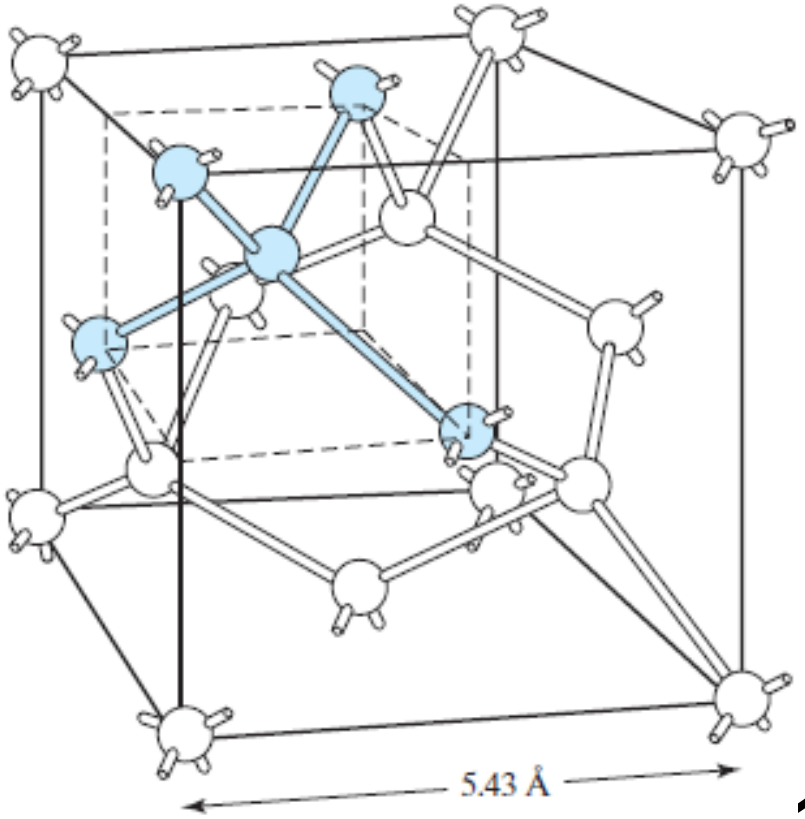
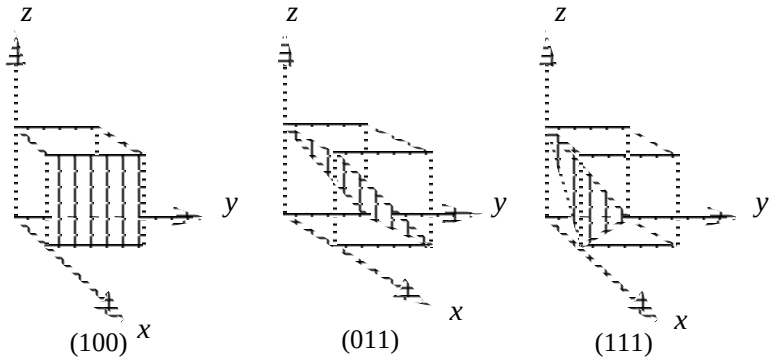
Chapter 1 Electrons and Holes in Semiconductors

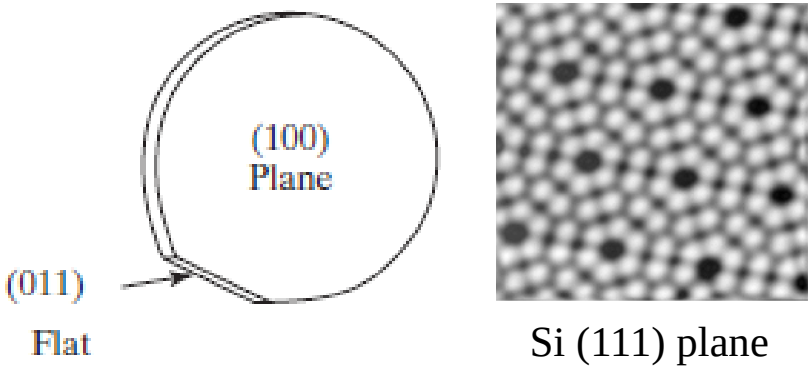
1.1 Silicon Crystal Structure

- ***Unit cell*** of silicon crystal is cubic.
- ***Each Si atom has 4 nearest neighbors.***



Silicon Wafers and Crystal Planes

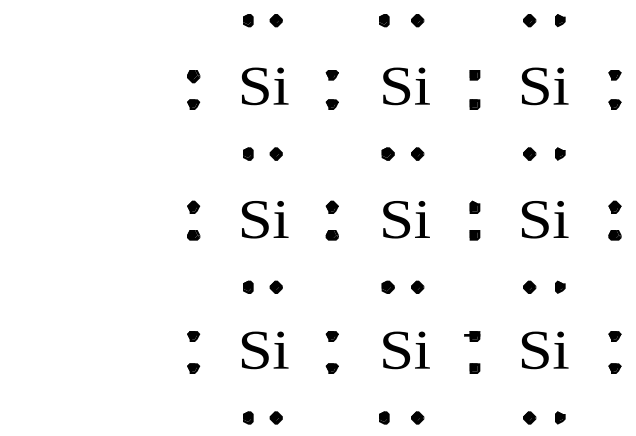




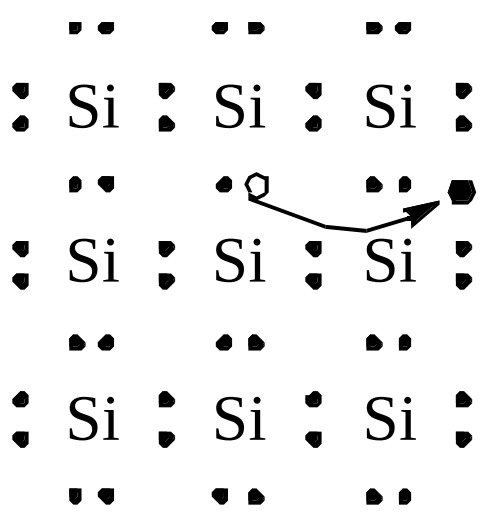
Si (111) plane

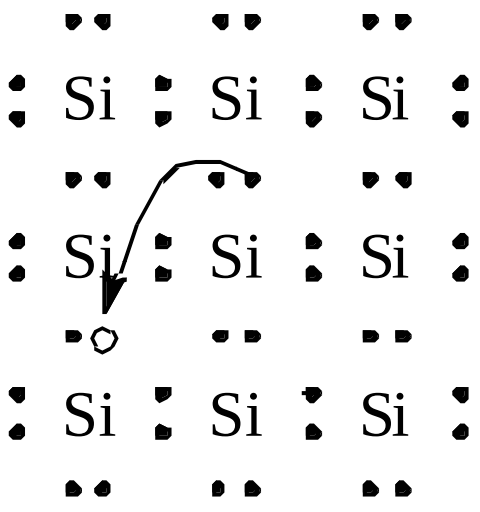
- The standard notation for crystal planes is based on the cubic unit cell.
- Silicon wafers are usually cut along the (100) plane with a flat or notch to help orient the wafer during IC fabrication.

1.2 Bond Model of Electrons and Holes



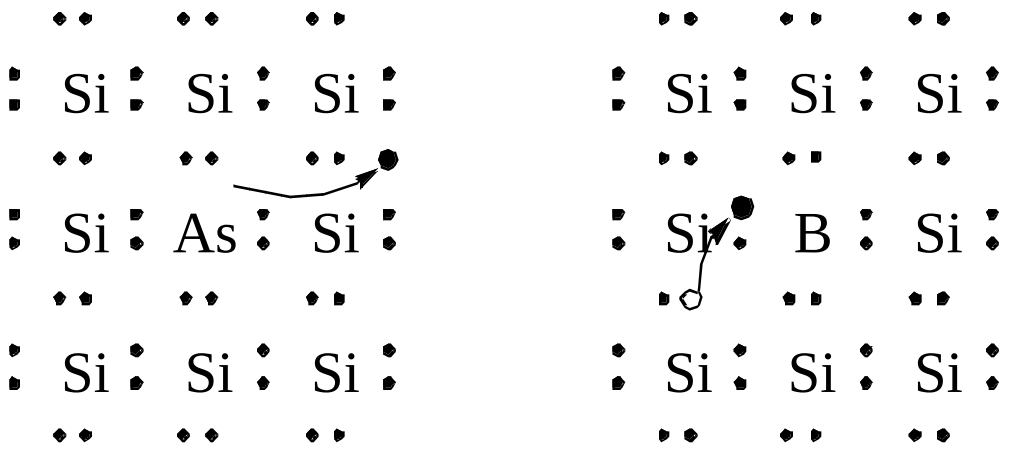
- Silicon crystal in a two-dimensional representation.





- When an electron breaks loose and becomes a ***conduction electron***, a ***hole*** is also created.

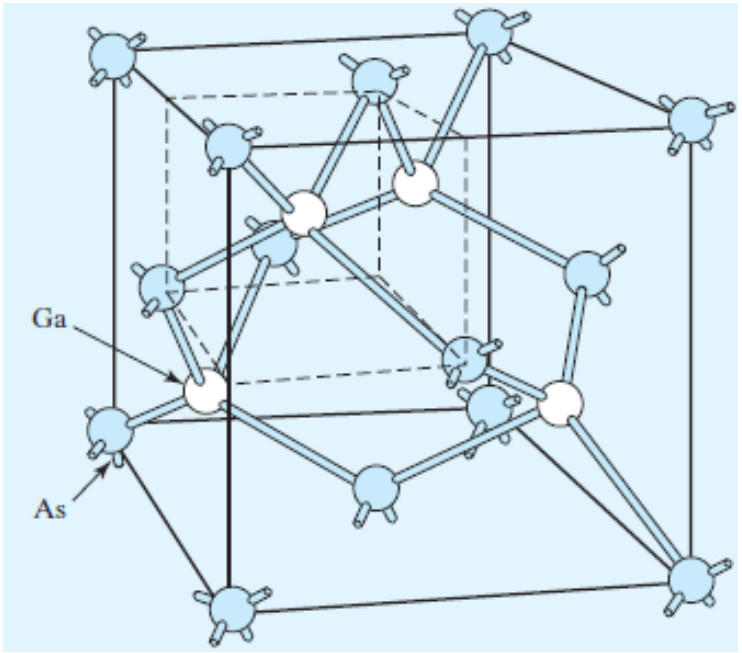
Dopants in Silicon

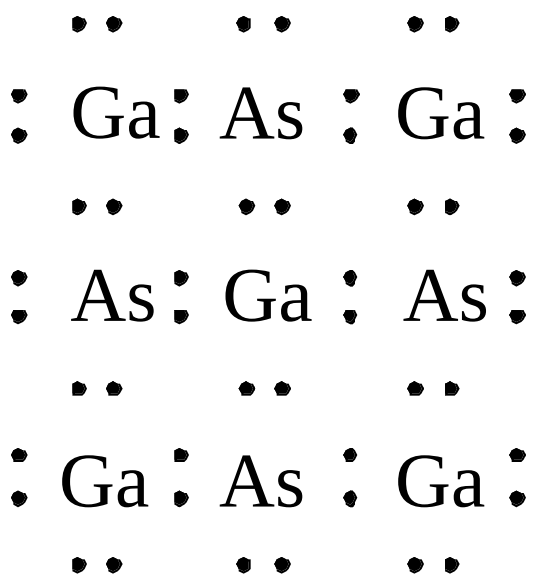


- As, a Group V element, introduces conduction electrons and creates ***N-type silicon,*** and is called a ***donor.***
- B, a Group III element, introduces holes and creates ***P-type silicon***, and is called an ***acceptor***.
- Donors and acceptors are known as dopants. Dopant ionization energy ~50meV (very low).

