure system, ultraviolet light shines through the glass plate and the image **is projected onto** the wafer to **expose photoresist** for one die in the array of circuits, then steps to the next. (Photograph courtesy of IBM Corp.)

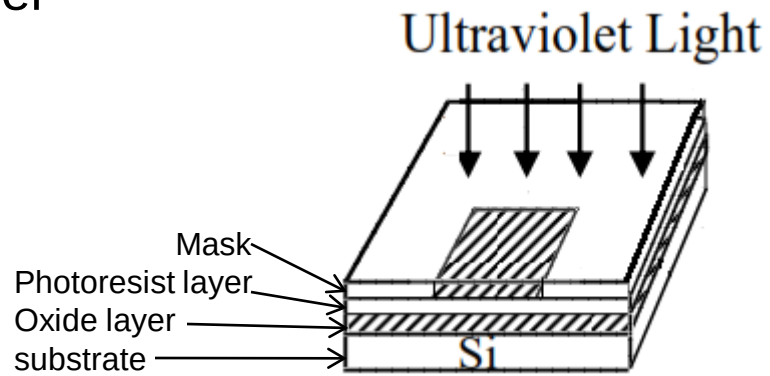
Fabrication of P-N Junction

Photolythography: photoresist

• Photosensitive thin organic polymer transparent layer (called photoresist layer) is added on the surface of the wafer before the mask is applied.
• Photoresist layer is used to define the mask pattern on the wafer by chemical changes occurs on it when it is exposed to energetic particles such as electrons or photons



•There are two types of photoresist; positive and negative
• Positive photoresist: material exposed to light or electrons is removed
• Negatiuve photoresist: Unexposed material is removed
• Mask is then applied on photoresist to transfer the mask pattern on wafer surface after exposure to UV light



Fabrication of P-N Junction

Photolythography: UV light system

The exposure of the wafers to UV light is achieved by a system called a *stepper (Fig. 5-7b)* , the ultraviolet light shines selectively through the mask onto a single die location (small area). After the photoexposure is done, the wafer mechanically translates on a precisely controlled *x-y translation stage to the next die location and is exposed* again

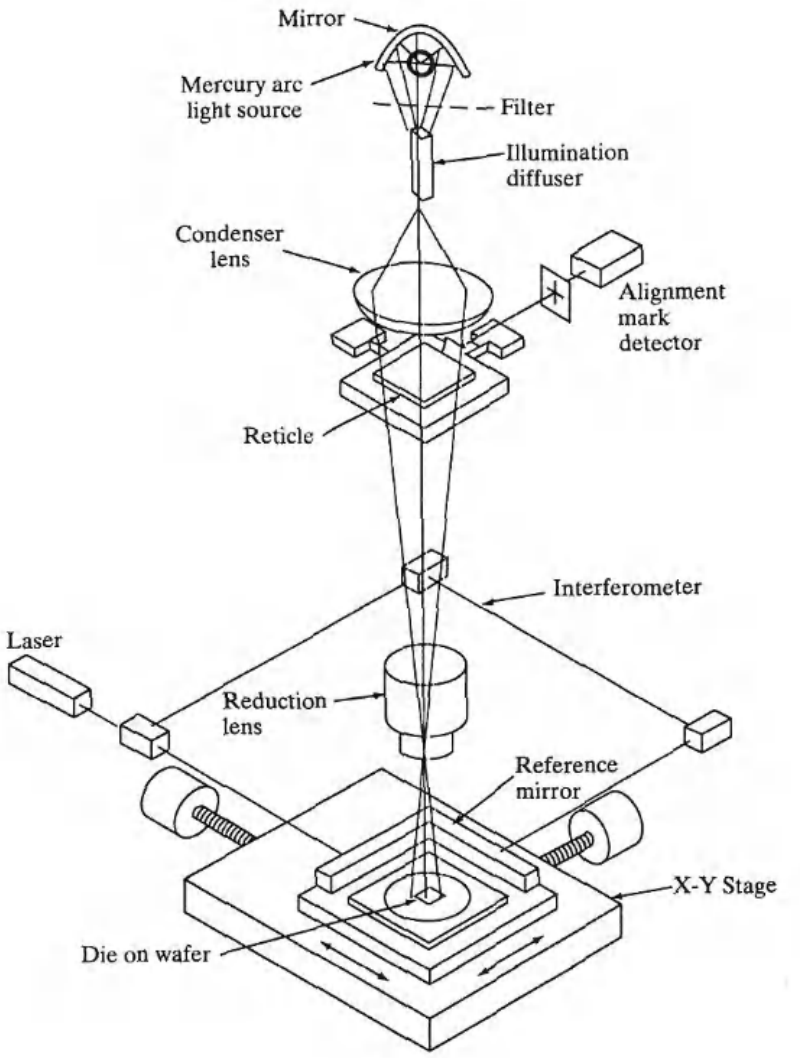
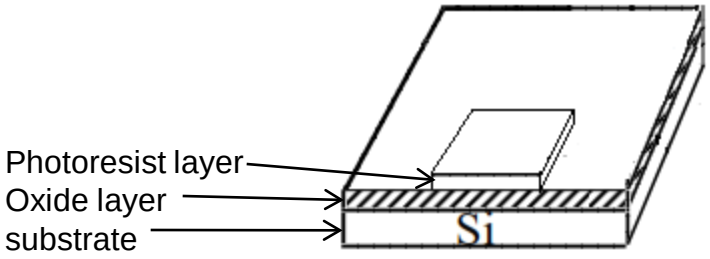


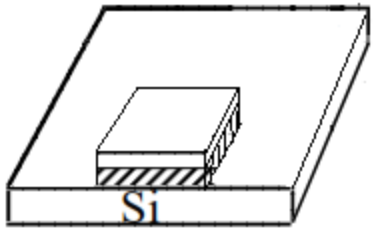
Figure 5—7b
Schematic diagram of **an optical stepper**

Fabrication of P-N Junction

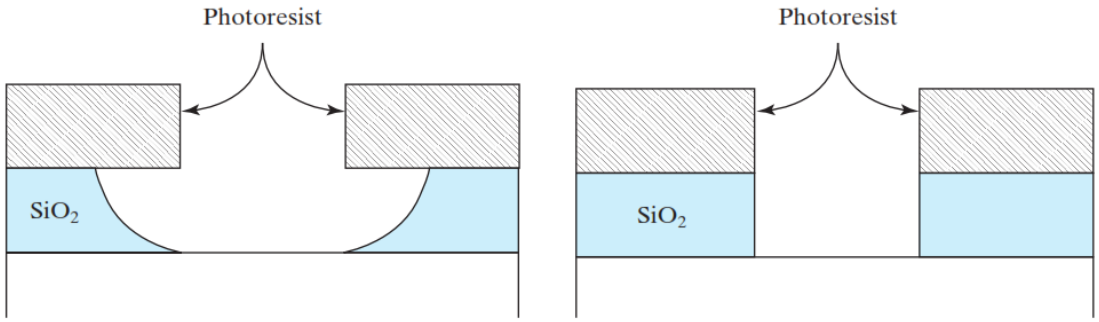
Etching

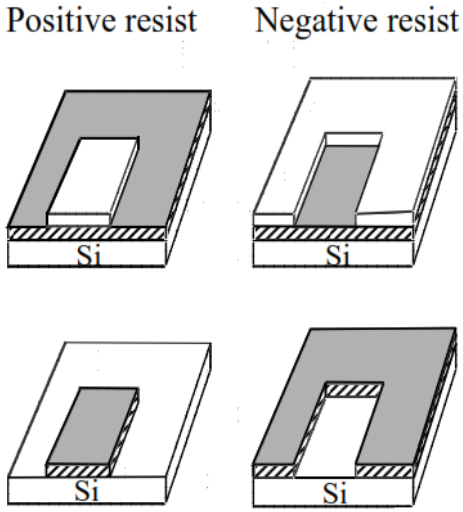
• The exposed photoresist layer (if +ve) is removed by NaOH solution → unexposed pattern remains on the wafer





• After the photoresist pattern is formed, it can be used as a mask to etch the unpatterned material (SiO_2 in this case).
• Etching was done using wet chemicals (dilute HF to etch SiO_2) . Note that HF attacks SiO_2, but does not affect the Si substrate or the photoresist mask.
• Wet chemical method is *isotropic, which means that they etch as fast laterally as they etch vertically* → unacceptable for ultra-small features. Hence, wet etching has been largely displaced





Etching and Resist Strip

Fabrication of P-N Junction

Dry Etching

• Dry, plasma-based etching which can be made both selective and *anisotropic (etches vertically but not laterally along the surface).*
• The most popular type of plasma based etching is known as *reactive ion etching* (RIE) (Fig. 5-8).
•Appropriate etch gases such as chlorofluorocarbons (CFCs) flow into the chamber at reduced pressure (~1-100 mTorr), and a plasma is struck by applying an rf voltage across a cathode and an anode.
• The rf voltage accelerates the light electrons in the system to much higher kinetic energies (~10 eV) than the heavier ions. The high energy electrons collide with neutral atoms and molecules to create ions and molecular fragments called radicals.
• The wafers are held on the rf powered cathode, while the grounded chamber walls act as the anode.

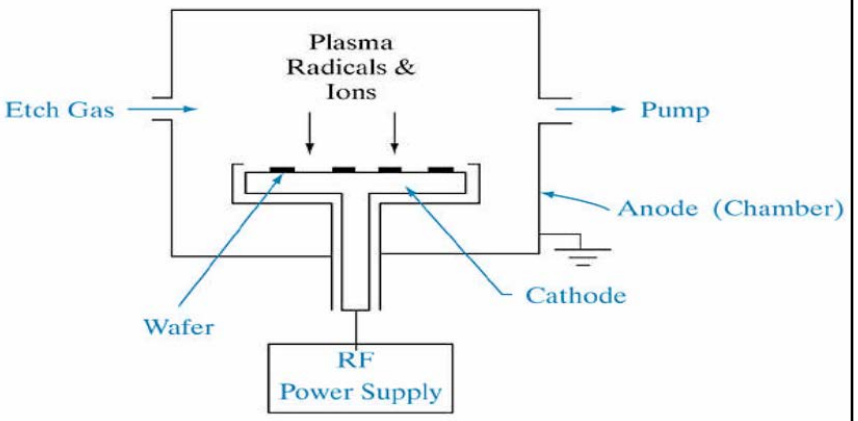


Figure 5—8
Reactive ion etcher (RIE). Single or multiple wafers are placed on the rf powered cathode to maximize the ion bombardment. *The most* commonly used rf frequency is 13.56 MHz, which is a frequency dedicated to industrial use so that there is minimal interference with radio communications.

• positive ions gain kinetic energy by being accelerated and bombard the wafer normal to the surface. This bombardment at normal incidence contributes a physical component to the etch that makes it anisotropic

Fabrication of P-N Junction

Metallization

• After the semiconductor devices are made by the processing methods described previously, they have to be connected to each other, and ultimately to the IC package, by metallization.

• Metal films are generally deposited by a physical vapor deposition technique such as evaporation (e.g., Au on GaAs) or sputtering (e.g., Al on Si).

• Sputtering of Al is achieved by immersing an Al target in an Ar plasma. Argon ions bombard the Al and physically dislodge Al atoms by momentum transfer (Fig. 5-9). Many of the Al atoms ejected from the target deposit on the Si wafers.

• The Al is then patterned using the metallization reticle and subsequently etched by RIE. Finally, it is sintered at ~450°C for ~30 minutes to form a good electrical, ohmic contact to the Si.

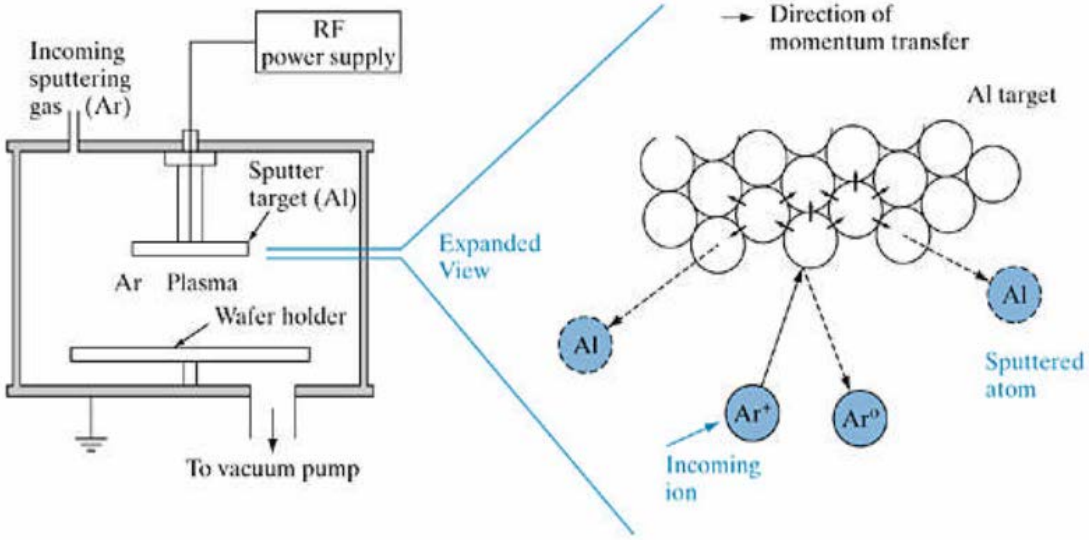


Figure 5—9
Aluminum sputtering by Ar+ ions. The Ar ions with energies of ~1–3 keV physically dislodge Al atoms which end up depositing on the Si wafers held in close proximity. The chamber pressures are kept low such that the mean free path of the ejected Al atoms is long compared to the target to wafer separation.