Hydrogen: $E_{ion} = \dfrac{m_0 q^4}{8\varepsilon_0^2 h^2} = 13.6 \text{ eV}$

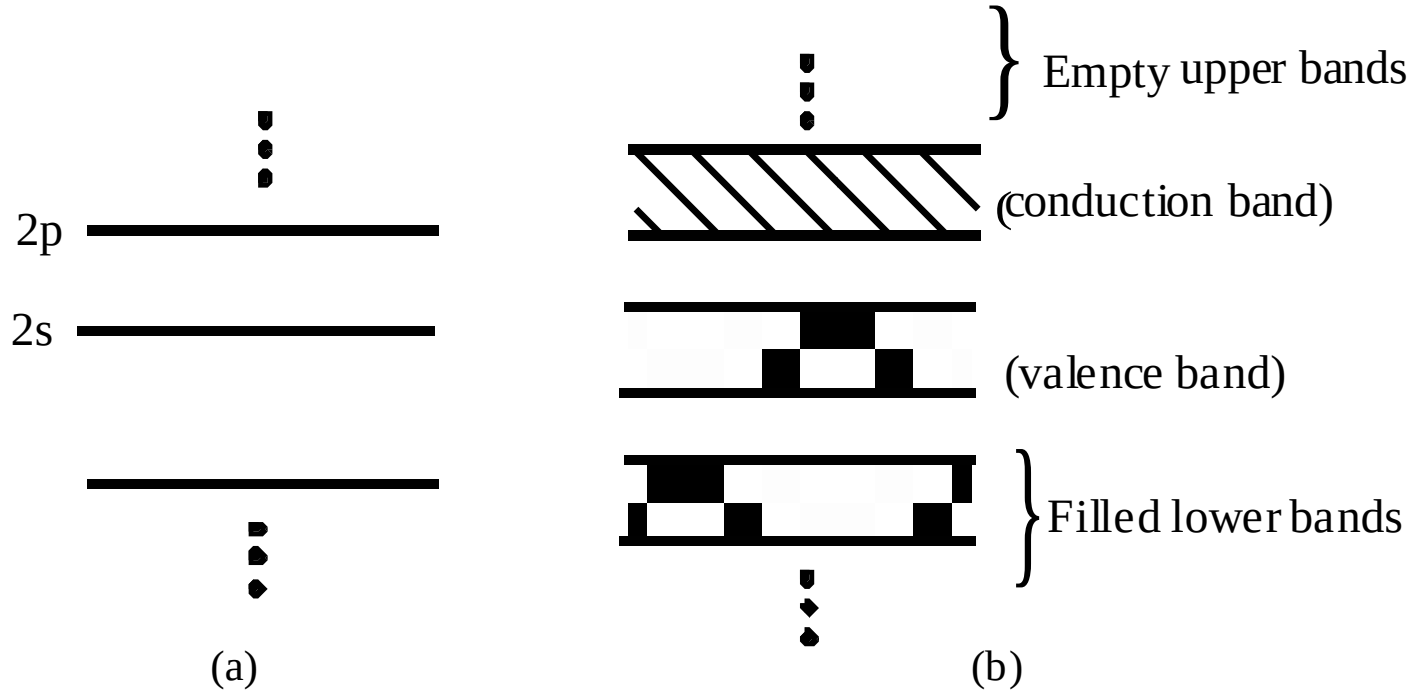
GaAs, III-V Compound Semiconductors, and Their Dopants





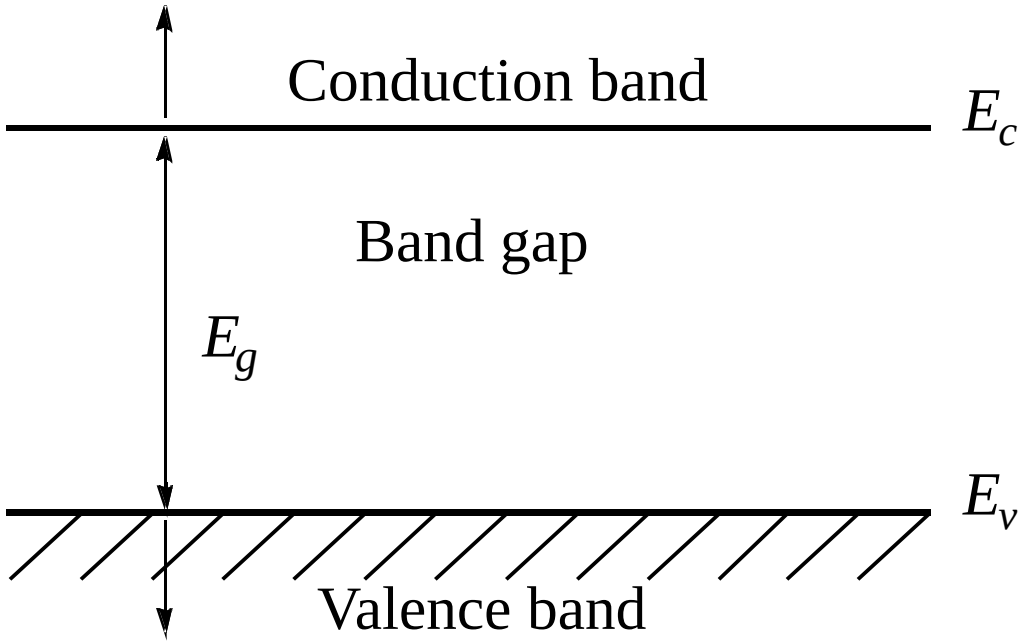
- GaAs has the same crystal structure as Si.
- GaAs, GaP, GaN are III-V compound semiconductors, important for optoelectronics.
- Wich group of elements are candidates for donors? acceptors?

1.3 Energy Band Model



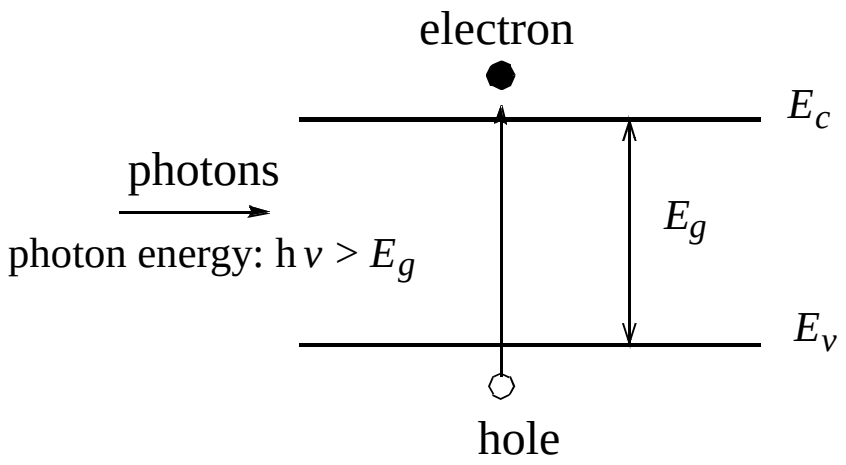
- Energy states of Si atom (a) expand into energy bands of Si crystal (b).
- The lower bands are filled and higher bands are empty in a semiconductor.
- The highest filled band is the ***valence band.***
- The lowest empty band is the ***conduction band*** .

1.3.1 Energy Band Diagram



- ***Energy band diagram*** shows the bottom edge of conduction band, E_c , and top edge of valence band, E_v .
- E_c and E_v are separated by the ***band gap energy, E_g*** .

Measuring the Band Gap Energy by Light Absorption

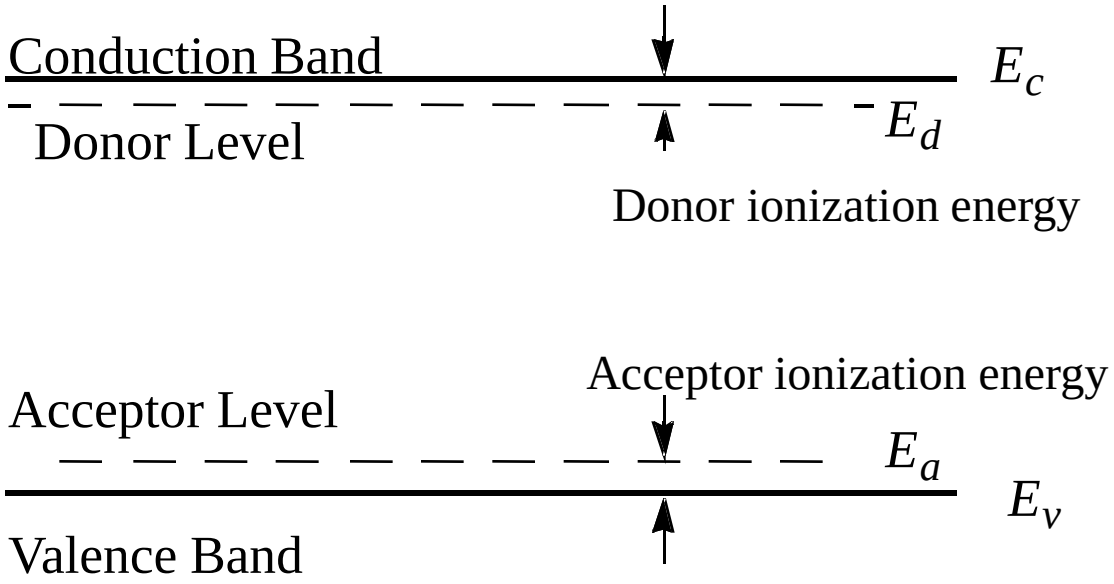


• E_g can be determined from the minimum energy ($h\nu$) of photons that are absorbed by the semiconductor.

Bandgap energies of selected semiconductors

Semi-conductor	InSb	Ge	Si	GaAs	GaP	ZnSe	Diamond
Eg (eV)	0.18	0.67	1.12	1.42	2.25	2.7	6

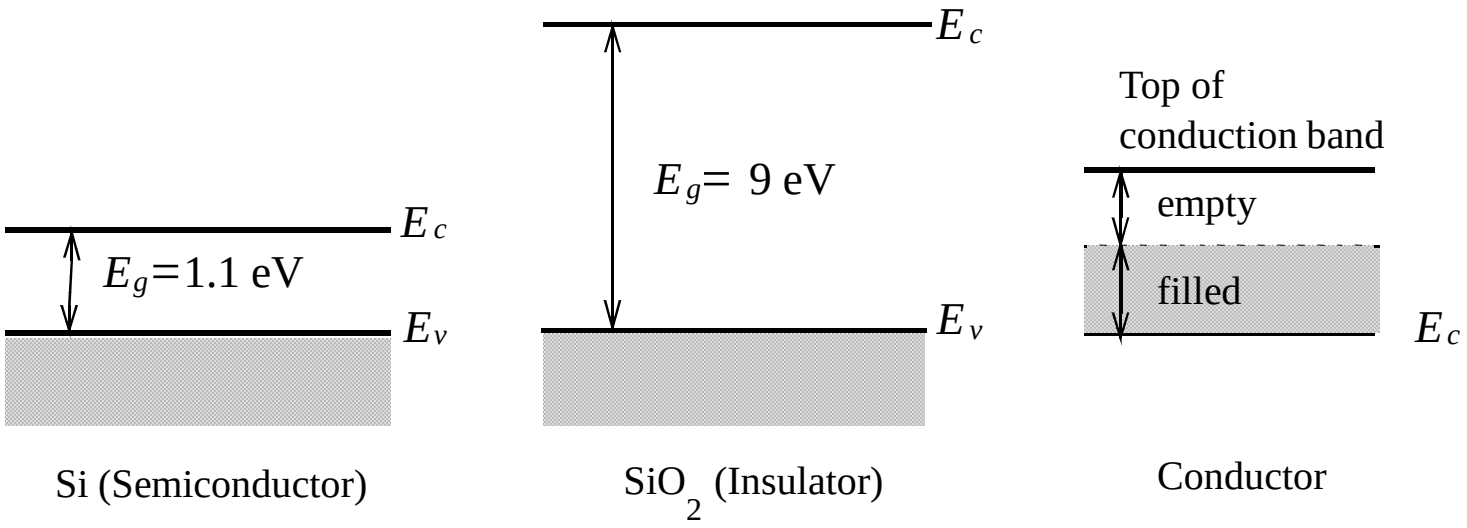
1.3.2 Donor and Acceptor in the Band Model



Ionization energy of selected donors and acceptors in silicon

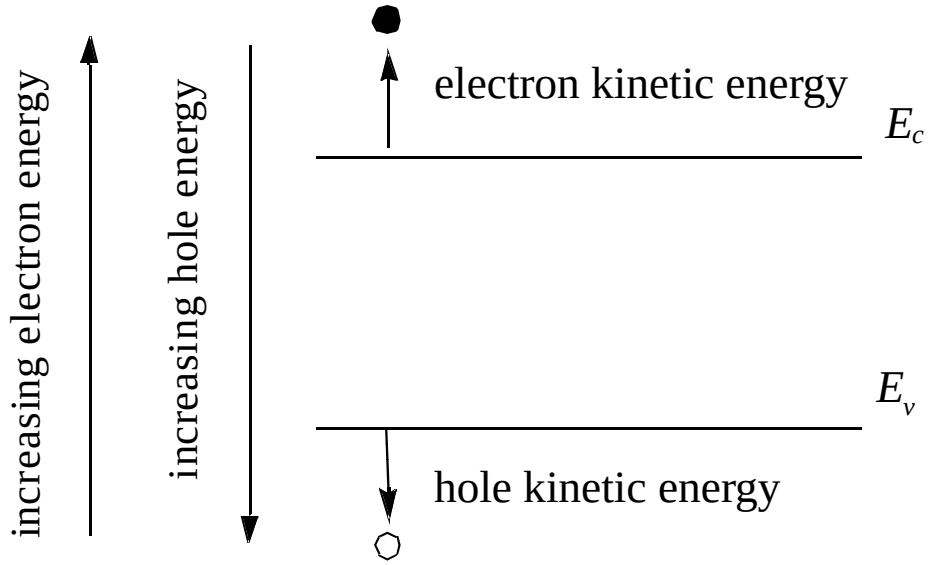
	Donors			Acceptors		
Dopant	Sb	P	As	B	Al	In
Ionization energy, $E_c - E_d$ or $E_a - E_v$ (meV)	39	44	54	45	57	160

1.4 Semiconductors, Insulators, and Conductors



- Totally filled bands and totally empty bands do not allow current flow. (Just as there is no motion of liquid in a totally filled or totally empty bottle.)
- Metal conduction band is half-filled.
- Semiconductors have lower E_g's than insulators and can be doped.

1.5 Electrons and Holes



- Both electrons and holes tend to seek their lowest energy positions.
- Electrons tend to fall in the energy band diagram.
- Holes float up like bubbles in water.

1.5.1 Effective Mass

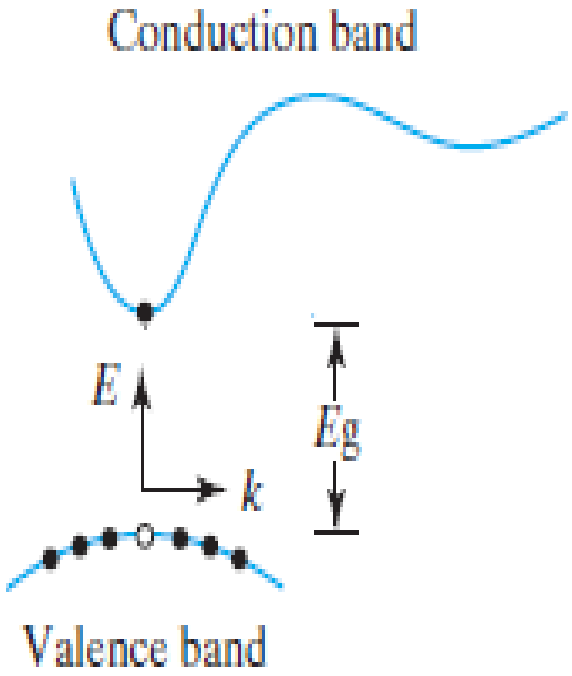
The electron wave function is the solution of the three dimensional Schrodinger wave equation

$$-\frac{\hbar^2}{2m_0}\nabla^2\psi + V(r)\psi = \psi$$

The solution is of the form $\exp(\pm \mathbf{k}\cdot\mathbf{r})$

k = wave vector = 2π/electron wavelength

For each k, there is a corresponding E.



$$\text{acceleration} = -\frac{q\varepsilon}{\hbar^2}\frac{d^2E}{dk^2} = \frac{F}{m}$$

$$\text{effective mass} \equiv \frac{\hbar^2}{d^2E/dk^2}$$

1.5.1 Effective Mass

In an electric field, **E**, an electron or a hole accelerates.

$$a = \frac{-q\mathcal{E}}{m_n} \qquad \text{electrons}$$

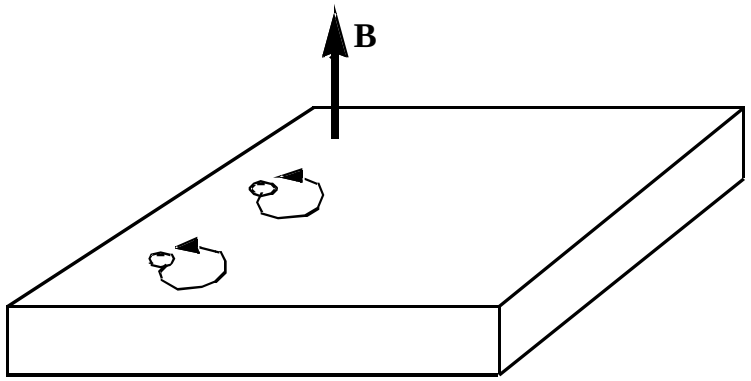
$$a = \frac{q\mathcal{E}}{m_p} \qquad \text{holes}$$

Electron and hole effective masses

	Si	Ge	GaAs	InAs	AlAs
m_n/m_0	0.26	0.12	0.068	0.023	2
m_p/m_0	0.39	0.3	0.5	0.3	0.3

1.5.2 How to Measure the Effective Mass

Cyclotron Resonance Technique

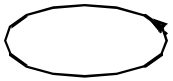


Centripetal force = Lorentzian force

$$\frac{m_n v^2}{r} = qvB$$

$$v = \frac{qBr}{m_n}$$

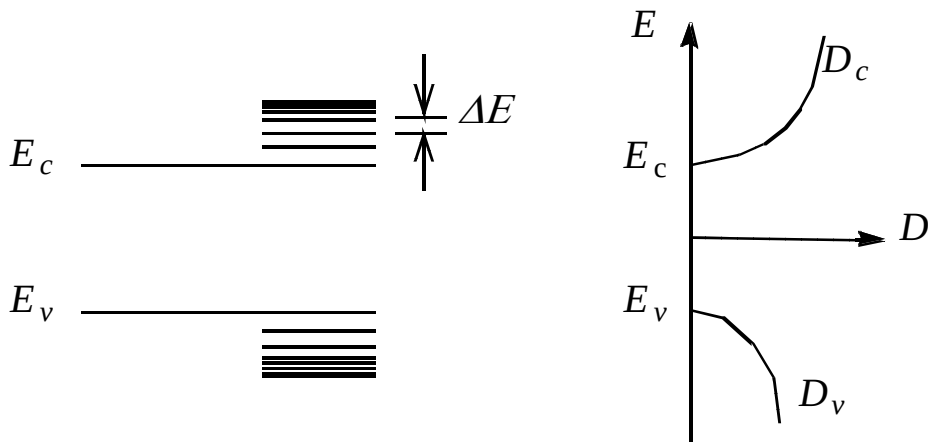
$$f_{cr} = \frac{v}{2\pi r} = \frac{qB}{2\pi m_n}$$



Microwave

- f_{cr} is the Cyclotron resonance frequency.
- It is independent of v and r.
- Electrons strongly absorb microwaves of that frequency.
- By measuring f_{cr}, m_n can be found.

1.6 Density of States



$$D_c(E) \equiv \frac{\text{number of states in } \Delta E}{\Delta E \cdot \text{volume}} \quad \left(\frac{1}{\text{eV} \cdot \text{cm}^3} \right)$$

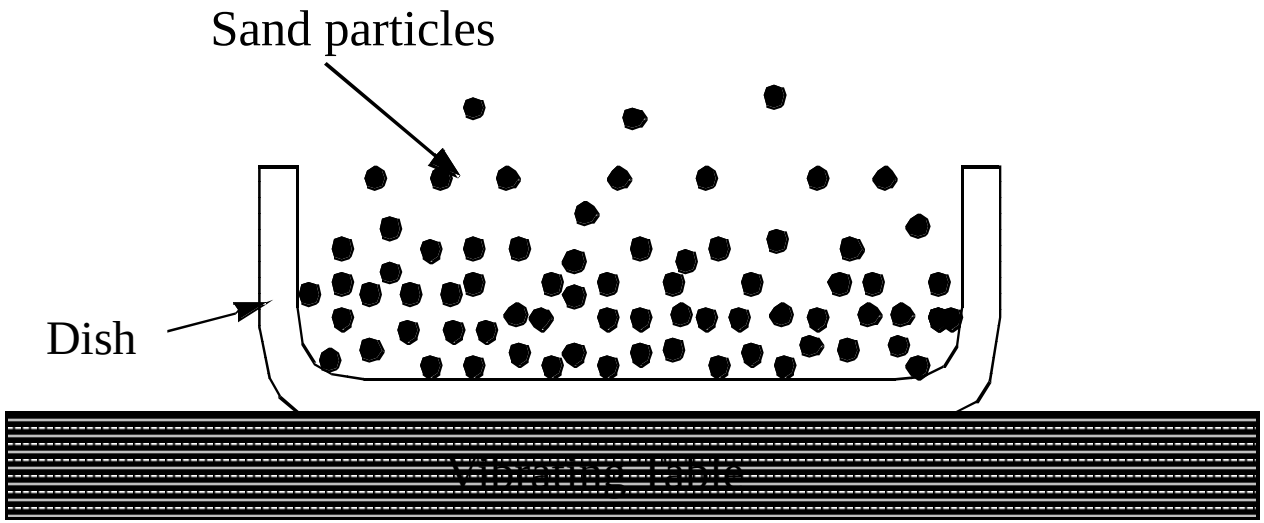
$$D_c(E) \equiv \frac{8\pi m_n \sqrt{2m_n (E - E_c)}}{h^3}$$

$$D_v(E) \equiv \frac{8\pi m_p \sqrt{2m_p (E_v - E)}}{h^3}$$

Derived in Appendix I

1.7 Thermal Equilibrium and the Fermi Function

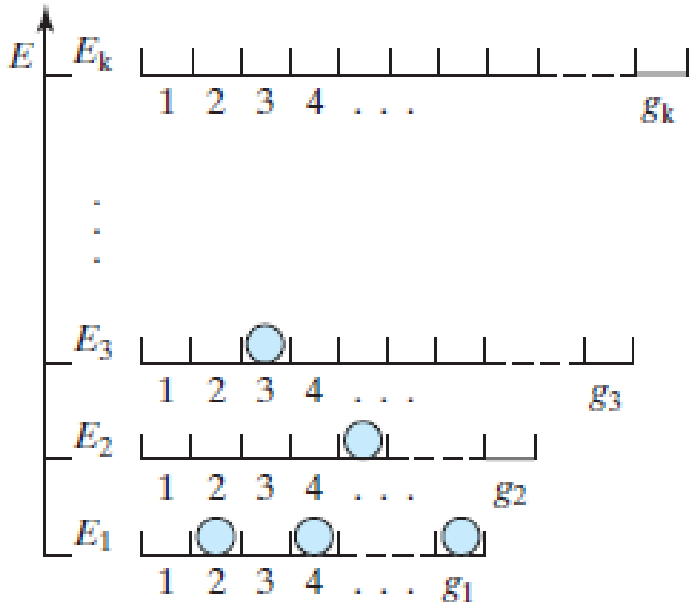
1.7.1 An Analogy for Thermal Equilibrium



- There is a certain probability for the electrons in the conduction band to occupy high-energy states under the agitation of thermal energy.

Appendix II. Probability of a State at E being Occupied

- There are g_1 states at E_1, g_2 states at E_2… There are N electrons, which constantly shift among all the states but the average electron energy is fixed at 3kT/2.
- There are many ways to distribute N among n_1, n_2, n_3….and satisfy the 3kT/2 condition.



- The equilibrium distribution is the distribution that maximizes the number of combinations of placing n_1 in g_1 slots, n_2 in g_2 slots…. :

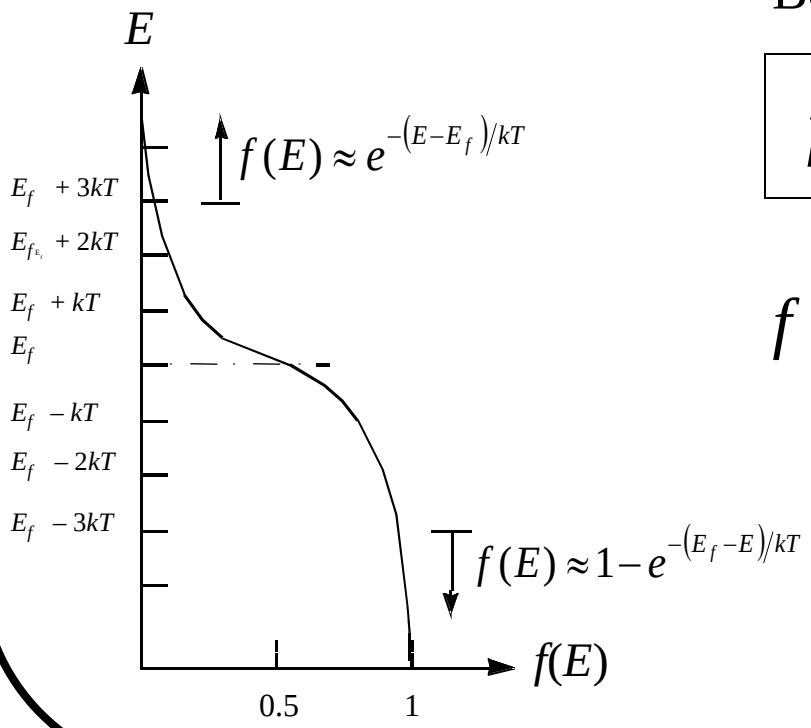
$$n_i/g_i = \frac{1}{1 + e^{(E - E_F)/kT}}$$

E_F is a constant determined by the condition $\sum n_i = N$

1.7.2 Fermi Function–The Probability of an Energy State Being Occupied by an Electron

$$f(E) = \frac{1}{1 + e^{(E - E_f)/kT}}$$

E_f is called the ***Fermi energy*** or the ***Fermi level.***



Boltzmann approximation:

$$f(E) \approx e^{-(E - E_f)/kT} \qquad E - E_f >> kT$$

$$f(E) \approx 1 - e^{-(E_f - E)/kT} \qquad E - E_f << -kT$$

Remember: there is only one Fermi-level in a system at equilibrium.