Fabrication of P-N Junction

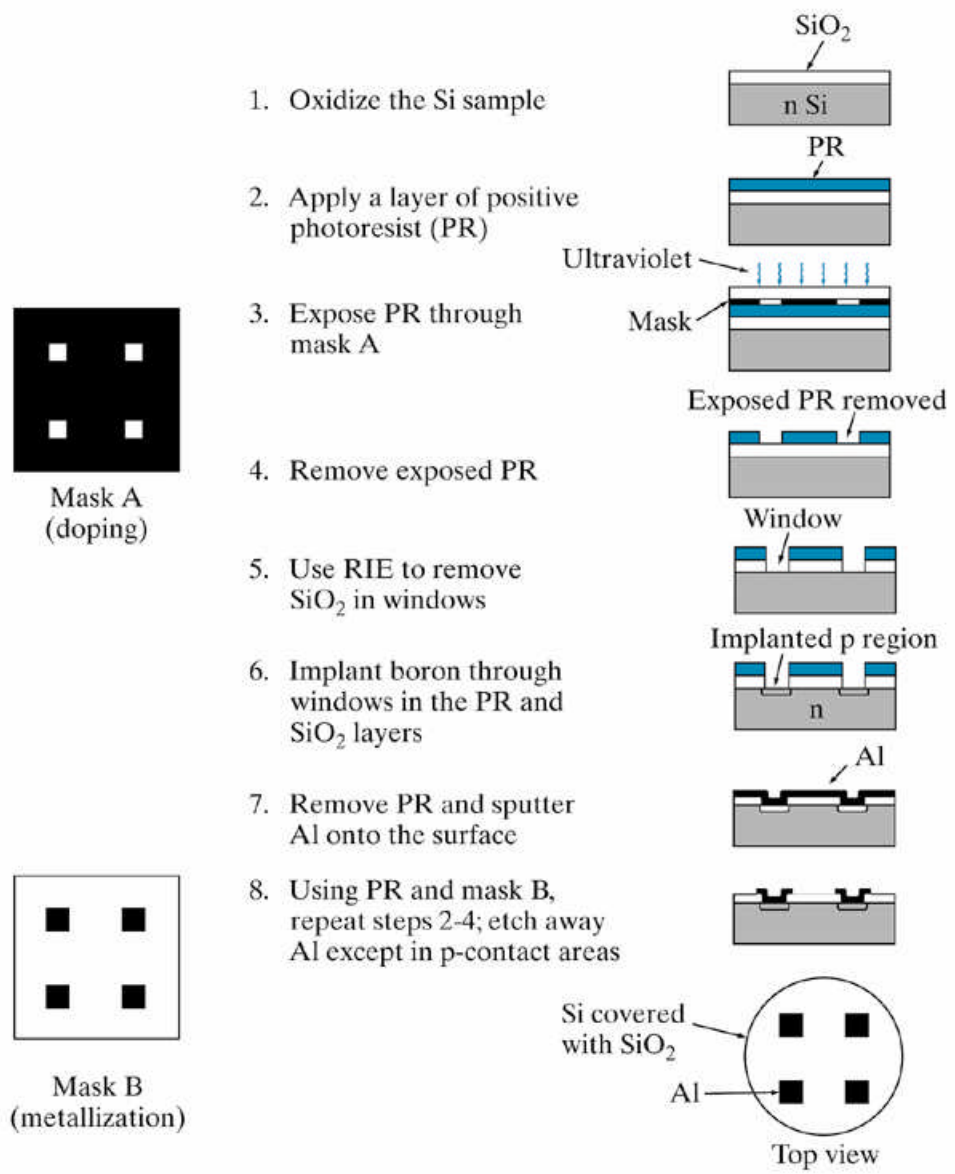
Finalization of IC fabrication

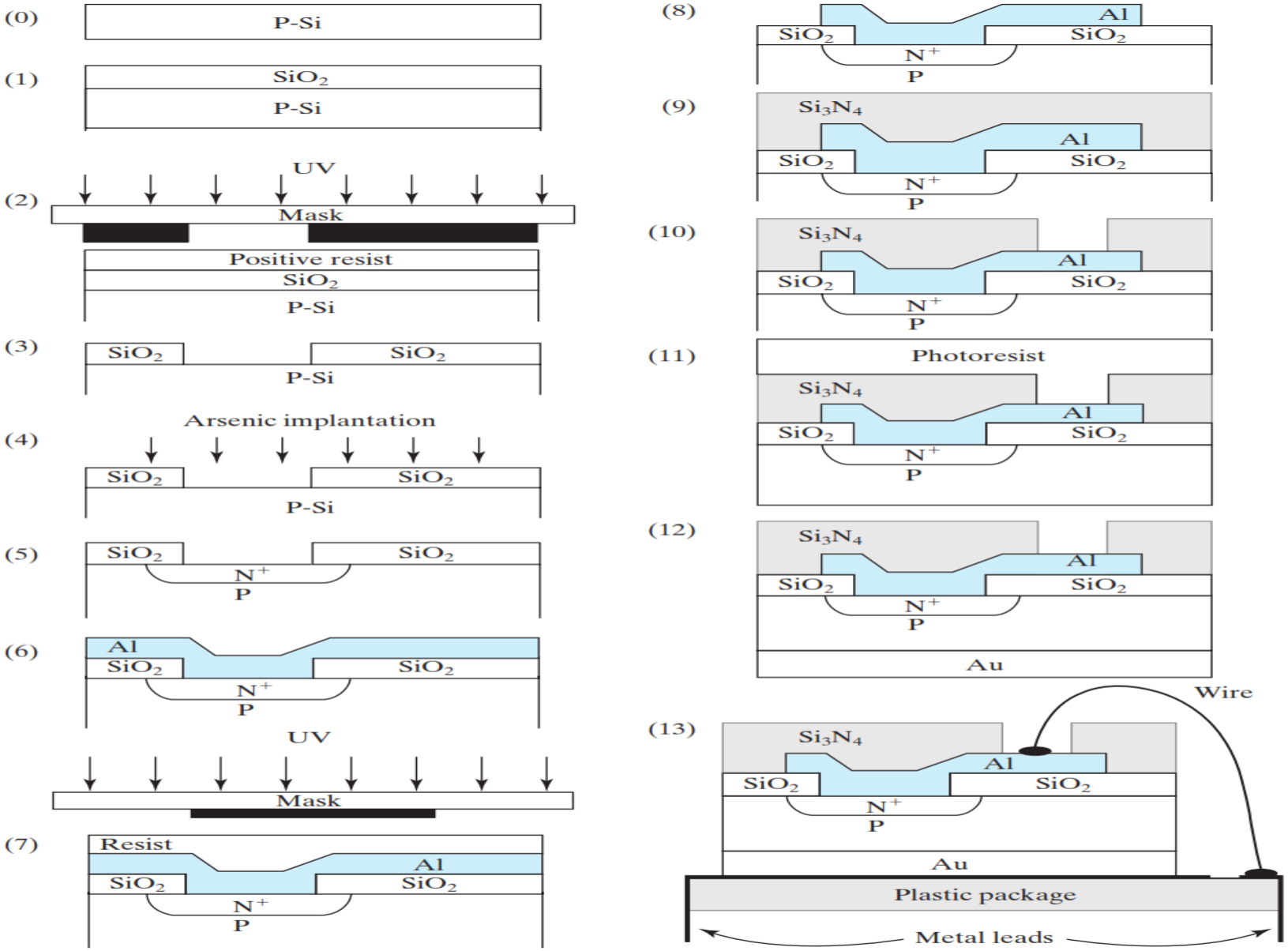
- After the interconnection metallization is complete, a protective overcoat of silicon nitride is deposited using plasma-enhanced CVD.
- Then the individual integrated circuits can be separated by sawing or by scribing and breaking the wafer.
- The final steps of the process are mounting individual devices in appropriate packages and connecting leads to the Al contact regions. Very precise lead bonders are available for bonding Au or Al wire

Fabrication of P-N Junction

Summary for p-n junction fabrication

Figure 5-10
Simplified description of steps in the fabrication of p-n junctions.
For simplicity, only four diodes per wafer are shown, and the relative thicknesses of the oxide, PR, and the Al layers are exaggerated.



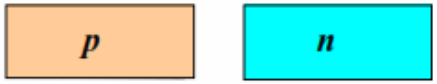


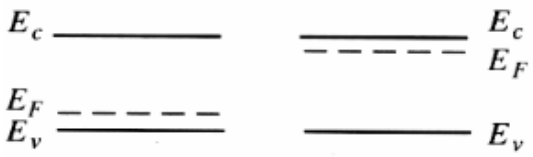
Graphical summary of the major processing steps in the formation of a PN junction diode. (0) Start; (1) oxidation; (2) lithography; (3) oxide etching; (4) As implantation; (5) annealing and diffusion; (6) sputtering Al; (7) lithography; (8) metal etching; (9) CVD nitride deposition; (10) lithography and bonding window etching; (11) removal of oxide from back side of wafer; (12) deposition of Au on back side; and (13) dicing and packaging.

P-N Junction Electrostatics

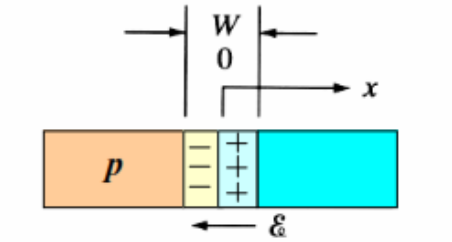
When p- and n-type materials are joined, the electrons and holes diffuse due to their large carrier concentration gradients at the junction where the holes diffuse from the p-side into the n-side, and the electron diffuse from n to p. Thus, the concentration gradients creates a diffusion component of current from p to n region. (**Recall** that electron current is opposite to the direction of electron flow).

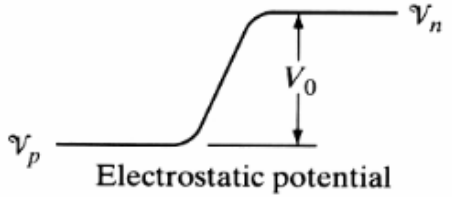
(i) Before electronic contact

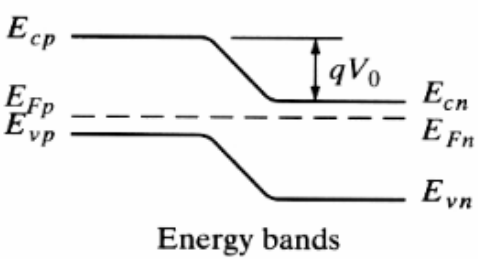




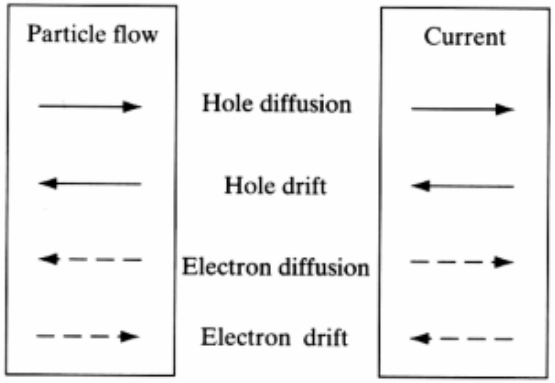
(i) After electronic contact







At transient state:



At equilibrium:

$$J_p(drift) + J_p(diffusion) = 0$$

$$J_n(drift) + J_n(diffusion) = 0$$

P-N Junction Electrostatics

The electrons diffusing from n to p region leave behind uncompensated donor ions (N_d^+) in the n material and holes diffusing from p to n region leave behind uncompensated acceptor ions (N_a^-) in the p material.

However, **the diffusion current can not build up indefinitely** because an opposing electric field is created at the p-n junction, and this electric field is in the direction opposite to that of the diffusion for each type of carrier. The created electric field due to the diffusion, in turn, creates a drift component of current from n to p region, thus opposing the diffusion current. (***See Fig. 5-11***)

Contact Potential

At equilibrium, no net current can flow across the junction and no net buildup of electrons or holes on either side can occur as a function of time.

$$J_p(drift) + J_p(diff) = 0$$

$$J_n(drift) + J_n(diff) = 0$$

Electric field builds up to the point where the net current is zero at equilibrium in a ***space charge region*** or ***transition region*** **(W) around the junction** and there is **an equilibrium potential difference (V_0) across W** according to the fundamental relation of $\mathcal{E}(x) = -dV(x)/dx$

Assume that **the electric field is zero in the neutral regions outside *W*** and then there will be a constant potential V_n in the neutral *n* region, a constant V_p in the neutral ***p*** region, and a potential difference $V_0 = V_n - V_p$ between the two regions.

Contact potential : $V_0 = V_n - V_p$

This contact potential separates the energy band in p-n junction by qV_0. And the separation of the bands at equilibrium is just that required to **make the Fermi level flat throughout the device**.