1.8 Electron and Hole Concentrations

1.8.1 Derivation of *n* and *p* from *D*(*E*) and *f*(*E*)

$$n = \int_{E_c}^{\text{top of conduction band}} f(E) D_c(E) dE$$

$$n = \frac{8\pi m_n \sqrt{2m_n}}{h^3} \int_{E_c}^{\infty} \sqrt{E - E_c}\, e^{-\left(E - E_f\right)/kT} dE$$

$$= \frac{8\pi m_n \sqrt{2m_n}}{h^3} e^{-\left(E_c - E_f\right)/kT} \int_0^{E-Ec} \sqrt{E - E_c}\, e^{-\left(E - E_c\right)/kT} d(E - Ec)$$

Electron and Hole Concentrations

$$n = N_c e^{-(E_c - E_f)/kT}$$

$$N_c \equiv 2\left[\frac{2\pi m_n kT}{h^2}\right]^{3/2}$$

N_c is called the ***effective density of states (of the conduction band) .***

$$p = N_v e^{-(E_f - E_v)/kT}$$

$$N_v \equiv 2\left[\frac{2\pi m_p kT}{h^2}\right]^{3/2}$$

N_v is called the ***effective density of states of the valence band.***

Remember: the closer E_f moves up to N_c, the larger *n* is; the closer E_f moves down to E_v , the larger *p* is.

For Si, $N_c = 2.8 \times 10^{19}\,cm^{-3}$ and $N_v = 1.04 \times 10^{19}\,cm^{-3}$.

1.8.2 The Fermi Level and Carrier Concentrations

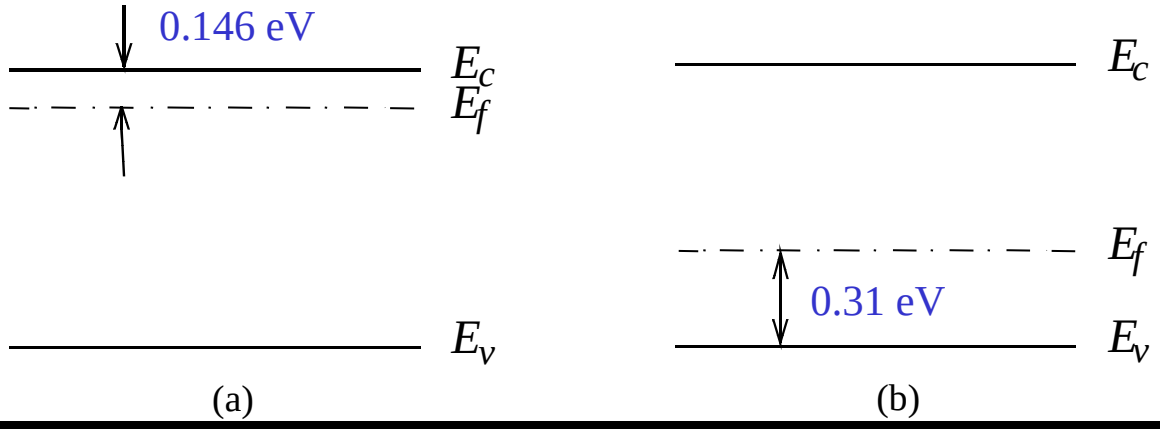
Where is E_f for $n = 10^{17}\,cm^{-3}$? And for $p = 10^{14}\,cm^{-3}$?

***Solution**: (a)* $n = N_c e^{-(E_c - E_f)/kT}$

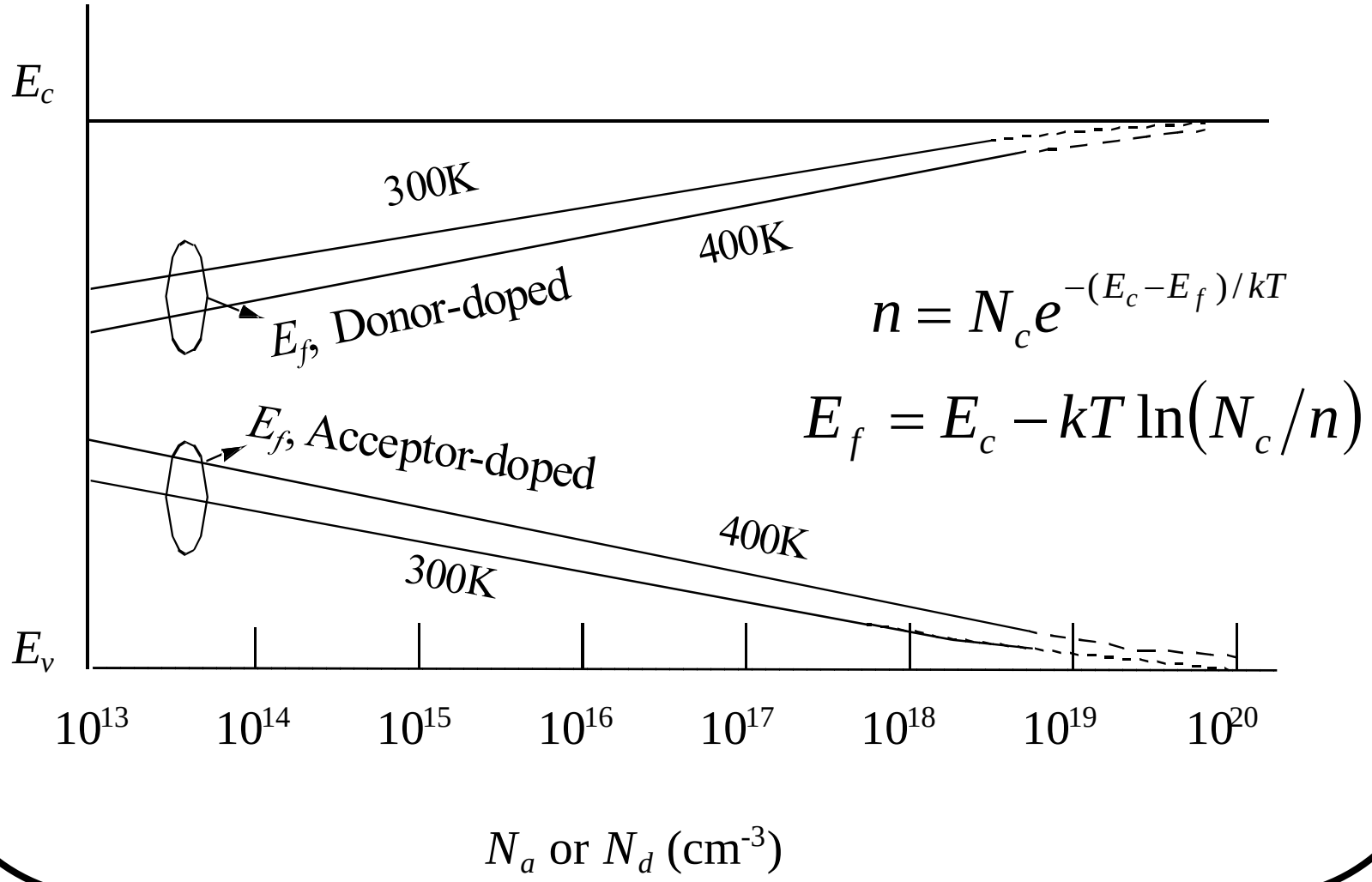
$$E_c - E_f = kT \ln\left(N_c/n\right) = 0.026 \ln\left(2.8\times10^{19}/10^{17}\right) = 0.146\,\text{eV}$$

(b) For $p = 10^{14}cm^{-3}$, from Eq.(1.8.8),

$$E_f - E_v = kT \ln\left(N_v/p\right) = 0.026 \ln\left(1.04\times10^{19}/10^{14}\right) = 0.31\,\text{eV}$$



1.8.2 The Fermi Level and Carrier Concentrations



$$n = N_c e^{-(E_c - E_f)/kT}$$

$$E_f = E_c - kT \ln\left(N_c / n\right)$$

1.8.3 The *np* Product and the Intrinsic Carrier Concentration

Multiply $n = N_c e^{-(E_c - E_f)/kT}$ and $p = N_v e^{-(E_f - E_v)/kT}$

$$np = N_c N_v e^{-(E_c - E_v)/kT} = N_c N_v e^{-E_g/kT}$$

$$\boxed{np = n_i^2}$$

$$\boxed{n_i = \sqrt{N_c N_v}\, e^{-E_g/2kT}}$$

- In an intrinsic (undoped) semiconductor, $n = p = n_i$.
- n_i is the ***intrinsic carrier concentration***, ~10^{10} cm^{-3} for Si.

EXAMPLE: Carrier Concentrations

Question: *What is the hole concentration in an N-type semiconductor with 10^{15} cm^{-3} of donors?*

Solution: *$n = 10^{15}$ cm^{-3}.*

$$p = \frac{n_i^{\,2}}{n} \approx \frac{10^{20}\,\mathrm{cm}^{-3}}{10^{15}\,\mathrm{cm}^{-3}} = 10^{5}\,\mathrm{cm}^{-3}$$

After increasing T by 60 ℃, n remains the same at 10^{15} cm^{-3} while p increases by about a factor of 2300 because $n_i^{\,2} \propto e^{-E_g/kT}$.

Question: *What is n if $p = 10^{17} cm^{-3}$ in a P-type silicon wafer?*

Solution:

$$n = \frac{n_i^{\,2}}{p} \approx \frac{10^{20}\,\mathrm{cm}^{-3}}{10^{17}\,\mathrm{cm}^{-3}} = 10^{3}\,\mathrm{cm}^{-3}$$

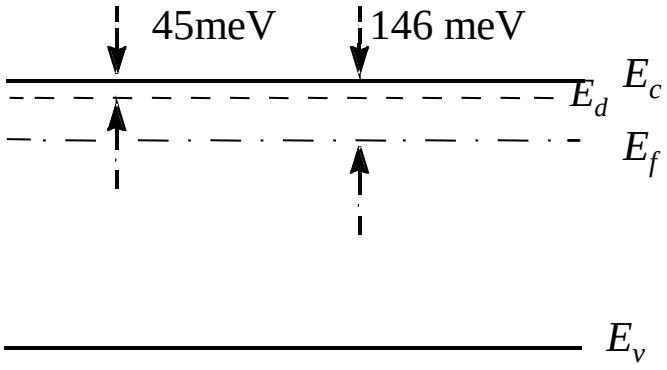
1.9 General Theory of n and p

EXAMPLE: Complete ionization of the dopant atoms

$N_d = 10^{17}$ cm^{-3}. What fraction of the donors are not ionized?

Solution: *First assume that all the donors* ***are*** *ionized.*

$$n = N_d = 10^{17}\,\text{cm}^{-3} \Rightarrow E_f = E_c - 146\text{meV}$$



Probability of not being ionized $\approx \dfrac{1}{1+\frac{1}{2}e^{(E_d - E_f)/kT}} = \dfrac{1}{1+\frac{1}{2}e^{((146-45)\text{meV})/26\text{meV}}} = 0.04$

Therefore, it is reasonable to assume complete ionization, i.e., $n = N_d$.

1.9 General Theory of n and p

Charge neutrality: $n + N_a = p + N_d$

$$np = n_i^2$$

$$p = \frac{N_a - N_d}{2} + \left[\left(\frac{N_a - N_d}{2} \right)^2 + n_i^2 \right]^{1/2}$$

$$n = \frac{N_d - N_a}{2} + \left[\left(\frac{N_d - N_a}{2} \right)^2 + n_i^2 \right]^{1/2}$$

1.9 General Theory of on n and p

I. $N_d - N_a >> n_i$ (i.e., N-type)

$$n = N_d - N_a$$
$$p = n_i^2 / n$$

If $N_d >> N_a$, $n = N_d$ and $p = n_i^2 / N_d$

II. $N_a - N_d >> n_i$ (i.e., P-type)

$$p = N_a - N_d$$
$$n = n_i^2 / p$$

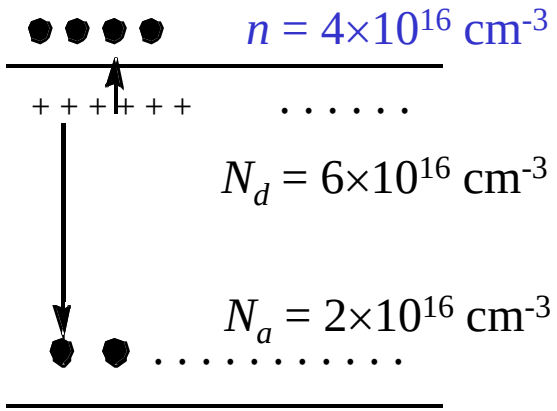
If $N_a >> N_d$, $p = N_a$ and $n = n_i^2 / N_a$

EXAMPLE: Dopant Compensation

What are n and p in Si with (a) $N_d = 6\times10^{16}$ cm^{-3} and $N_a = 2\times10^{16}$ cm^{-3} and (b) additional 6×10^{16} cm^{-3} of N_a?

(a) $n = N_d - N_a = 4\times10^{16}\,\text{cm}^{-3}$

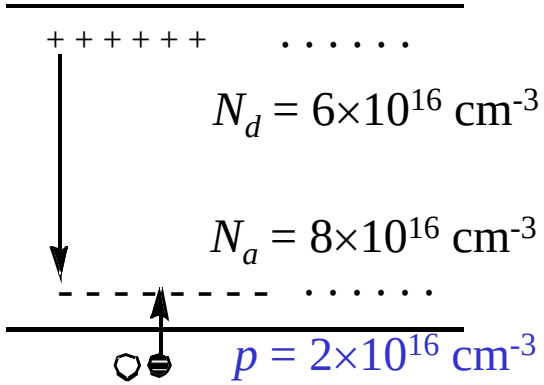
$p = n_i^{\,2} / n = 10^{20} / 4\times10^{16} = 2.5\times10^{3}\,\text{cm}^{-3}$



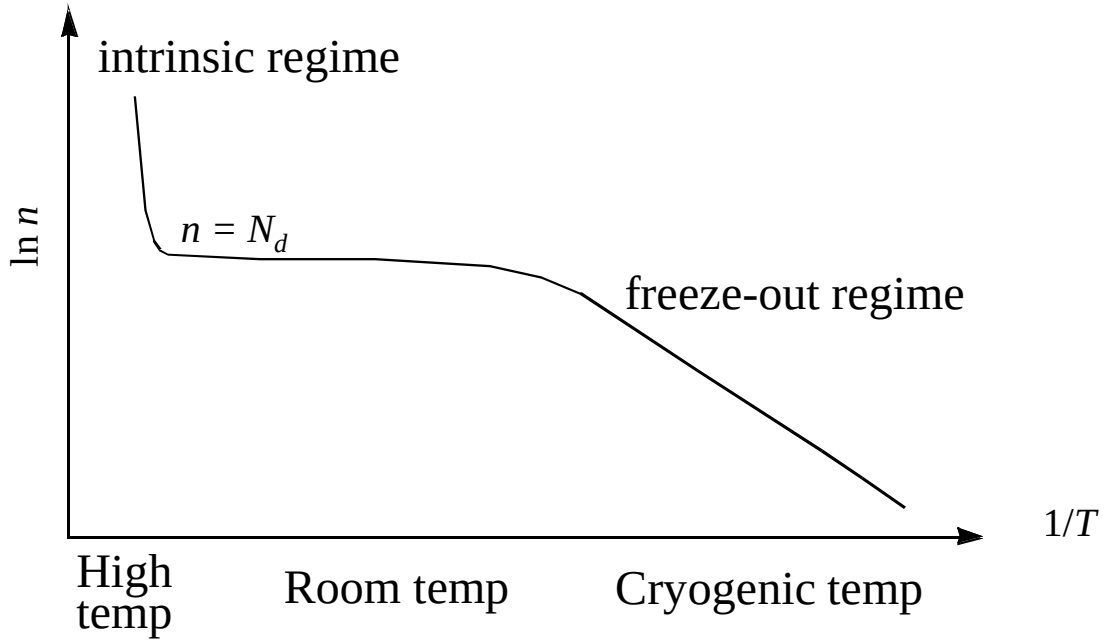
(b) $N_a = 2\times10^{16} + 6\times10^{16} = 8\times10^{16}\,\text{cm}^{-3} > N_d$

$p = N_a - N_d = 8\times10^{16} - 6\times10^{16} = 2\times10^{16}\,\text{cm}^{-3}$

$n = n_i^{\,2} / p = 10^{20} / 2\times10^{16} = 5\times10^{3}\,\text{cm}^{-3}$



1.10 Carrier Concentrations at Extremely High and Low Temperatures



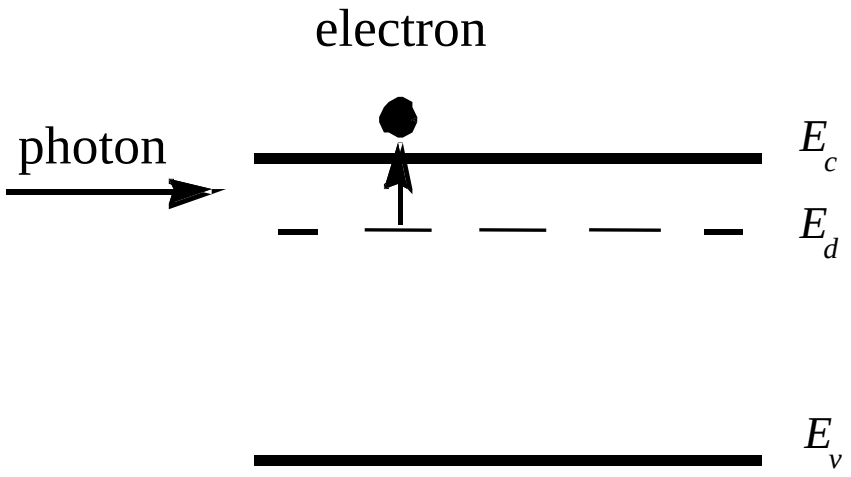
high T: $n = p = n_i = \sqrt{N_c N_v}\, e^{-E_g/2kT}$

low T: $n = \left[\frac{N_c N_d}{2}\right]^{1/2} e^{-(E_c - E_d)/2kT}$

Infrared Detector Based on Freeze-out

•To image the black-body radiation emitted by tumors requires a photodetector that responds to $h\nu$'s around 0.1 eV.

•In doped Si operating in the freeze-out mode, conduction electrons are created when the infrared photons provide the energy to ionized the donor atoms.



1.11 Chapter Summary

Energy band diagram. Acceptor. Donor. m_n, m_p. Fermi function. E_f .

$$n = N_c e^{-(E_c - E_f)/kT}$$

$$p = N_v e^{-(E_f - E_v)/kT}$$

$$n = N_d - N_a$$

$$p = N_a - N_d$$

$$np = n_i^2$$

Chapter 2 Motion and Recombination of Electrons and Holes

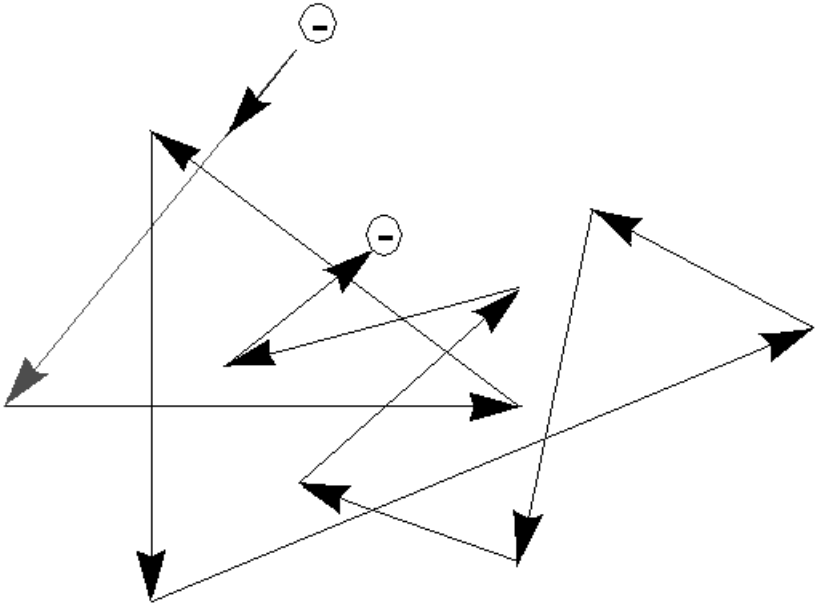
2.1 Thermal Motion

Average electron or hole kinetic energy $= \frac{3}{2}kT = \frac{1}{2}mv_{th}^2$

$$v_{th} = \sqrt{\frac{3kT}{m_{eff}}} = \sqrt{\frac{3\times 1.38\times 10^{-23}\,\mathrm{JK^{-1}}\times 300\mathrm{K}}{0.26\times 9.1\times 10^{-31}\,\mathrm{kg}}}$$

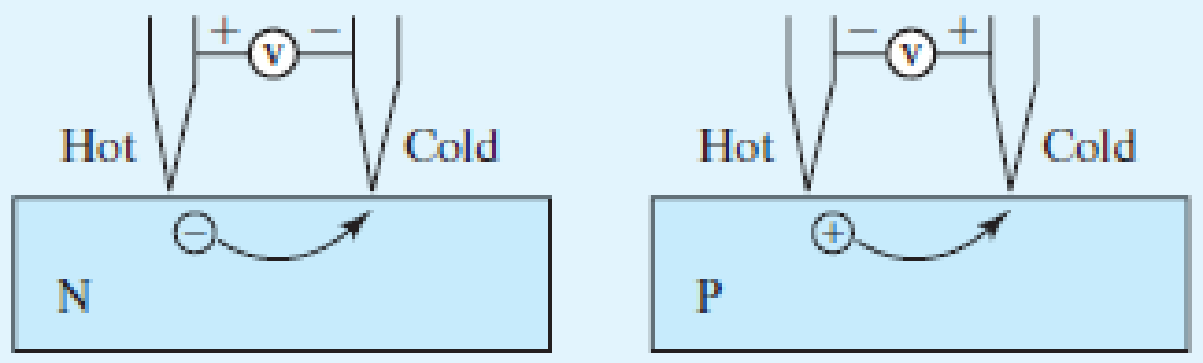
$$= 2.3\times 10^{5}\,\mathrm{m/s} = 2.3\times 10^{7}\,\mathrm{cm/s}$$

2.1 Thermal Motion



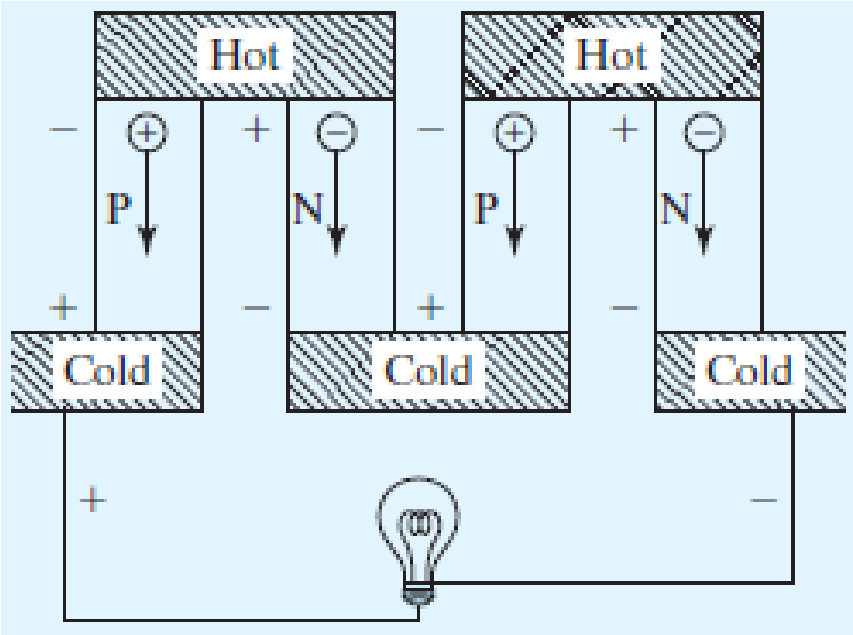
- Zig-zag motion is due to collisions or scattering with imperfections in the crystal.
- Net thermal velocity is zero.
- Mean time between collisions is $\tau_m \sim 0.1$ps