Contact Potential

At equilibrium,

$$J_n(x) = q\mu_n n(x)\mathcal{E}(x) + qD_n \frac{dn(x)}{dx} = 0 \qquad \left(\frac{dn(x)}{dx} > 0\right)$$

$$J_p(x) = q\mu_p p(x)\mathcal{E}(x) - qD_p \frac{dp(x)}{dx} = 0 \qquad \left(\frac{dp(x)}{dx} < 0\right)$$

For holes $\longrightarrow$ $\dfrac{\mu_p}{D_p}\varepsilon(x) = \dfrac{1}{p(x)}\dfrac{dp(x)}{dx}$ (x : p → n direction)

Using $\mathcal{E}(x) = -dV(x)/dx$ & $\dfrac{D}{\mu} = \dfrac{kT}{q}$

$$-\frac{q}{kT}\frac{dV(x)}{dx} = \frac{1}{p(x)}\frac{dP(x)}{dx}$$

$$-\frac{q}{kT}\int_{V_p}^{V_n} dV = \int_{P_p}^{P_n}\frac{1}{p}dp$$

P_p: holes in p-region (majority)
P_n: holes in n-region (minority)

Contact Potential

$$-\frac{q}{kT}\int_{V_p}^{V_n} dV = \int_{P_p}^{P_n} \frac{1}{p}\,dp$$

$$-\frac{q}{kT}(V_n - V_p) = \ln p_n - \ln p_p = \ln \frac{p_n}{p_p}$$

Since $V_n - V_p = V_0$ →

$$V_0 = \frac{kT}{q}\ln\frac{p_p}{p_n}$$

If we assume electron current →

$$V_0 = \frac{kT}{q}\ln\frac{n_n}{n_p}$$

For a step junction **made up of** N_a acceptors and N_d donors in p and n sides, respectively :

$$V_0 = \frac{kT}{q}\ln\frac{N_a}{n_i^2/N_d} = \frac{kT}{q}\ln\frac{N_a N_d}{n_i^2}$$

$$\frac{p_p}{p_n} = e^{qV_0/kT}$$

At equilibrium,

We know that $p_p n_p = p_n n_n = n_i^2$ so that

$$\frac{p_p}{p_n} = \frac{n_n}{n_p} = e^{qV_0/kT}$$

Equilibrium Fermi Level

Assume that the p_n and p_p are the hole equilibrium concentration values in n and p regions outside W, respectively:

$$\frac{p_p}{p_n} = e^{qV_0/kT}$$

$$p_0 = N_v e^{-(E_F - E_v)/kT} \qquad \textbf{(from Eq. 3-19)}$$

$$\frac{p_p}{p_n} = e^{qV_0/kT} = \frac{N_v e^{-(E_{Fp} - E_{vp})/kT}}{N_v e^{-(E_{Fn} - E_{vn})/kT}}$$

At equilibrium, $E_{Fn} = E_{Fp}$

$$qV_0 = E_{vp} - E_{vn} = E_{cp} - E_{vp} = E_{ip} - E_{in}$$

Example

An abrupt Si *p-n* junction has $N_a = 10^{17}$ cm^{-3} on the *p* side and $N_d = 10^{16}$ cm^{-3} on the *n* side. At 300K, (a) calculate the Fermi levels with respect to E_i on each side, draw an equilibrium band diagram and find V_0 from the diagram; (b) compare the result from (a) with V_0 calculated from Eq. (5-8) in textbook. Assume that for Si $n_i = 1.5\text{x}10^{10}$ cm^{-3}.

Solution:

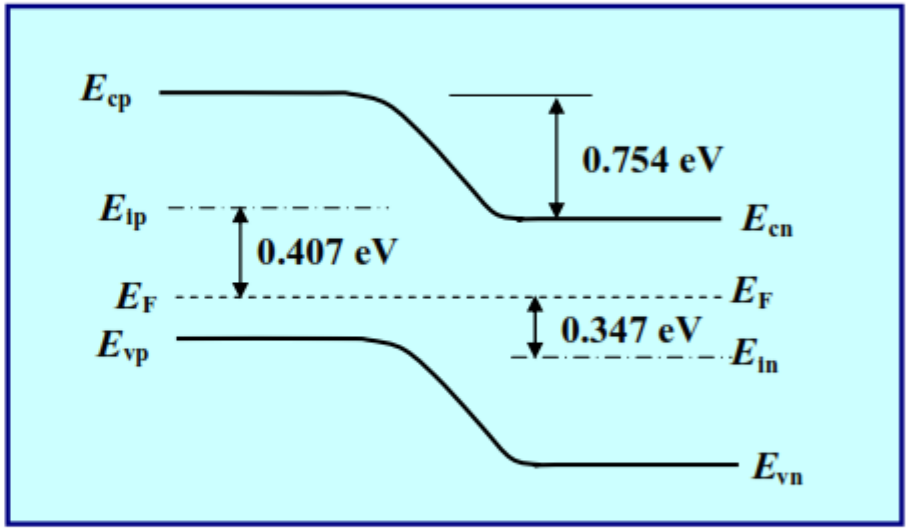
(a)

$$E_{ip} - E_F = kT \ln \frac{p_p}{n_i} = 0.0259 \ln \frac{10^{17}}{(1.5 \cdot 10^{10})} = 0.407\,\text{eV}$$

$$E_F - E_{in} = kT \ln \frac{n_n}{n_i} = 0.0259 \ln \frac{10^{16}}{(1.5 \cdot 10^{10})} = 0.347\,\text{eV}$$

$$qV_0 = E_{ip} - E_{in} = E_{ip} - E_F + E_F - E_{in}$$

$$= 0.407 + 0.347 = 0.754\,\text{eV}$$



(b)

$$qV_0 = kT \ln \frac{N_a N_d}{n_i^2} = 0.0259 \ln \frac{10^{33}}{2.25 \cdot 10^{20}} = 0.754\,\text{eV}$$

Space Charge at a p-n junction

Within the transition region (***W***), electrons and holes are in transit from one side of the junction to the other side. Some electrons diffuse from ***n*** to ***p***, and some are swept by the electric field from ***p*** to ***n*** and **conversely for holes**. However, there are very few carriers within ***W*** at any given time since the electric field serves to sweep out carriers which have wandered into ***W***.

Consider the space charge within ***W*** due only to the uncompensated donor and acceptor ions (***See Fig. 5-12***) and neglect carriers within ***W***. Using the ***depletion approximation*** (***i.e.,*** carrier depletion within **W** and neutrality outside ***W***).

The charge density on the n side = qN_d

The charge density on the p side = $-qN_a$

The cross-sectional area = A

The depletion width: $W = x_{p0} + x_{n0}$

The space charge region in p region = x_{p0}

The space charge region in n region = x_{n0}

Space Charge at a p-n junction

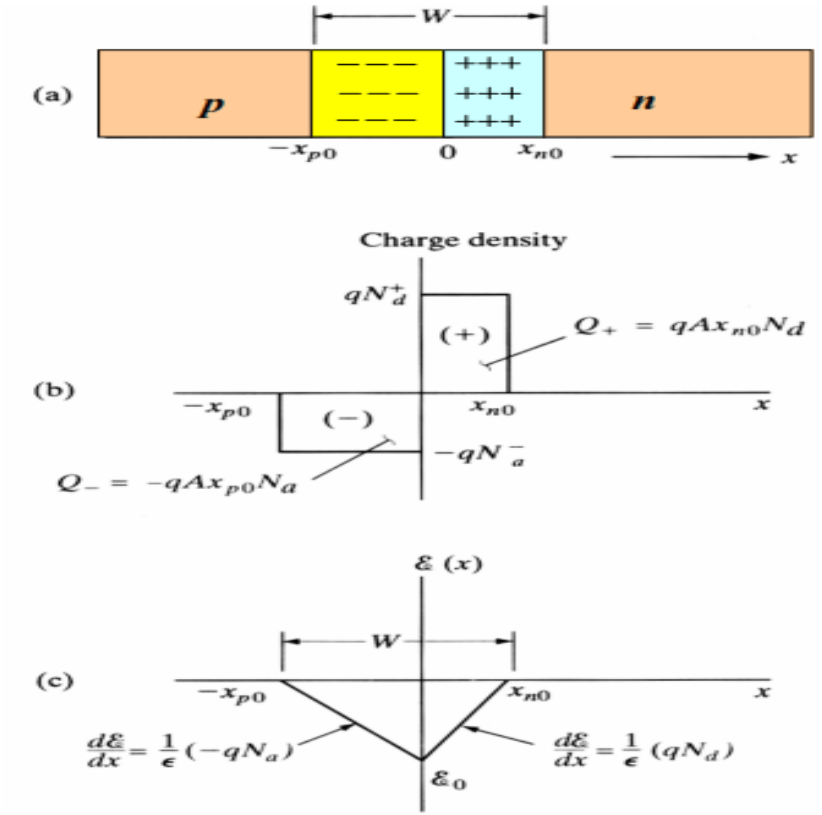


Figure 5—12
Space charge and electric field distribution within the transition region of a p-n junction with $N_d > N_a$: (a) the transition region, with x = 0 defined at the *metallurgical junction; (b) charge density within the transition region, neglecting the free carriers;* (c) the electric field distribution, where the reference direction for the electric field is arbitrarily taken as the +*x*-*direction.*

Space Charge at a p-n junction

$$Q_+ = |Q_-| \longleftarrow$$

Charge magnitudes are equal on both sides
Charge Magnitude = charge density × volume
Q = nV

$$qAx_{p0}N_a = qAx_{n0}N_d \longrightarrow x_{p0}N_a = x_{n0}N_d$$

Calculate the electric field distribution within **W** with ***Poisson's equation***, which relates the gradient of the electric field to the local space charge at any point ***x***:

$$\frac{d\mathcal{E}(x)}{dx} = \frac{q}{\varepsilon}(p - n + N_d^+ - N_a^-)$$

Note here : ε = dielectric constant

Assume complete ionization of the impurities and neglect the contribution of the carriers (**p** and **n**) to the space charge :

$$\frac{d\mathcal{E}(x)}{dx} = \frac{q}{\varepsilon}N_d^+ = \frac{q}{\varepsilon}N_d \qquad (0 < x < x_{n0})$$

$$\frac{d\mathcal{E}(x)}{dx} = -\frac{q}{\varepsilon}N_a^- = -\frac{q}{\varepsilon}N_a \qquad (-x_{p0} < x < 0)$$

Space Charge at a p-n junction

The maximum electric field of $\mathcal{E}_0$ can be found as follows:

$$\int_{\mathcal{E}_0}^{0} d\mathcal{E} = \frac{q}{\varepsilon} N_d \int_0^{x_{n0}} dx \qquad (0 < x < x_{n0})$$

$$\int_0^{\mathcal{E}_0} d\mathcal{E} = -\frac{q}{\varepsilon} N_a \int_{-x_{p0}}^{0} dx \qquad (-x_{p0} < x < 0)$$

$$\boxed{\mathcal{E}_0 = -\frac{q}{\varepsilon} N_d x_{no} = -\frac{q}{\varepsilon} N_a x_{po}}$$

Relate the electric field to the contact potential :

$$\mathcal{E}(x) = -\frac{dV(x)}{dx} \quad or \quad -V_0 = \int_{-x_{p0}}^{x_{n0}} \mathcal{E}(x)dx$$ (Area under the graph)

$$\boxed{V_0 = -\frac{1}{2}\mathcal{E}_0 W = \frac{1}{2}\frac{q}{\varepsilon} N_d x_{n0} W \qquad \rightarrow \mathcal{E}_0 = \frac{-2V_0}{W}}$$

Since $x_{p0} N_a = x_{n0} N_d$ **and** $W = x_{p0} + x_{n0}$ $\rightarrow$ $x_{n0} = W \dfrac{N_a}{N_a + N_d}$

$$V_0 = \frac{1}{2}\frac{q}{\varepsilon} N_d x_{n0} W = \frac{1}{2}\frac{q}{\varepsilon}\frac{N_a N_d}{N_a + N_d} W^2$$

Space Charge at a p-n junction

$$W = \left[\frac{2\varepsilon V_0}{q}\left(\frac{N_a + N_d}{N_a N_d}\right)\right]^{1/2} = \left[\frac{2\varepsilon V_0}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}$$

Since $$V_0 = \frac{kT}{q}\ln\frac{N_a}{n_i^2 / N_d} = \frac{kT}{q}\ln\frac{N_a N_d}{n_i^2}$$

$$W = \left[\frac{2\varepsilon}{q}\left(\frac{kT}{q} ln\ \frac{N_a N_d}{n_i^2}\right)\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}$$

$$x_{n0} = W\frac{N_a}{N_a + N_d} = \left\{\frac{2\varepsilon V_0}{q}\left[\frac{N_a}{N_d(N_a + N_d)}\right]\right\}^{1/2}$$

$$x_{p0} = W\frac{N_d}{N_a + N_d} = \left\{\frac{2\varepsilon V_0}{q}\left[\frac{N_d}{N_a(N_a + N_d)}\right]\right\}^{1/2}$$

Space Charge at a Junction (1)

Depletion Approximation: Space charge within the transition region is only due to the *uncompensated donor and acceptor ions*. The carrier concentration inside the space charge region is neglected, and the semiconductor outside the space charge region is neutral.

Poisson's equation: Relationship between the derivative of electric field at any point x, and the charge density at that point. This is one of the fundamental laws of semiconductor electrostatics

Mathematically, $$\frac{dE(x)}{dx} = \frac{q}{\varepsilon}(p - n + N_d^+ - N_a^-)$$

Charge Neutrality condition: Even though the charges flow from one semiconductor to the other, the two semiconductors as a whole remains charge neutral. So the total charge in the n-type region must balance the total charge in the p-type region.