Hot-point Probe can determine sample doing type



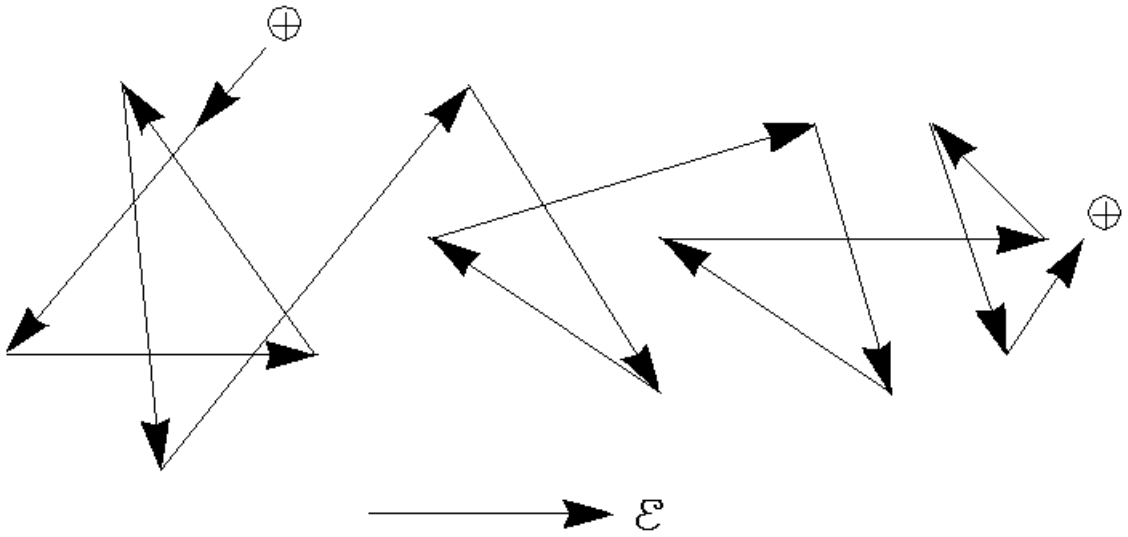
***Hot-point Probe** distinguishes N and P type semiconductors.*

***Thermoelectric Generator** (from heat to electricity) and Cooler (from electricity to refrigeration)*



2.2 Drift

2.2.1 Electron and Hole Mobilities



- ***Drift*** is the motion caused by an electric field.

2.2.1 Electron and Hole Mobilities

$$m_p v = q\mathcal{E}\tau_{mp}$$

$$v = \frac{q\mathcal{E}\tau_{mp}}{m_p}$$

$$v = \mu_p \mathcal{E}$$

$$\mu_p = \frac{q\tau_{mp}}{m_p}$$

$$v = -\mu_n \mathcal{E}$$

$$\mu_n = \frac{q\tau_{mn}}{m_n}$$

- μ_p is the hole mobility and μ_n is the electron mobility

2.2.1 Electron and Hole Mobilities

$v = \mu \mathbf{E}$; μ has the dimensions of $v/\mathbf{E}$ $\left[\frac{\text{cm/s}}{\text{V/cm}} = \frac{\text{cm}^2}{\text{V}\cdot\text{s}}\right]$.

Electron and hole mobilities of selected semiconductors

	Si	**Ge**	**GaAs**	**InAs**
μ_n ($\text{cm}^2/\text{V}\cdot\text{s}$)	1400	3900	8500	30000
μ_p ($\text{cm}^2/\text{V}\cdot\text{s}$)	470	1900	400	500

Based on the above table alone, which semiconductor and which carriers (electrons or holes) are attractive for applications in high-speed devices?

Drift Velocity, Mean Free Time, Mean Free Path

EXAMPLE: *Given $\mu_p = 470\ cm^2/V{\cdot}s$, what is the hole drift velocity at $\mathcal{E} = 10^3\ V/cm$? What is τ_{mp} and what is the distance traveled between collisions (called the **mean free path**)? Hint: When in doubt, use the MKS system of units.*

Solution: $v = \mu_p \mathcal{E} = 470\ \text{cm}^2/\text{V}{\cdot}\text{s} \times 10^3\ \text{V/cm} = 4.7\times 10^5\ \text{cm/s}$

$$\tau_{mp} = \mu_p m_p/q = 470\ \text{cm}^2/\text{V}\cdot\text{s} \times 0.39 \times 9.1\times10^{-31}\ \text{kg}/1.6\times10^{-19}\ \text{C}$$

$$= 0.047\ \text{m}^2/\text{V}\cdot\text{s} \times 2.2\times10^{-12}\ \text{kg/C} = 1\times10^{-13}\text{s} = 0.1\ \text{ps}$$

$$\textit{mean free path} = \tau_{mh} v_{th} \sim 1\times 10^{-13}\ \text{s} \times 2.2\times10^7\ \text{cm/s}$$

$$= 2.2\times10^{-6}\ \text{cm} = 220\ \text{Å} = 22\ \text{nm}$$

This is smaller than the typical dimensions of devices, but getting close.

2.2.2 Mechanisms of Carrier Scattering

There are two main causes of carrier scattering:

1. Phonon Scattering
2. Ionized-Impurity (Coulombic) Scattering

Phonon scattering mobility decreases when temperature rises:

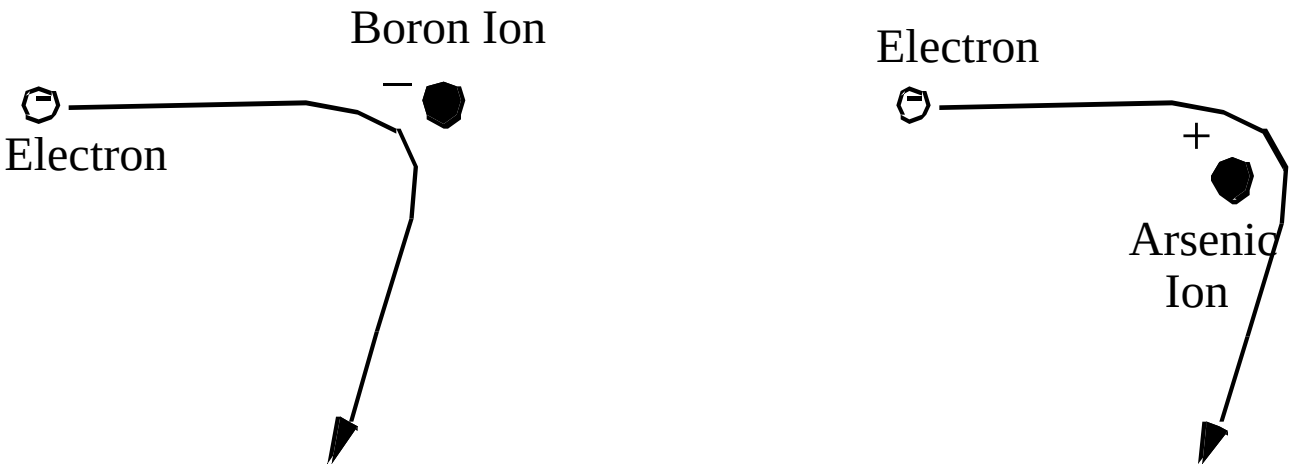
$$\mu_{phonon} \propto \tau_{phonon} \propto \frac{1}{phonon \quad density \times carrier \quad thermal \quad velocity} \propto \frac{1}{T \times T^{1/2}} \propto T^{-3/2}$$

$\mu = q\tau/m$

$\propto T$

$v_{th} \propto T^{1/2}$

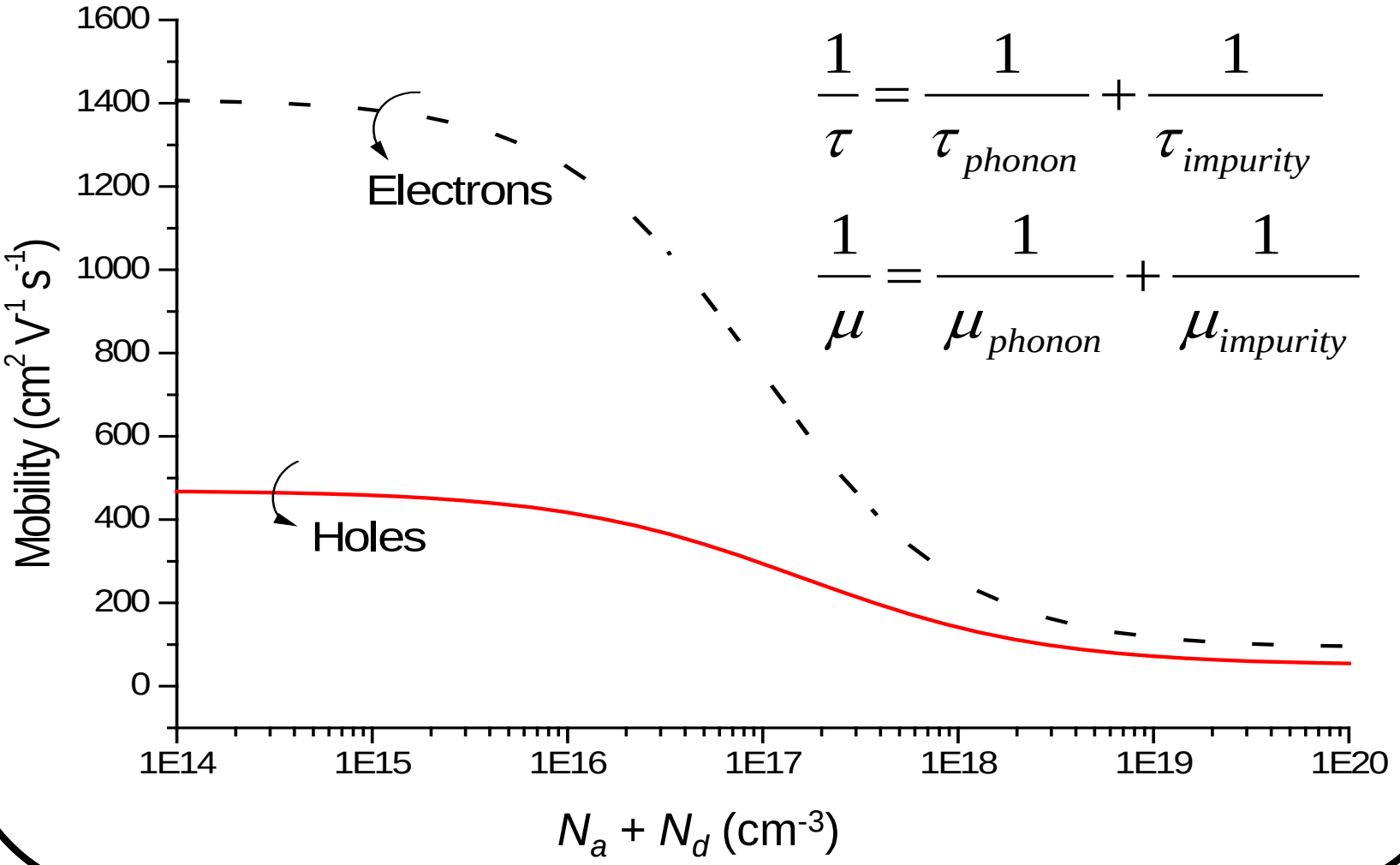
Impurity (Dopant)-Ion Scattering or Coulombic Scattering



There is less change in the direction of travel if the electron zips by the ion at a higher speed.

$$\mu_{impurity} \propto \frac{v_{th}^3}{N_a + N_d} \propto \frac{T^{3/2}}{N_a + N_d}$$

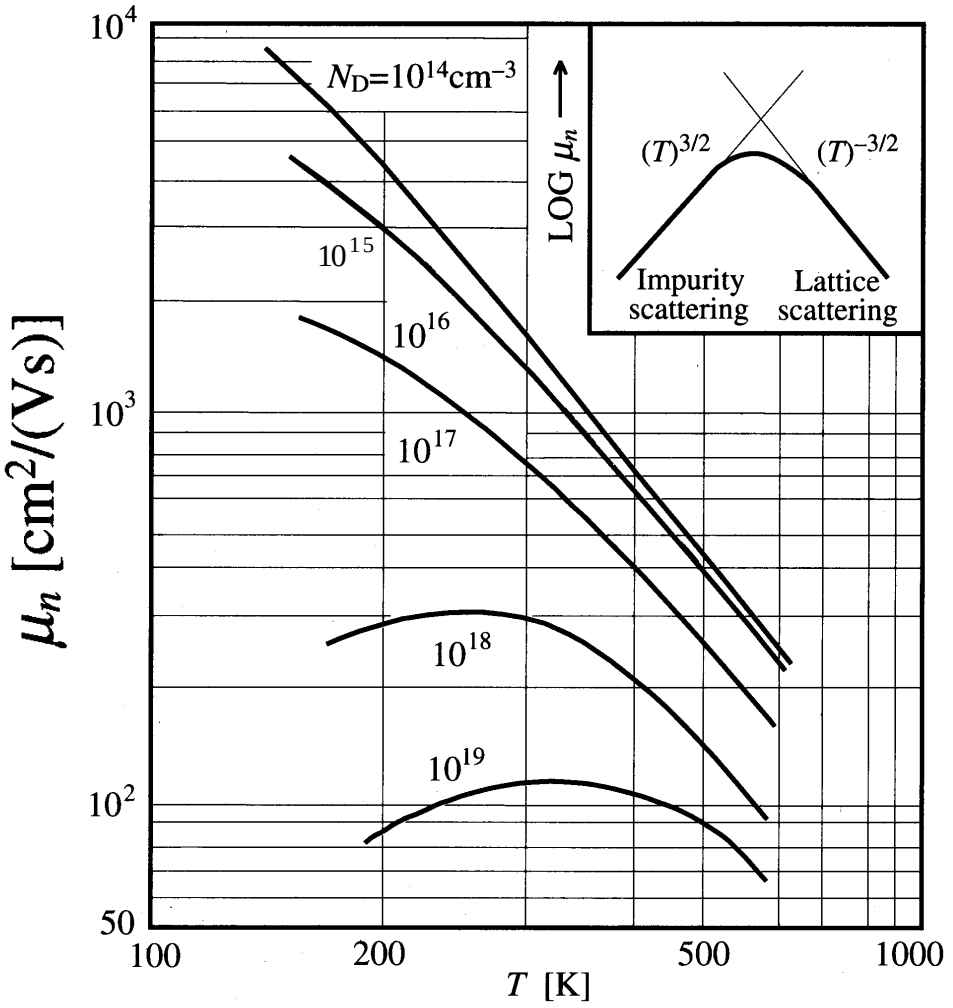
Total Mobility



$$\frac{1}{\tau} = \frac{1}{\tau_{phonon}} + \frac{1}{\tau_{impurity}}$$

$$\frac{1}{\mu} = \frac{1}{\mu_{phonon}} + \frac{1}{\mu_{impurity}}$$

Temperature Effect on Mobility



Question:
What N_d will make $d\mu_n/dT = 0$ at room temperature?

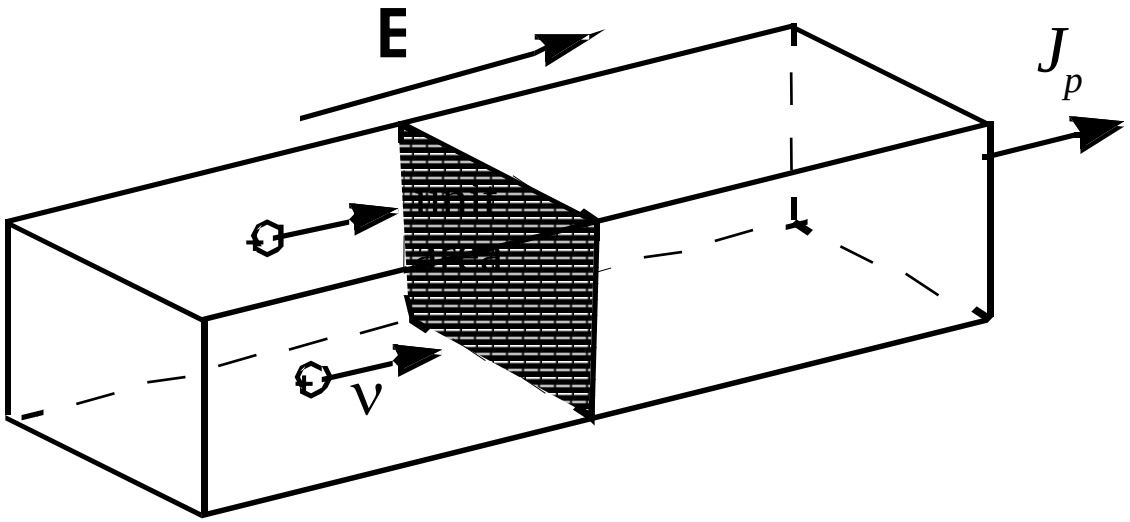
Velocity Saturation

• When the kinetic energy of a carrier exceeds a critical value, it generates an optical phonon and loses the kinetic energy.

• Therefore, the kinetic energy is capped at large **E,** and the velocity does not rise above a saturation velocity, v_{sat}.

• ***Velocity saturation*** has a deleterious effect on device speed as shown in Ch. 6.

2.2.3 Drift Current and Conductivity



Hole current density $\boxed{J_p = qpv}$ A/cm^2 or C/cm^2·sec

EXAMPLE: If $p = 10^{15}$cm^{-3} and $v = 10^4$ cm/s, then

$$J_p = 1.6\times10^{-19}\text{C} \times 10^{15}\text{cm}^{-3} \times 10^4\text{cm/s}$$

$$= 1.6\ \text{C/s}\cdot\text{cm}^2 = 1.6\ \text{A/cm}^2$$

2.2.3 Drift Current and Conductivity

$$J_{p,drift} = qpv = qp\mu_p \mathbf{E}$$

$$J_{n,drift} = -qnv = qn\mu_n \mathbf{E}$$

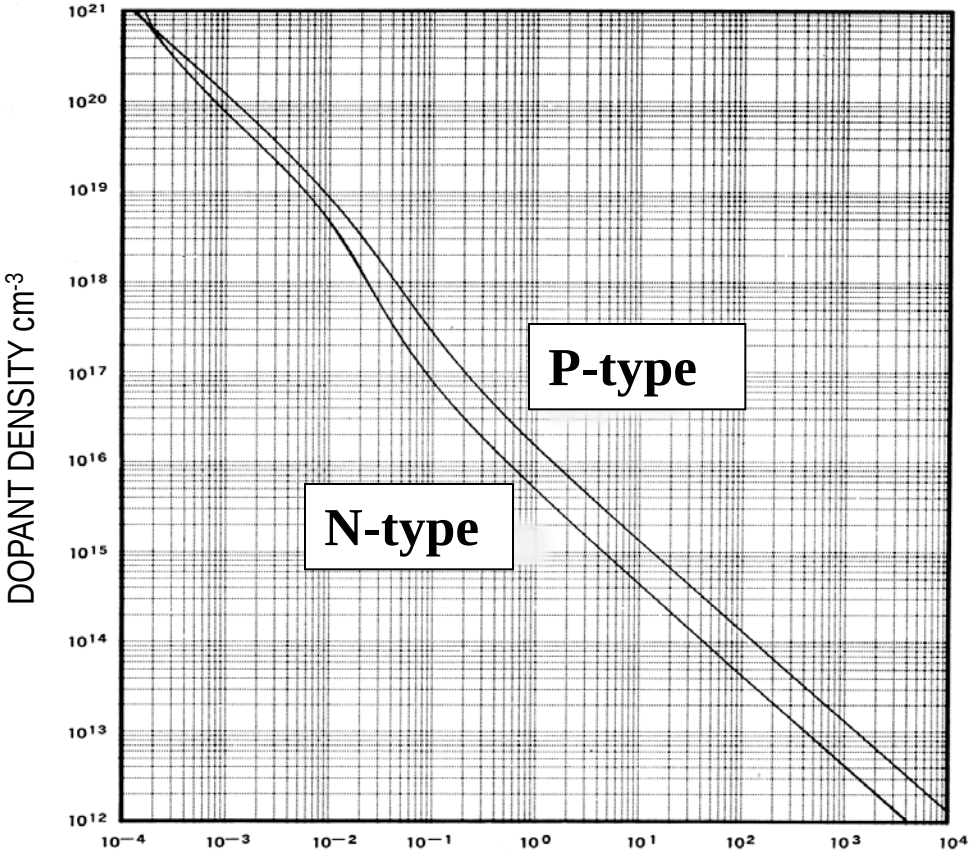
$$J_{drift} = J_{n,drift} + J_{p,drift} = \sigma \mathbf{E} = (qn\mu_n + qp\mu_p)\mathbf{E}$$

$\therefore$ ***conductivity*** *(1/ohm-cm) of a semiconductor is*

$$\sigma = qn\mu_n + qp\mu_p$$

$1/\sigma$ = is resistivity (ohm-cm)

Relationship between Resistivity and Dopant Density



$$\rho = 1/\sigma$$

EXAMPLE: Temperature Dependence of Resistance

(a) What is the resistivity (ρ) of silicon doped with $10^{17}cm^{-3}$ of arsenic?

(b) What is the resistance (R) of a piece of this silicon material 1μm long and 0.1 μm^2 in cross-sectional area?

Solution:

(a) Using the N-type curve in the previous figure, we find that ρ = 0.084 Ω-cm.

(b) $R = \rho L/A = 0.084\ \Omega\text{-cm} \times 1\ \mu\text{m} / 0.1\ \mu\text{m}^2$

$= 0.084\ \Omega\text{-cm} \times 10^{-4}\ \text{cm}/\ 10^{-10}\ \text{cm}^2$

$= 8.4 \times 10^{-4}\ \Omega$

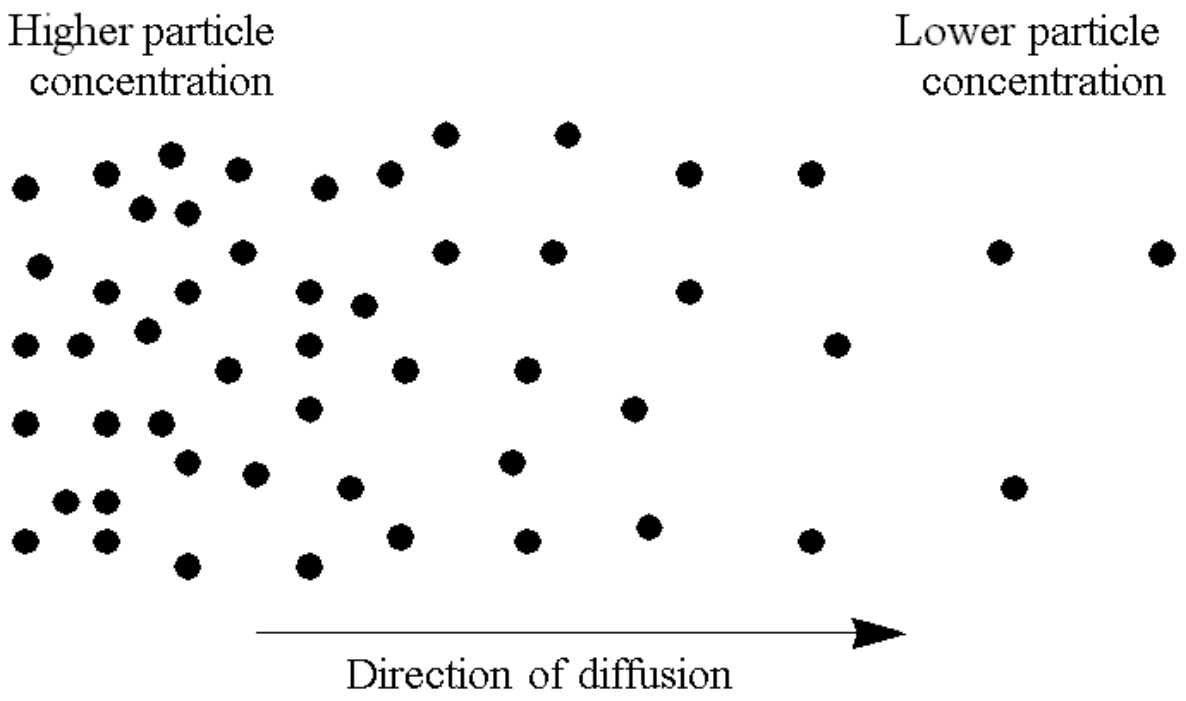
EXAMPLE: Temperature Dependence of Resistance

By what factor will R increase or decrease from T=300 K to T=400 K?