Space Charge at a Junction (2)

Mathematically, from charge neutrality,

$$qAx_{p0}N_a = qAx_{n0}N_d$$

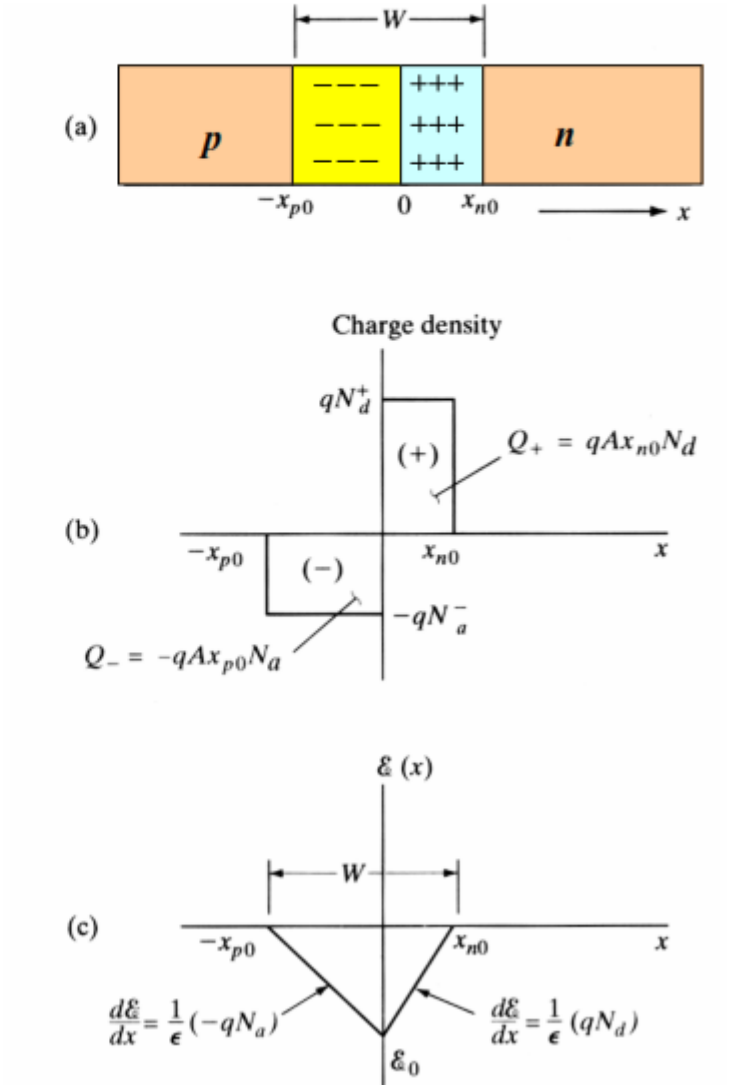
$$\Rightarrow x_{p0}N_a = x_{n0}N_d$$

From application of Poisson's equation in the two regions:

$$\frac{dE(x)}{dx} = -\frac{q}{\varepsilon}N_a, -x_{p0} < x < 0$$

$$\frac{dE(x)}{dx} = \frac{q}{\varepsilon}N_d, 0 < x < x_{n0}$$

$$E_0 = -\frac{q}{\varepsilon}N_d x_{n0} = -\frac{q}{\varepsilon}N_a x_{p0}$$



Overall junction quantities

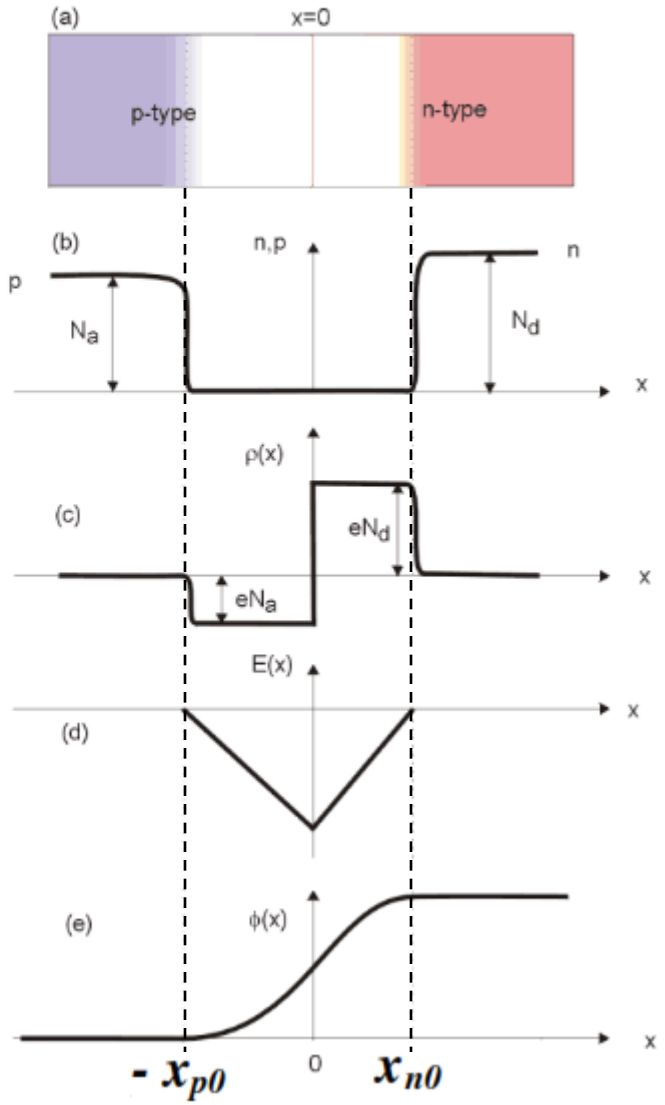
Junction charge color code → Blue: holes
Red: electrons

p and n side carrier density

Junction space charge

Variation of junction electric field

Variation of junction potential



Calculation of depletion depth

$$E(x) = -\frac{dV(x)}{dx} \Rightarrow -V_0 = \int_{-x_{p0}}^{x_{n0}} E(x)dx$$ = Area under the E(x) vs. x triangle

So, $V_0 = -\frac{1}{2}E_0 W = \frac{1}{2}\frac{q}{\varepsilon}N_d x_{n0} W$ From balance of charge: $x_{n0}N_d = x_{p0}N_a$

Also, $W = x_{n0} + x_{p0}$ $x_{n0} = W\, N_a/(N_a + N_d)$

Substituting the value of x_{n0}, $V_0 = \frac{1}{2}\frac{q}{\varepsilon}\frac{N_a N_d}{N_a + N_d}W^2$

So, $W = \left[\frac{2\varepsilon V_0}{q}\left(\frac{N_a + N_d}{N_a N_d}\right)\right]^{1/2} = \left[\frac{2\varepsilon V_0}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}$

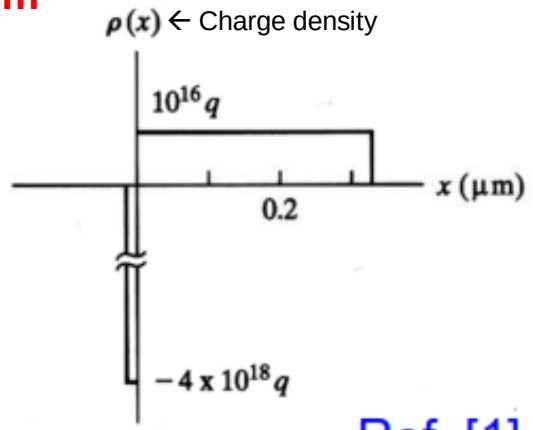
$x_{p0} = \frac{WN_d}{N_a + N_d} = \frac{W}{1 + N_a/N_d}$ $x_{n0} = \frac{WN_a}{N_a + N_d} = \frac{W}{1 + N_d/N_a}$

$= \left\{\frac{2\varepsilon V_0}{q}\left[\frac{N_d}{N_a(N_a + N_d)}\right]\right\}^{1/2}$ $= \left\{\frac{2\varepsilon V_0}{q}\left[\frac{N_a}{N_d(N_a + N_d)}\right]\right\}^{1/2}$

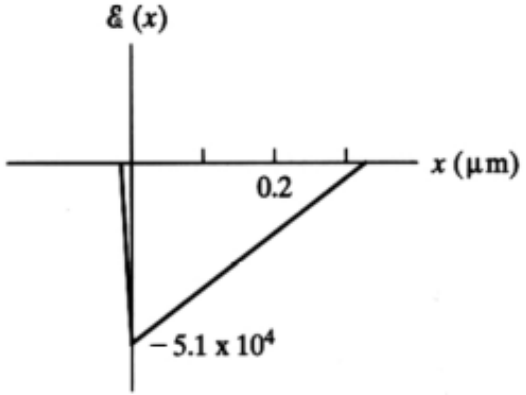
Junction depth as a function of the built-in voltage. **Very important!!**

Example

Boron is implanted into an *n*-type Si sample ($N_d = 10^{16}$ cm^{-3}), forming an abrupt junction of square section, with area = 2×10^{-3} cm^2. Assume that the acceptor concentration in the *p*-type region is $N_a = 4\times10^{18}$ cm^{-3}. Calculate V_0, x_{n0}, x_{p0}, Q_+, and $\mathcal{E}_0$ for this junction at equilibrium (300K). Sketch E (x) and charge density to scale. For Si $\varepsilon_r = 11.8$ and $\varepsilon_0 = 8.85\times10^{-14}$ F/cm



Ref. [1]



Solution:

$$V_0 = \frac{kT}{q}\ln\frac{N_a N_d}{n_i^2} = 0.0259\ln\frac{4\cdot10^{34}}{2.25\cdot10^{20}} = 0.85\text{V}$$

$$W = \left[\frac{2\varepsilon V_0}{q}\left(\frac{1}{N_a}+\frac{1}{N_d}\right)\right]^{1/2}$$

$$= \left[\frac{2(11.8\cdot8.85\cdot10^{-14})\cdot0.85}{1.6\cdot10^{-19}}(0.25\cdot10^{-18}+10^{-16})\right]^{1/2} = 3.34\cdot10^{-5}\text{cm} = 0.334\mu\text{m}$$

$$x_{n0} = \frac{W}{1+N_d/N_a} = \frac{0.334\,\mu\text{m}}{1+0.0025} = 0.333\,\mu\text{m}$$

$$x_{p0} = \frac{W}{1+N_a/N_d} = \frac{0.334\,\mu\text{m}}{1+400} = 8.3\cdot10^{-4}\,\mu\text{m} = 8.3\,\text{Å}.$$

$$Q_+ = -Q_- = qAx_{n0}N_d = 1.07\cdot10^{-10}\text{C}$$

$$\mathcal{E}_0 = \frac{-qN_d x_{n0}}{\varepsilon} = \frac{-(1.6\cdot10^{-19})(10^{16})(3.3\cdot10^{-5})}{(11.8)(8.85\cdot10^{-14})} = -5.1\cdot10^4\,\text{V/cm}$$

$\rightarrow = -2V_0/W$

Forward & Reverse Biased Junctions; Steady State

One useful feature of a ***p-n*** junction:

- Current flows quite freely in the ***p*** to ***n*** direction when ***p*** region has a positive external voltage bias relative to ***n*** region: ***forward bias and forward current***

- Whereas, virtually no current flows when ***p*** region is made negative relative to ***n*** region: ***reverse bias and reverse current***

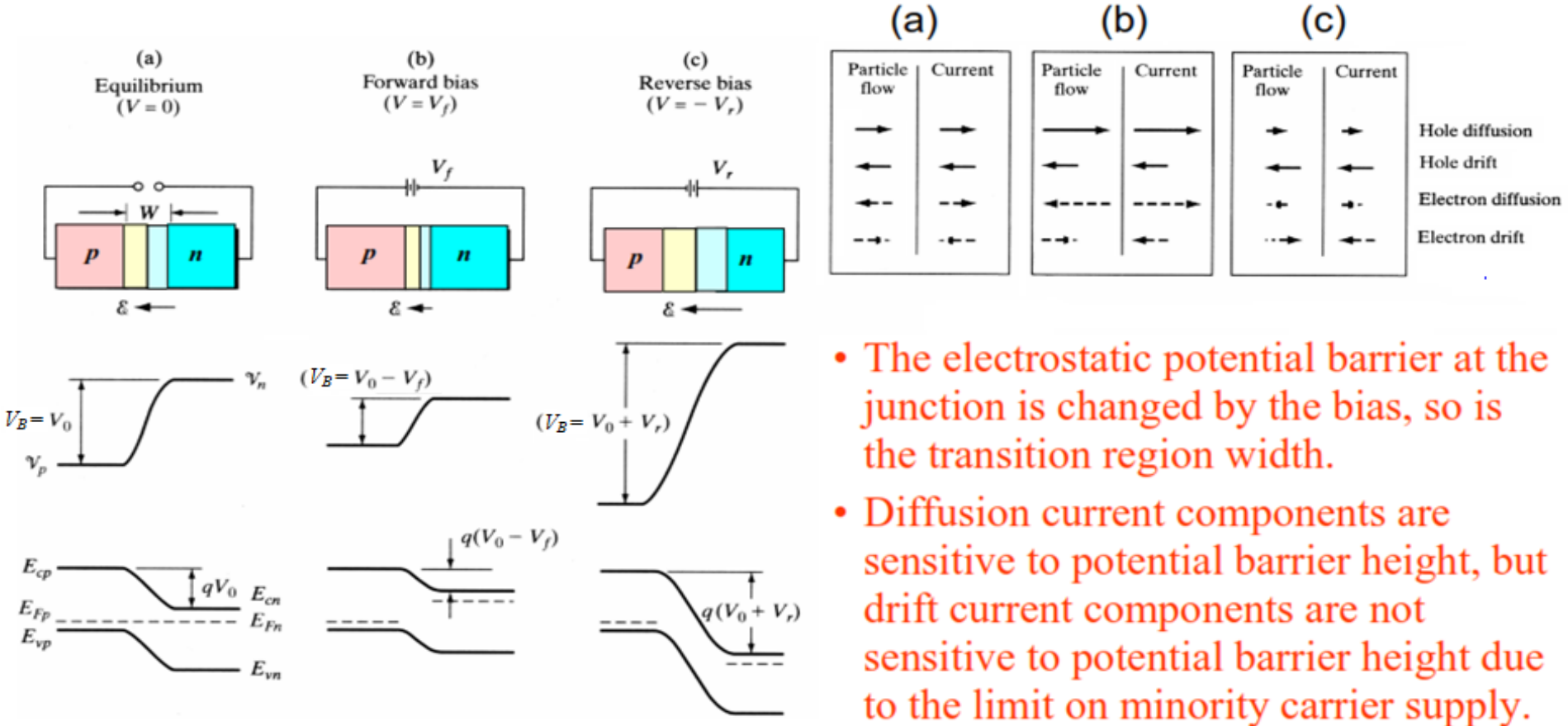
Applications of a ***p-n*** junction:

- Asymmetry of the current flow makes the p-n junction diode very useful as a rectifier.

- Biased p-n junctions can be used as voltage-variable capacitors, photocells, light emitters, many more devices which are basic to modern electronics.

- Two or more junctions can be used to form transistors and controlled switches.

Current Flow at a Junction: Qualitative



- The electrostatic potential barrier at the junction is changed by the bias, so is the transition region width.
- Diffusion current components are sensitive to potential barrier height, but drift current components are not sensitive to potential barrier height due to the limit on minority carrier supply.

E_{Fp}-E_{Fn}=0

No current will flow
J(diff)=J(drift)

E_{Fp}-E_{Fn}= qV_f

Many electrons and holes will diffuse through the junction → forward current (mainly diffusion)

E_{Fp}-E_{Fn}= qV_r

No diffusion due to large E-field. But minority carriers swept thorugh the barrier → reverse drift current (small)

Effects of a bias at a p-n junction; transition region width and electric field, electrostatic potential, energy band diagram, and particle flow and current directions within W for (a) equilibrium, (b) forward bias, and (c) reverse bias.

Diffusion current is sensitive to the height of potential barrier:

(a) Equilibrium (V = 0) : energy barrier = qV_0

(b) Forward Bias (V = V_f) : energy barrier = $q(V_0-Vf)$

(c) Reverse Bias (V = $-V_r$) : energy barrier = $q(V_0+Vr)$

Drift current is insensitive to the height of the potential barrier:

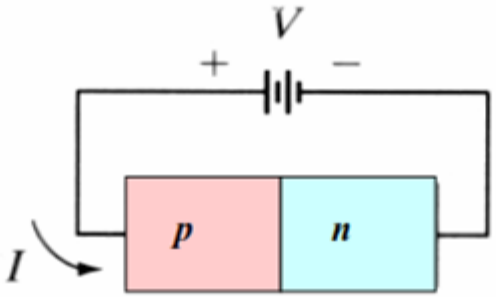
Drift current is limited not by how fast carriers are swept away down the barrier, but rather how often.

For example, minority carrier electrons on the **p** side which wander into the transition region (**W**) will be swept down the barrier by the electric field. This gives rise to the electron drift current ! However, this current is small not because of the size of the barrier, but because there are very few minority carrier electrons in the **p** side to participate. Every electrons on the **p** side which diffuses to the transition region will be swept down the potential energy hill regardless of the size of the hill. The electron drift current does not depend on how fast an individual electron is swept from p to n, but rather on how many electrons are swept down the barrier per second.

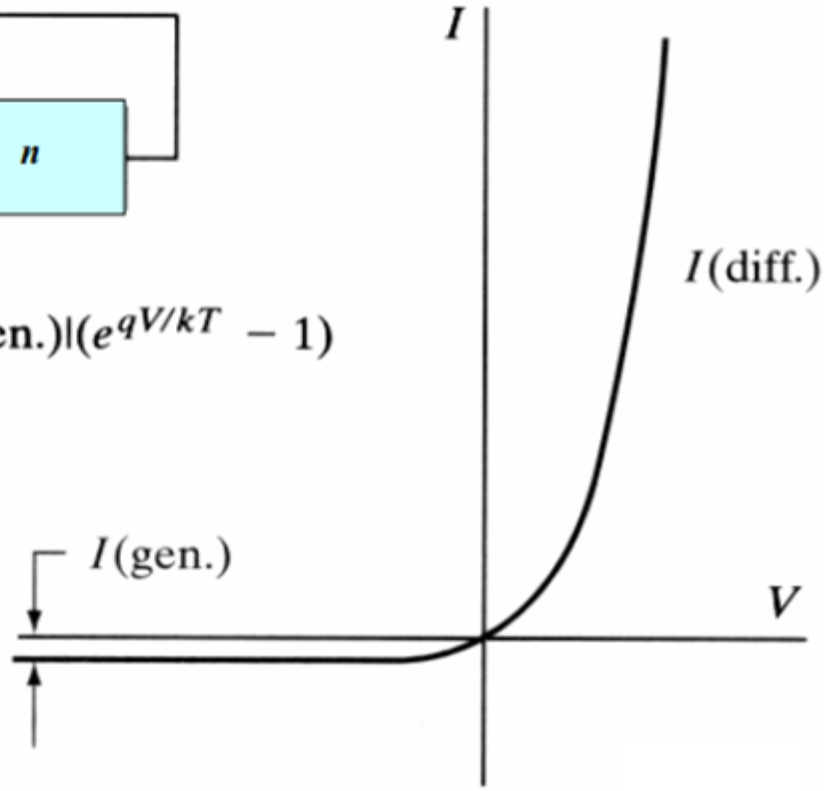
Similar comments may be applied regarding the drift of the minority carrier holes from the **n** side to the **p** side of the junction. Approximately, electron and hole drift currents **at the junction** are independent of the applied voltage.

I-V Characteristics of a P-N Junction

1. V > 0, diffusion current is dominant.
2. V = 0, diffusion current and drift current cancels each other.
3. V < 0, drift current is dominant.



$$I = |I(\text{gen.})|(e^{qV/kT} - 1)$$



I-V characteristics:

$$I = I_0\left(\exp\left(\frac{qV}{kT} - 1\right)\right)$$

Equilibrium ($V = 0$)

Forward bias ($V = V_f$)

Reverse bias ($V = -V_r$)

I_0 = reverse saturation current, which is drift current under reverse bias (also same as generation current)

Carrier Injection (1)

At equilibrium (from earlier results): $\frac{p_p}{p_n} = \exp\left(\frac{qV_0}{kT}\right)$ ← For holes

Under applied bias V: $\frac{p(-x_{p0})}{p(x_{n0})} = \exp\left(\frac{q(V_0 - V)}{kT}\right)$ ←Injection of holes from p- to n-region across W

Taking ratio of the above equations, $\frac{p(x_{n0})}{p_n} = \exp(\frac{qV}{kT})$ [assuming that the majority carrier does not change much due to carrier injection, $p_p = p(-x_{p0})$]

$\Longrightarrow \quad p(x_{n0}) = p_n \exp(\frac{qV}{kT})$

So, the ratio of the non-equilibrium carrier density to the equilibrium carrier density is dependent on the applied voltage

Change in carrier density, $\Delta p_n = p(x_{n0}) - p_n = p_n[\exp(\frac{qV}{kT}) - 1]$

Similarly, $\Delta n_p = n(x_{p0}) - n_p = n_p[\exp(\frac{qV}{kT}) - 1]$

Carrier Injection (2)

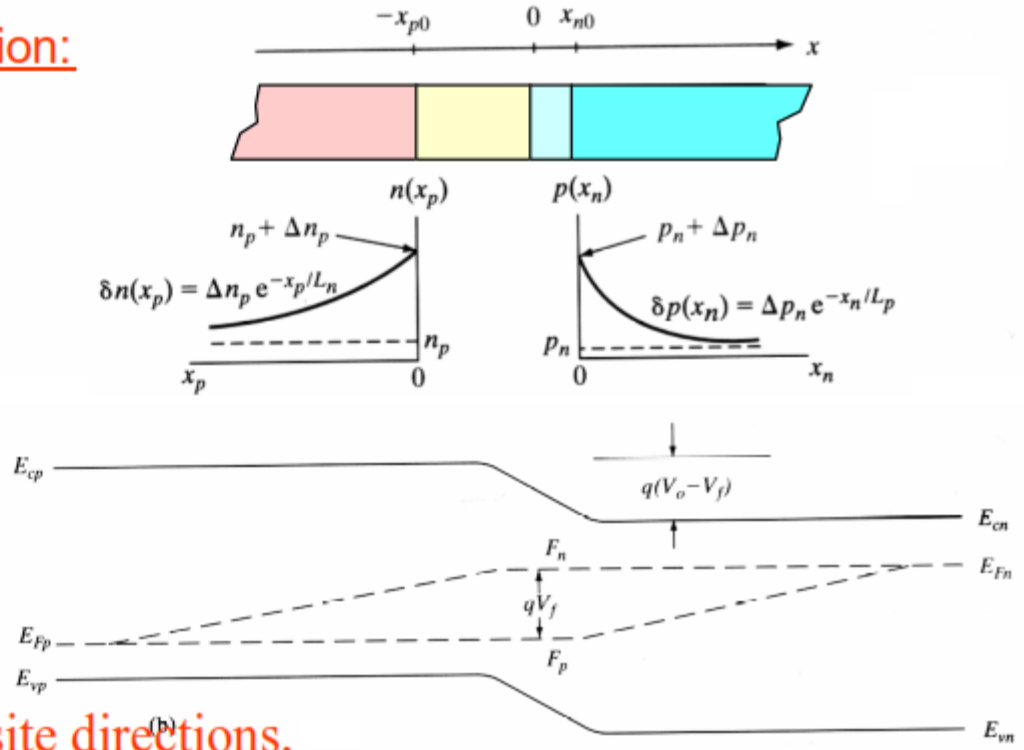
From the solution of the diffusion equation:

$$\delta p(x_n) = \Delta p_n \exp(-x_n / L_p)$$

$$= p_n[\exp(\frac{qV}{kT}) - 1]\exp(-x_n / L_p)$$

$$\delta n(x_p) = \Delta n_p \exp(-x_p / L_n)$$

$$= n_p[\exp(\frac{qV}{kT}) - 1]\exp(-x_p / L_n)$$



[Note: Distances x_p and x_n are in opposite directions, starting from x_{p0} and x_{n0} respectively]

Hence the diffusion current (I = JA) for both holes and electrons will be:

$$I_p(x_n) = -qAD_p \frac{d\delta p(x_n)}{dx_n} = qA\frac{D_p}{L_p}\Delta p_n \exp(-x_n / L_p) = qA\frac{D_p}{L_p}\delta p(x_n)$$

$$I_n(x_p) = qAD_n \frac{d\delta n(x_p)}{dx_p} = qA\frac{D_n}{L_n}\Delta n_p \exp(-x_p / L_p) = -qA\frac{D_n}{L_n}\delta n(x_p)$$

Carrier Injection (3)

The current in the transition region (W) will be only diffusion current, and can be calculated from previous equations at the edges of transition region x_{n0} and x_{p0}

$$I_p(x_n = 0) = qA\frac{D_p}{L_p}\Delta p_n = qA\frac{D_p}{L_p}p_n[\exp(qV/kT) - 1]$$

$$I_n(x_p = 0) = -qA\frac{D_n}{L_n}\Delta n_p = -qA\frac{D_n}{L_n}n_p[\exp(qV/kT) - 1]$$

Assuming no recombination in space charge region, the total current is give as

$$I = I_p(x_n = 0) - I_n(x_p = 0) = qA\frac{D_p}{L_p}\Delta p_n + qA\frac{D_n}{L_n}\Delta n_p$$

(negative sign in the second term to account for the different directions for x_n and x_p)

$$I = qA\left(\frac{D_p}{L_p}p_n + \frac{D_n}{L_n}n_p\right)[\exp(qV/kT) - 1] = I_0[\exp(qV/kT) - 1]$$

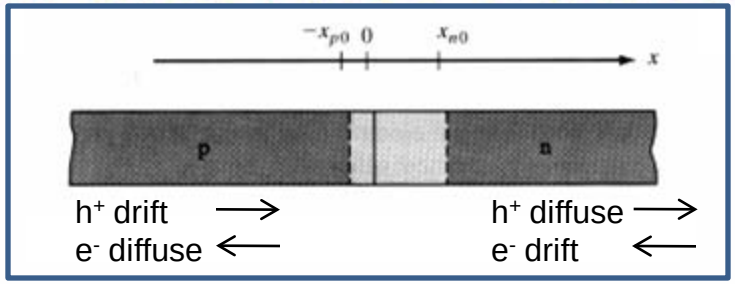
This is called the Diode Equation

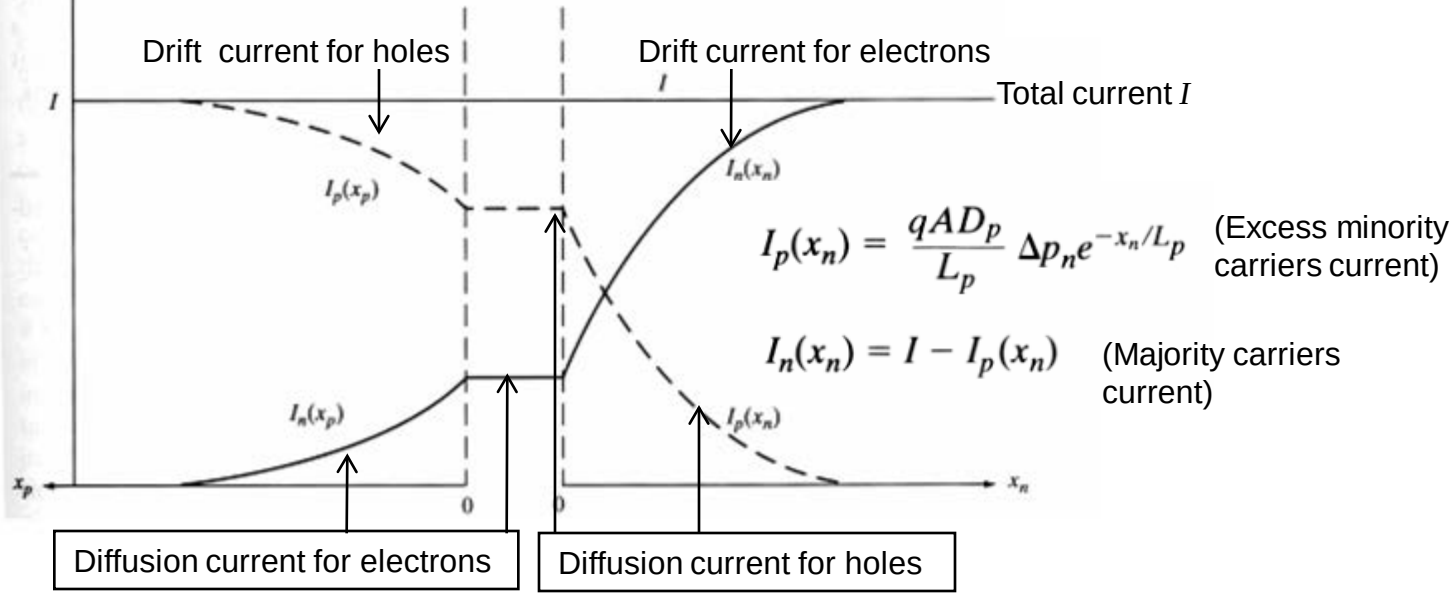
$$I_0 = qA\left(\frac{D_p}{L_p}p_n + \frac{D_n}{L_n}n_p\right)$$

I_0 = reverse saturation current, which is drift current under reverse bias (also same as generation current)

Electron and Hole Current Components

Under forward bias





$$I_p(x_n) = \frac{qAD_p}{L_p}\Delta p_n e^{-x_n/L_p}$$ (Excess minority carriers current)

$$I_n(x_n) = I - I_p(x_n)$$ (Majority carriers current)

- As the hole current decreases exponentially with distance, the electron current must increase so that the total current remains constant, and vice versa
- Also, the minority carrier current (predominant near the junction) is mostly diffusion, but the majority carrier current (away from the junction) is mostly drift.

Reverse Bias

When $V = -V_r < 0$ & $V_r >> kT/q$

$$\Delta p_n = p_n(e^{q(-V_r)/kT} - 1) = -p_n$$

$$\Delta n_p = n_p(e^{q(-V_r)/kT} - 1) = -n_p$$

Excess minority carrier concentrations in neutral regions

$$\delta p(x_n) = \Delta p_n e^{-x_n/L_p} = p_n(e^{qV/kT} - 1)e^{-x_n/L_p}$$

$$\delta n(x_p) = \Delta n_p e^{-x_p/L_n} = n_p(e^{qV/kT} - 1)e^{-x_p/L_n}$$

Depletion of carriers below the equilibrium values extends approximately a diffusion length beyond each side of the transition region.

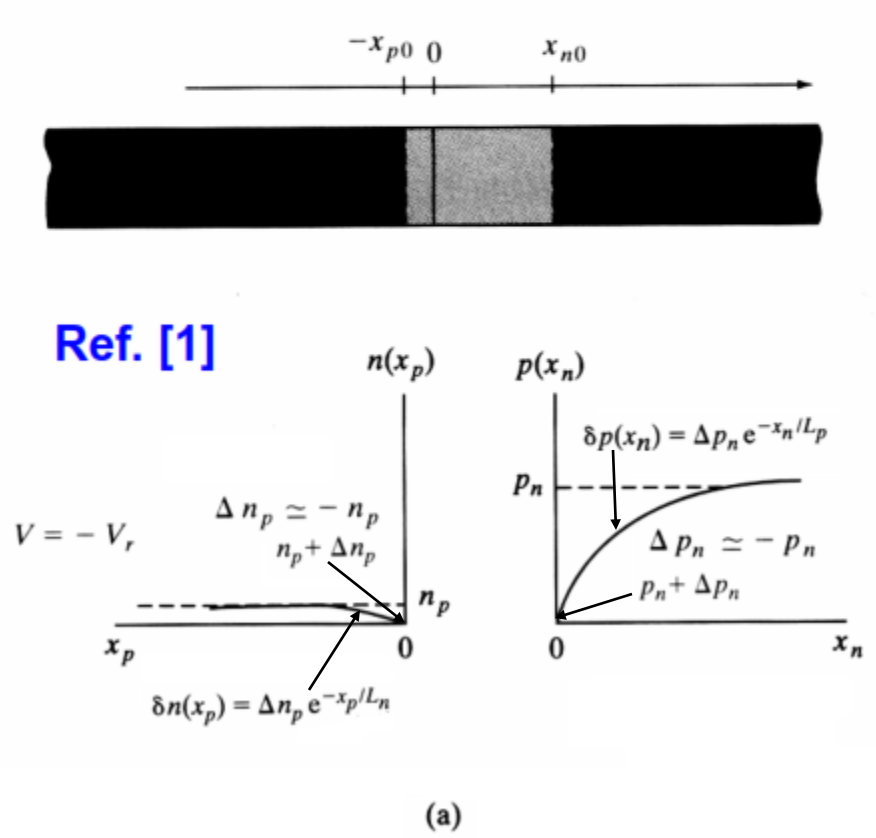
Reverse Bias

$$\Delta p_n = p(x_{n0}) - p_n = p_n[\exp(\frac{q(-V_r)}{kT}) - 1] \approx -p_n \Rightarrow p_n = 0$$

$$\Delta n_p = n(x_{p0}) - n_p = n_p[\exp(\frac{q(-V_r)}{kT}) - 1] = -n_p \Rightarrow n_p = 0$$

$$\Rightarrow \quad p(x_{n0}) = n(-x_{p0}) = 0$$

- Note that under reverse bias the minority carrier concentration reduces to almost zero near the edge of the depletion regions.
- The reverse saturation current is caused by the drift of the carriers across the depletion region, which are supplied by the diffusion of thermally generated minority carriers from outside the depletion region



EXAMPLE

An abrupt Si p-n junction ($A = 10^{-4}$ cm^2) has the following properties at 300 K:

p side	*n side*
$N_a = 10^{17}$ cm^{-3}	$N_d = 10^{15}$
$\tau_n = 0.1$ μs	$\tau_p = 10$ μs
$\mu_p = 200$ cm^2/V–s	$\mu_n = 1300$
$\mu_n = 700$	$\mu_p = 450$

The junction is forward biased by 0.5 V. What is the forward current? What is the current at a reverse bias of −0.5 V?

SOLUTION

$$I = qA\left(\frac{D_p}{L_p}p_n + \frac{D_n}{L_n}n_p\right)(e^{qV/kT} - 1) = I_0(e^{qV/kT} - 1)$$

$$p_n = \frac{n_i^2}{n_n} = \frac{(1.5 \times 10^{10})^2}{10^{15}} = 2.25 \times 10^5 \text{ cm}^{-3}$$

$$n_p = \frac{n_i^2}{p_p} = \frac{(1.5 \times 10^{10})^2}{10^{17}} = 2.25 \times 10^3 \text{ cm}^{-3}$$

For minority carriers,

$$D_p = \frac{kT}{q}\mu_p = 0.0259 \times 450$$

$$= 11.66 \text{ cm}^2/\text{s on the n side}$$

$$D_n = \frac{kT}{q}\mu_n = 0.0259 \times 700$$

$$= 18.13 \text{ cm}^2/\text{s on the p side}$$

$$L_p = \sqrt{D_p\tau_p} = \sqrt{11.66 \times 10 \times 10^{-6}} = 1.08 \times 10^{-2} \text{ cm}$$

$$L_n = \sqrt{D_n\tau_n} = \sqrt{18.13 \times 0.1 \times 10^{-6}} = 1.35 \times 10^{-3} \text{ cm}$$

$$I_0 = qA\left(\frac{D_p}{L_p}p_n + \frac{D_n}{L_n}n_p\right)$$

$$= 1.6 \times 10^{-19} \times 0.0001\left(\frac{11.66}{0.0108}2.25 \times 10^5 + \frac{18.13}{0.00135}2.25 \times 10^3\right)$$

$$= 4.370 \times 10^{-15} \text{ A}$$

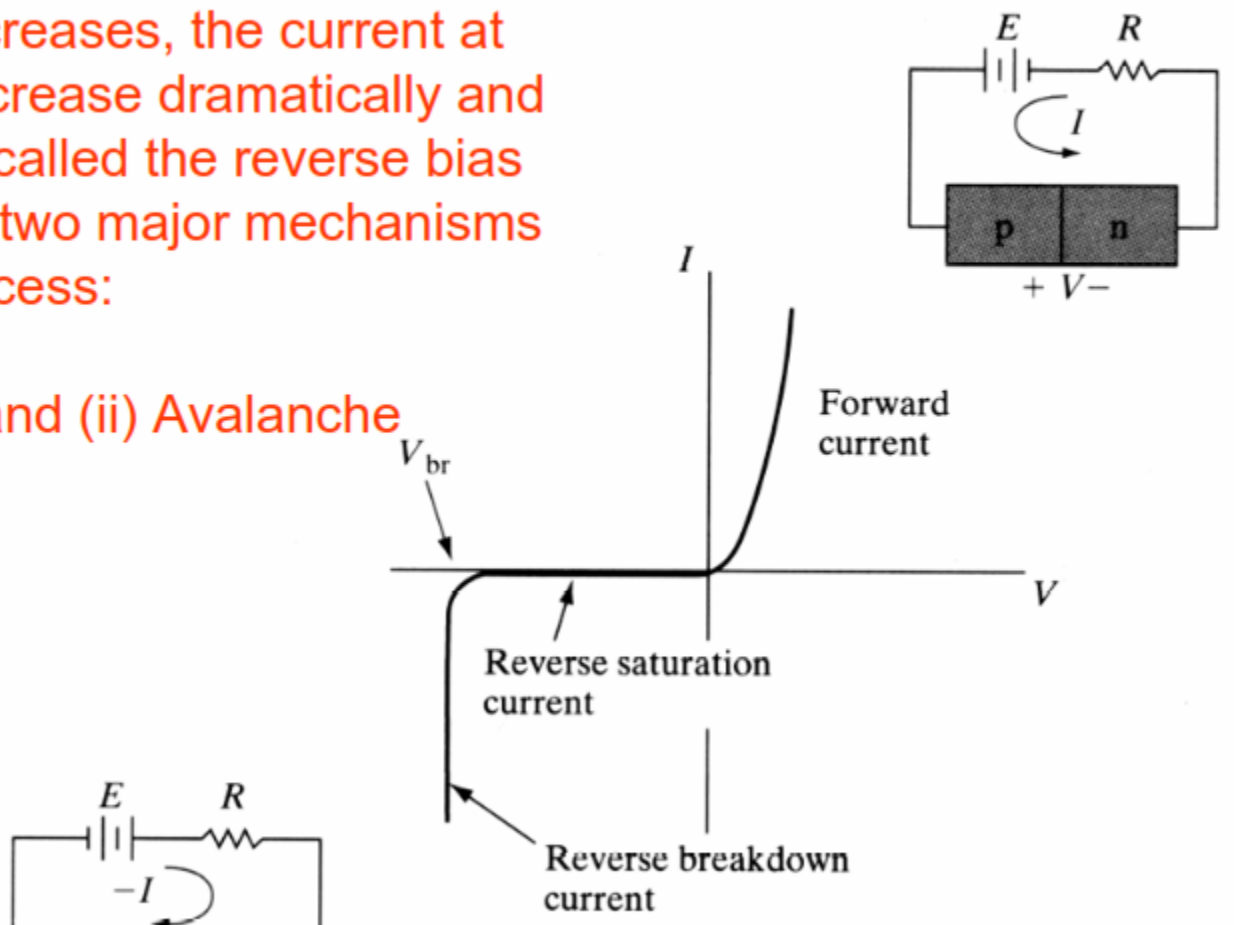
$I = I_0(e^{0.5/0.0259} - 1) \approx \mathbf{1.058 \times 10^{-6}}$ **A** in forward bias.

$I = -I_0 = \mathbf{-4.37 \times 10^{-15}}$ **A** in reverse bias.

Reverse-Bias Breakdown

As the reverse bias increases, the current at some point starts to increase dramatically and uncontrollably. This is called the reverse bias breakdown. There are two major mechanisms for this breakdown process:

(i) Zener Breakdown, and (ii) Avalanche Breakdown

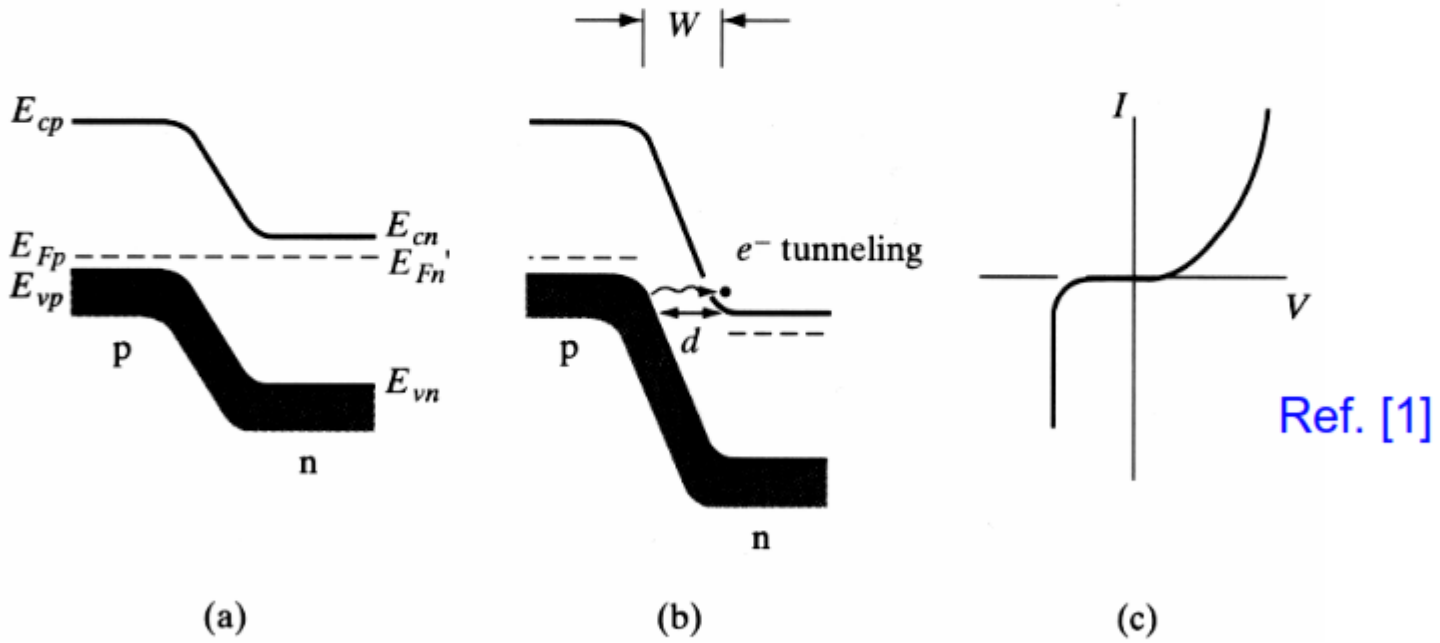


Ref. [1]

Zener Breakdown

Occurs at highly doped p-n junction (p^+-n or p-n^+ junctions)

Process: Zener breakdown process can be thought of as electric field induced ionization of the host atoms *at the junction*. The electric field required for this type of ionization is on the order of 10^6 V/cm. From band diagram, the electrons can be seen to go *through the forbidden gap* to reach the conduction band from the valence band, which is called tunneling



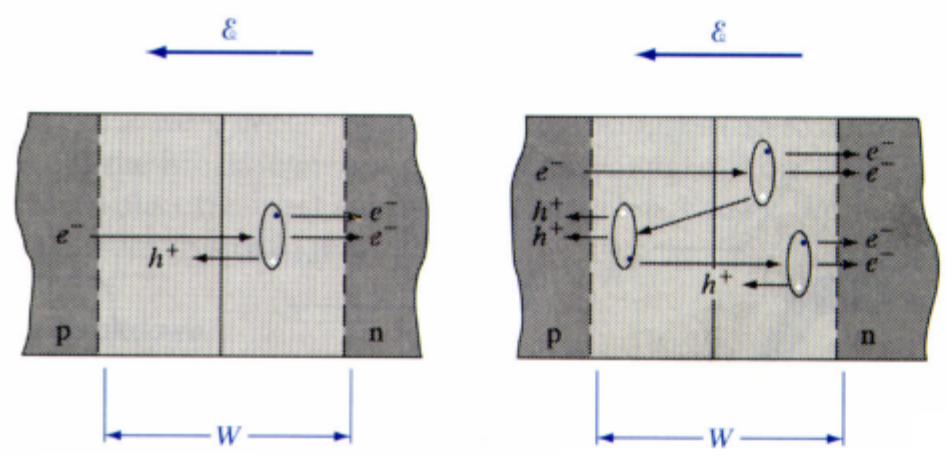
Ref. [1]

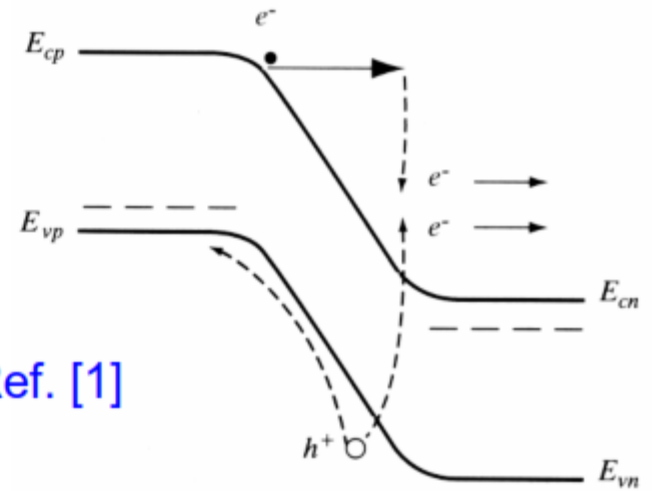
The Zener effect: (a) heavily doped junction at equilibrium: (b) reverse bias with electron tunneling from p to n; (c) I-V characteristic.

Avalanche Breakdown (1)

Process: The carriers moving under the effect of high electric field gets accelerated by it, and causes ionization of the atoms by hitting them (a process called "impact ionization"). The secondary carriers also create impact ionization, and so on, leading to an avalanche of carriers. This increases the current dramatically, and is called the avalanche breakdown, where the diode starts conducting in the reverse direction.

Impact ionization rate: $$M = \frac{1}{1-\left(\frac{V}{V_{br}}\right)^n}$$





Ref. [1]

Capacitance of p-n Junctions

General Definition for capacitance: $C = \left| \frac{dQ}{dV} \right|$

Two types of capacitances in a p-n junction:

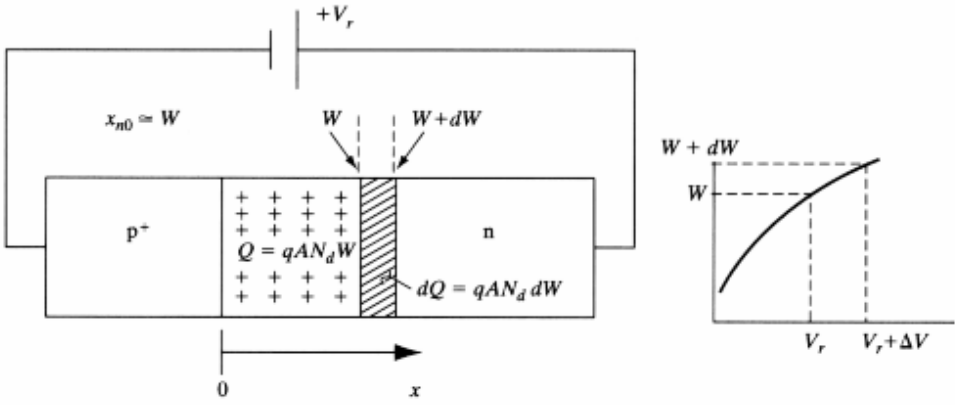
(1) Junction Capacitance (Depletion Capacitance):
Due to the charge dipole across the depletion region

(2) Charge Storage Capacitance (Diffusion Capacitance):
Due to minority carrier storage across the depletion region, which causes the voltage to lag behind as the current changes.

Junction Capacitance

With bias

$$W = \left[\frac{2\varepsilon(V_0 - V)}{q}\left(\frac{N_a + N_d}{N_a N_d}\right)\right]^{1/2}$$



$$|Q| = qAx_{n0}N_d = qAx_{p0}N_a$$

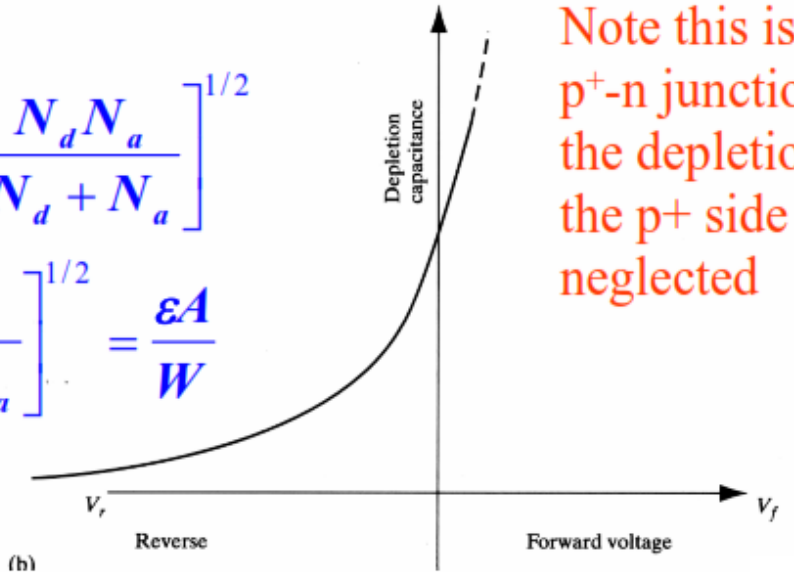
$$x_{n0} = \frac{N_a}{N_a + N_d}W \qquad x_{p0} = \frac{N_d}{N_a + N_d}W$$

$$|Q| = qA\frac{N_d N_a}{N_d + N_a}W = A\left[2q\varepsilon(V_0 - V)\frac{N_d N_a}{N_d + N_a}\right]^{1/2}$$

$$C_j = \left|\frac{dQ}{d(V_0 - V)}\right| = \frac{A}{2}\left[\frac{2q\varepsilon}{(V_0 - V)}\frac{N_d N_a}{N_d + N_a}\right]^{1/2} = \frac{\varepsilon A}{W}$$

$$C_j = \frac{A}{2}\left[\frac{2q\varepsilon}{V_0 - V}N_d\right]^{1/2}$$

For p^+-n junction ($N_a >> N_d$)



Note this is for p^+-n junction, so the depletion in the p+ side is neglected

For the diode of Example before, what is the total depletion capacitance at –4 V?

$$C_j = \sqrt{\epsilon A}\left[\frac{q}{2(V_0 - V)}\frac{N_d N_a}{N_d + N_a}\right]^{1/2}$$

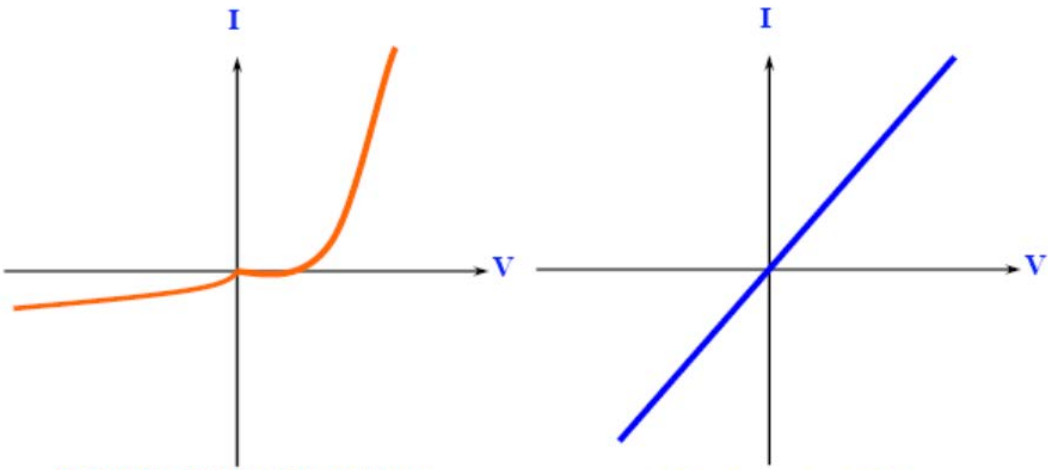
$$= \sqrt{(8.85 \times 10^{-14} \times 11.8)(10^{-4})}\left[\frac{1.6 \times 10^{-19}}{2(0.695 + 4)}\left(\frac{10^{15} \times 10^{17}}{10^{15} + 10^{17}}\right)\right]^{1/2}$$

$$= \mathbf{4.198 \times 10^{-13}\ F}$$

METAL-SEMICONDUCTOR JUNCTIONS

• In solid-state physics, a **metal–semiconductor (M–S) junction** is a type of junction in which a metal comes in close contact with a semiconductor material.

• M–S junctions can either be rectifying or non-rectifying. The rectifying metal–semiconductor junction forms a ***Schottky barrier***, making a device known as a Schottky diode, while the non-rectifying junction is called an ***ohmic contact***.



Schottky I-V just look like a pn diode I-V.

Ohmic contact I-V.

When you put a metal interconnect to a semiconductor, you need to make an ohmic contact unless you need a Schottky diode.

Schottky junction (Diode)

When a metal with work function $q\Phi_m$ is brought in contact with a semiconductor having a work function $q\Phi_s$, charge transfer occurs until the Fermi levels align at equilibrium

Work function: energy required to remove electron at fermi level to vacuum.

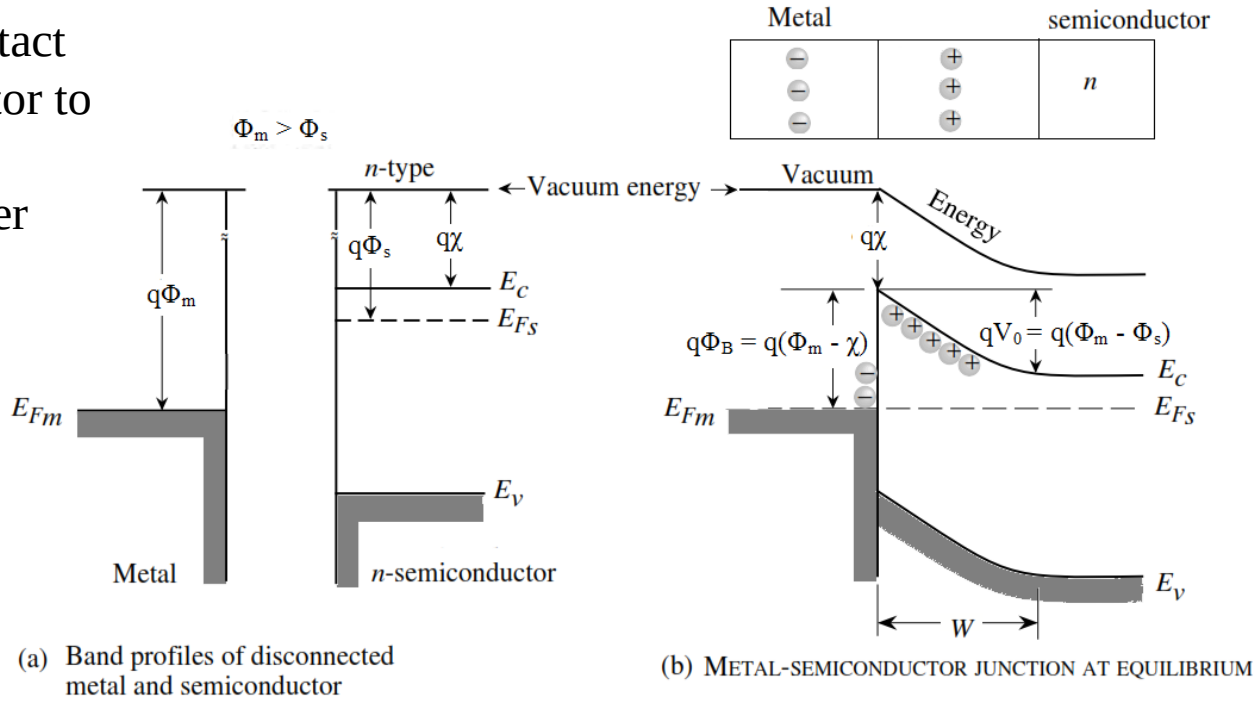
Example: if we have contact of metal with n-type semiconductor and the potential $\Phi_m > \Phi_s$ → To have equilibrium, E_{Fs} need to be lowered → electron flow from semiconductor to metal → depletion region (W) is formed near the junction in the n-region ($W \approx x_{n0}$) with contact potential V_0 from semiconductor to metal. From metal to semiconductor, Schottky barrier $q\Phi_B$ is formed.

Using p+-n junction ($N_a >> N_d$) approximation →

$$W = \left[\frac{2\varepsilon V_0}{qN_d}\right]^{1/2}$$

and junction capacitance

$$C_j = \frac{A}{2}\left[\frac{2q\varepsilon}{V_0 - V}N_d\right]^{1/2} = \frac{\varepsilon A}{W}$$



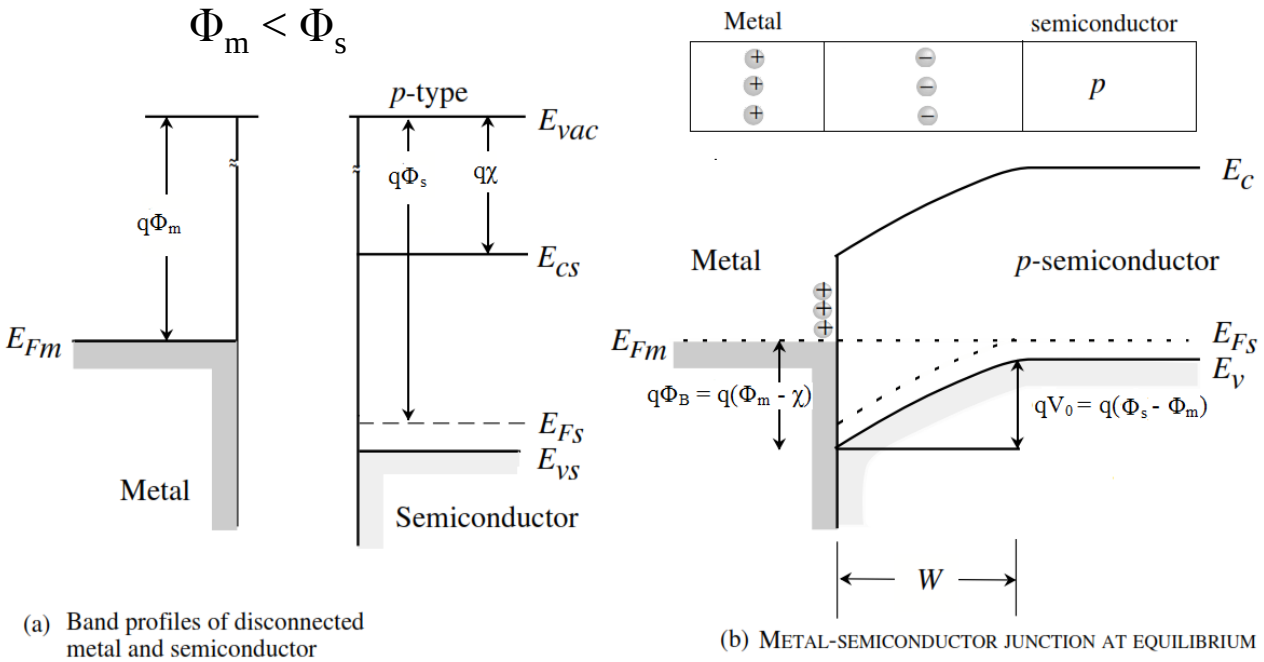
(a) Band profiles of disconnected metal and semiconductor

(b) METAL-SEMICONDUCTOR JUNCTION AT EQUILIBRIUM

Φ_m: metal work function
Φ_s: Semiconductor work function
χ : electron affinity
Φ_B: Schottky Barrier

Schottky junction (Diode)

If we have p-type semiconductor in contact with metal and $\Phi_m < \Phi_s$ → electrons will move from metal to semiconductor → this will lead to formation of negative charge in the depletion region and contact potential V_0 witch will not allow holes to move from semiconductor the metal and Φ_B from metal to semiconductor that will not allow holes to flow from metal to semiconductor at equilibrium

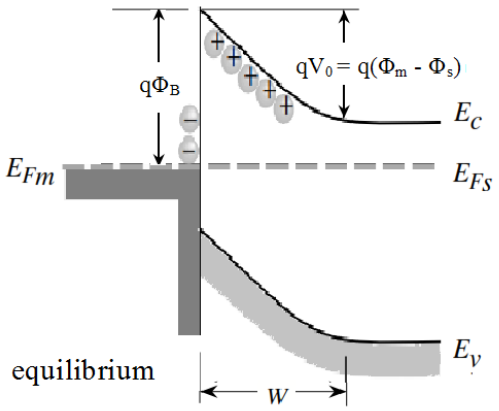


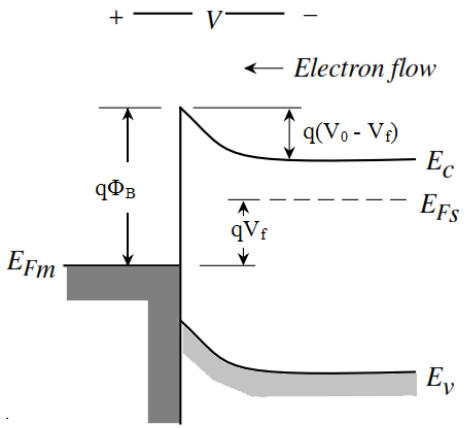
(a) Band profiles of disconnected metal and semiconductor

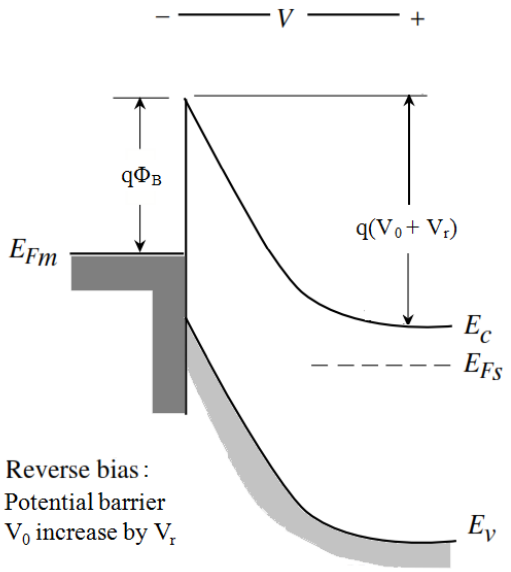
(b) METAL-SEMICONDUCTOR JUNCTION AT EQUILIBRIUM

Schottky junction (Diode)

Rectifying Schottky junction is the junction which allow current to pass only in one direction (forward bias). Energy diagram at equilibrium, forward and reverse bias for n-type Schottky diode is shown below.







Diode with area A:

$$I = I_s\left[\exp\left(\frac{qV}{k_BT}\right) - 1\right]$$

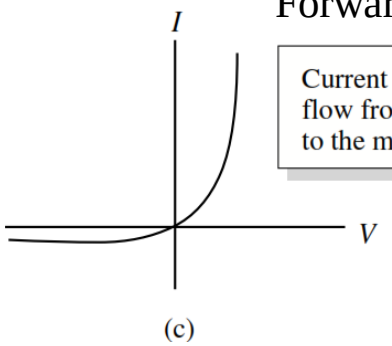
$$I_s = A\left(\frac{m^* q k_B^2}{2\pi^2 h^3}\right) T^2 \exp\left(\frac{-q\Phi_B}{k_BT}\right)$$

$$= A\,R^*\,T^2 \exp\left(\frac{-q\Phi_B}{k_BT}\right)$$

Richardson constant: $R^* = 120\ \frac{m^*}{m_o}\ \text{Acm}^{-2}\text{K}^{-2}$

Reverse bias current

Current dominated by electron flow from the metal to the semiconductor



Forward bias current

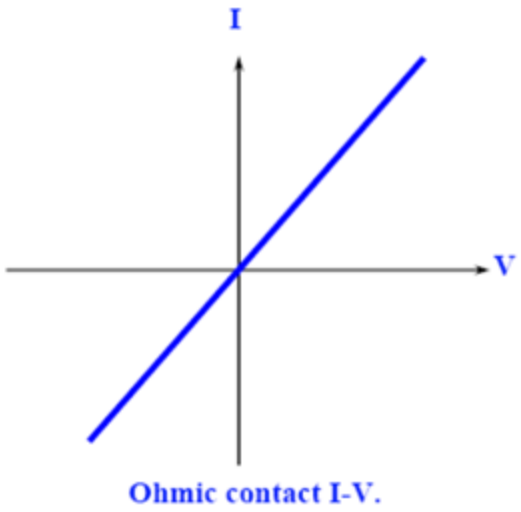
Current dominated by electron flow from the semiconductor to the metal

Ohmic contact (non-rectifying)

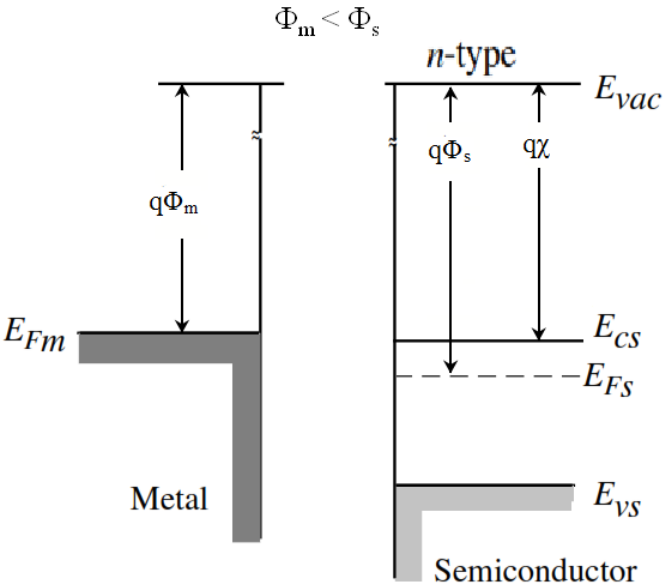
In semiconductor device applications, we need to have semiconducting material connected to outside by metal wires → we need non-rectifying junction (not like previously discussed Schottky diode) → linear IV relationship→ ohmic relationship

Example: if we have contact of metal with n-type semiconductor and the potential $\Phi_m < \Phi_s$ → Fermi levels are aligned by transition of electrons from metal to semiconductor → very small barrier is created at equilibrium → easily overcome the barrier at low voltage values.

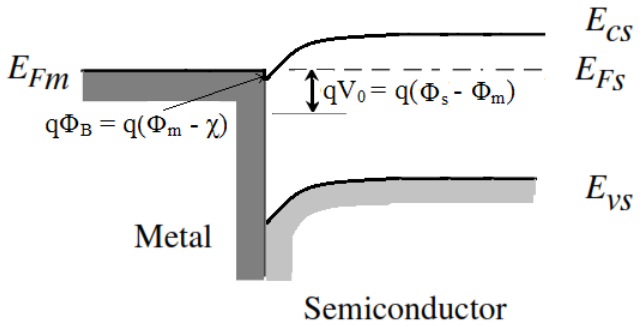
The current flow easily with applied potential in the two directions in an ohmic linear relationship V = IR



Ohmic contact I-V.



(a) Band profiles of disconnected metal and semiconductor



(b) METAL-SEMICONDUCTOR JUNCTION AT EQUILIBRIUM

Similarly contact of metal with p-type semiconductor with $\Phi_m > \Phi_s$ → formation of ohmic contact