Solution: *The temperature dependent factor in σ (and therefore ρ) is μ_n. From the mobility vs. temperature curve for $10^{17}cm^{-3}$, we find that μ_n decreases from 770 at 300K to 400 at 400K. As a result, R* ***increases*** *by*

$$\frac{770}{400} = 1.93$$

2.3 Diffusion Current

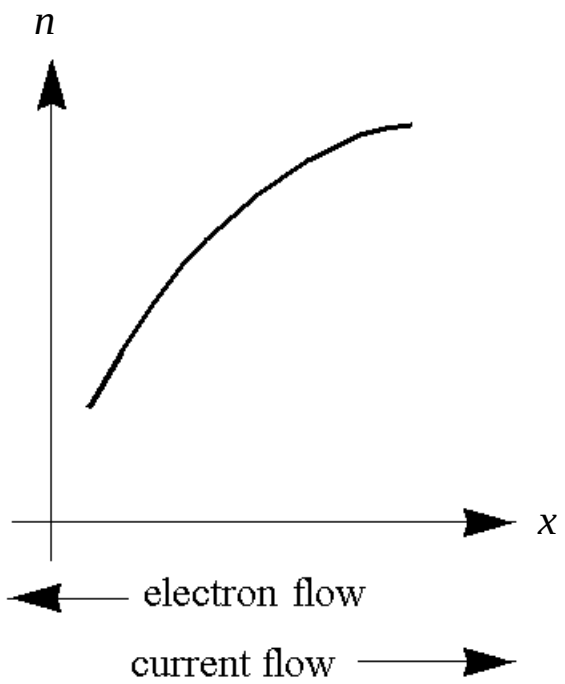


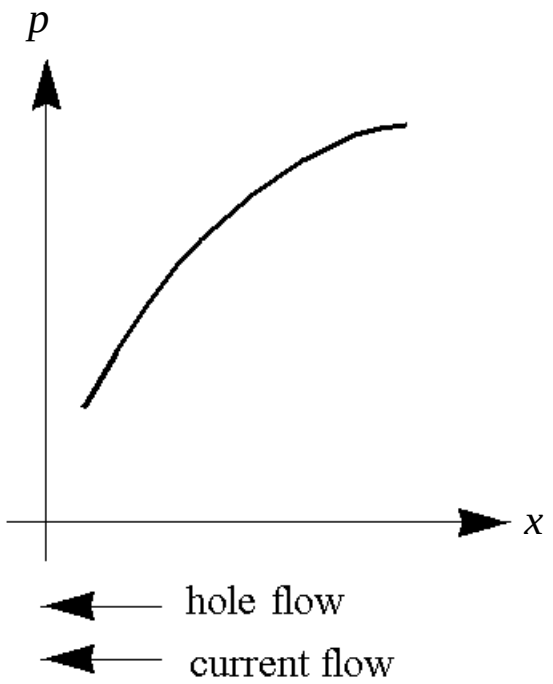
Particles diffuse from a higher-concentration location to a lower-concentration location.

2.3 Diffusion Current

$$J_{n,diffusion} = qD_n \frac{dn}{dx} \qquad J_{p,diffusion} = -qD_p \frac{dp}{dx}$$

D is called the diffusion constant. Signs explained:





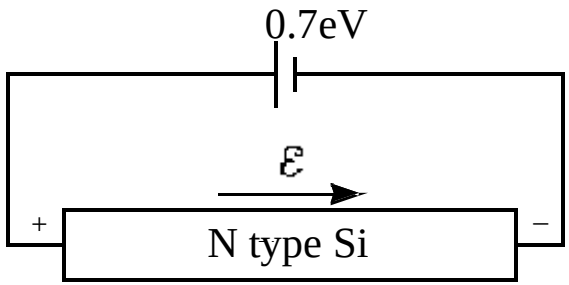
Total Current – Review of Four Current Components

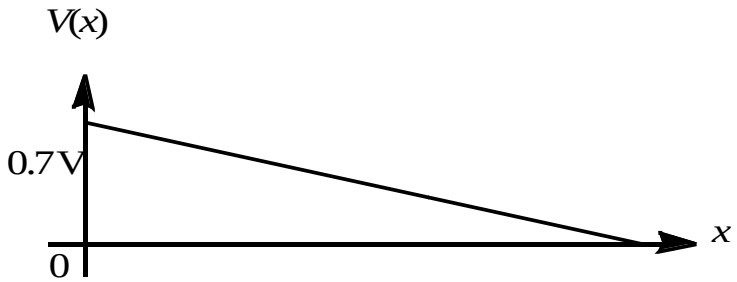
$$J_{TOTAL} = J_n + J_p$$

$$J_n = J_{n,drift} + J_{n,diffusion} = qn\mu_n \mathbf{E} + qD_n \frac{dn}{dx}$$

$$J_p = J_{p,drift} + J_{p,diffusion} = qp\mu_p \mathbf{E} - qD_p \frac{dp}{dx}$$

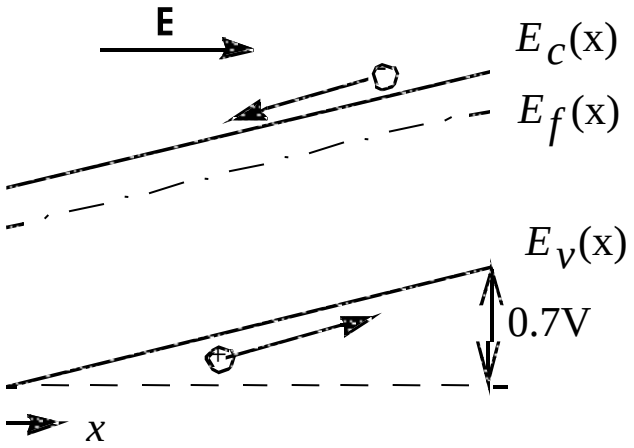
2.4 Relation Between the Energy Diagram and V, E





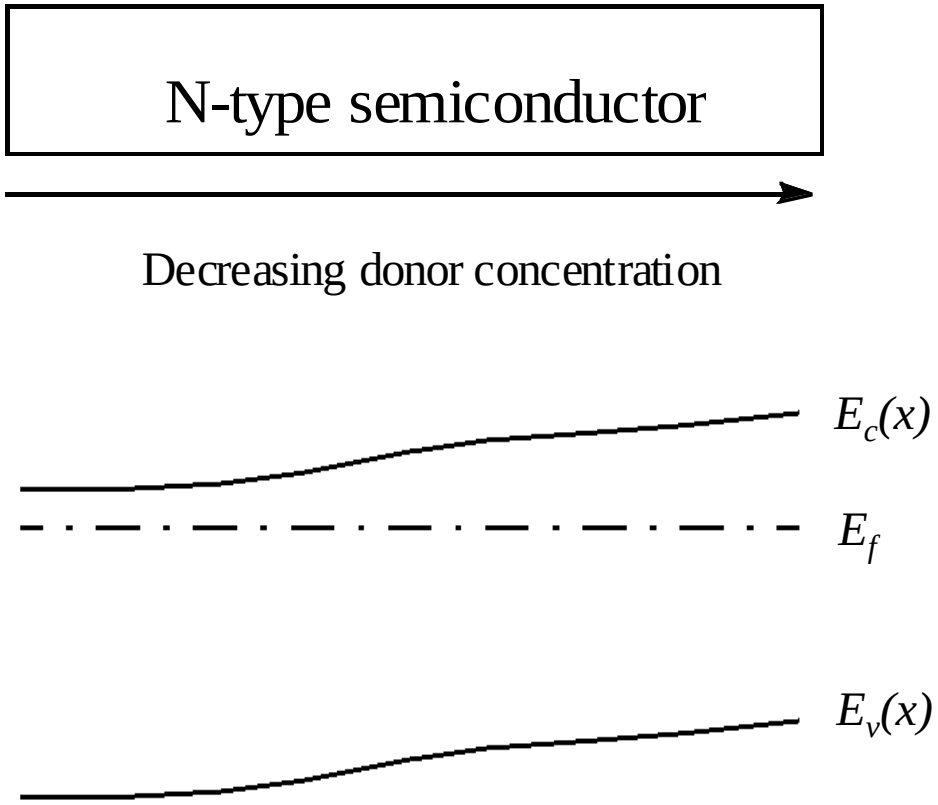
E_c and E_v vary in the opposite direction from the voltage. That is, E_c and E_v are higher where the voltage is lower.

$$\mathbf{E}(x) = -\frac{dV}{dx} = \frac{1}{q}\frac{dE_c}{dx} = \frac{1}{q}\frac{dE_v}{dx}$$



2.5 Einstein Relationship between D and μ

Consider a piece of non-uniformly doped semiconductor.



$$n = N_c e^{-(E_c - E_f)/kT}$$

$$\frac{dn}{dx} = -\frac{N_c}{kT} e^{-(E_c - E_f)/kT} \frac{dE_c}{dx}$$

$$= -\frac{n}{kT} \frac{dE_c}{dx}$$

$$= -\frac{n}{kT} q\mathcal{E}$$

2.5 Einstein Relationship between D and μ

$$\frac{dn}{dx} = -\frac{n}{kT} q\mathcal{E}$$

$$J_n = qn\mu_n \mathcal{E} + qD_n \frac{dn}{dx} = 0 \quad \text{at equilibrium.}$$

$$0 = qn\mu_n \mathcal{E} - qn\frac{qD_n}{kT}\mathcal{E}$$

$$\boxed{D_n = \frac{kT}{q}\mu_n} \quad \text{Similarly,} \quad \boxed{D_p = \frac{kT}{q}\mu_p}$$

These are known as the Einstein relationship.

EXAMPLE: Diffusion Constant

What is the hole diffusion constant in a piece of silicon with $\mu_p = 410\ cm^2\ V^{-1}s^{-1}$?

Solution:

$$D_p = \left(\frac{kT}{q}\right)\mu_p = (26\,\mathrm{mV})\cdot 410\,\mathrm{cm^2V^{-1}s^{-1}} = 11\,\mathrm{cm^2/s}$$

Remember: kT/q = 26 mV at room temperature.

2.6 Electron-Hole Recombination

- The equilibrium carrier concentrations are denoted with n_0 and p_0.
- The total electron and hole concentrations can be different from n_0 and p_0 .
- *T*he differences are called the ***excess carrier concentrations*** *n'* and *p'*.

$$n \equiv n_0 + n'$$
$$p \equiv p_0 + p'$$

Charge Neutrality

•Charge neutrality is satisfied at equilibrium ($n' = p' = 0$).

• When a non-zero n' is present, an equal p' may be assumed to be present to maintain charge equality and vice-versa.

•If charge neutrality is not satisfied, the net charge will attract or repel the (majority) carriers through the drift current until neutrality is restored.

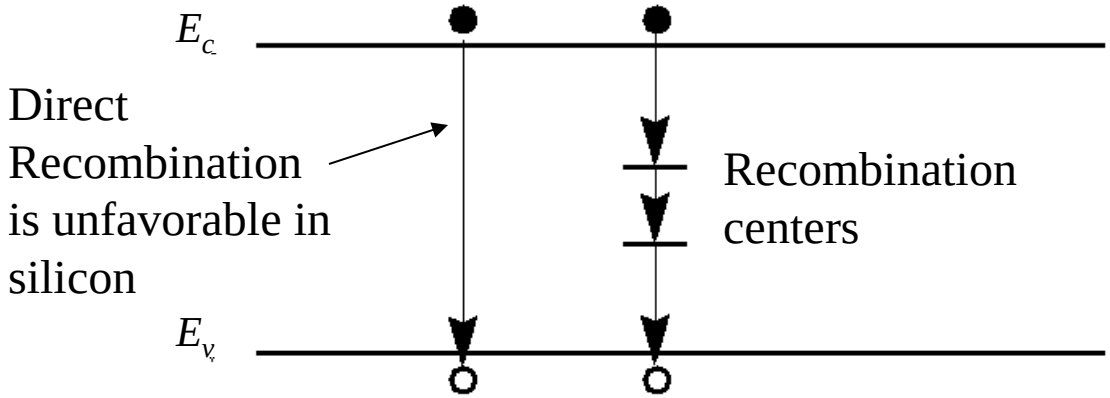
$$\boxed{n' = p'}$$

Recombination Lifetime

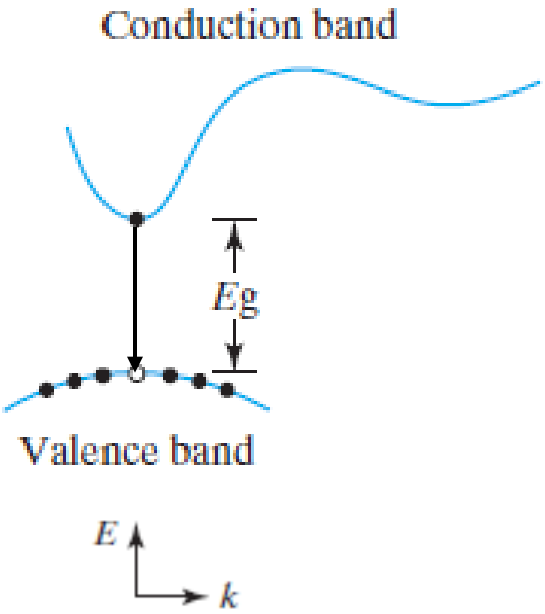
- Assume light generates *n'* and *p'*. If the light is suddenly turned off, *n'* and *p'* decay with time until they become zero.
- The process of decay is called ***recombination.***
- The time constant of decay is the ***recombination time*** or ***carrier lifetime,*** τ.
- Recombination is nature's way of restoring equilibrium ($n' = p' = 0$).

Recombination Lifetime

- τ ranges from 1ns to 1ms in Si and depends on the density of metal impurities (contaminants) such as Au and Pt.
- These ***deep traps*** capture electrons and holes to facilitate recombination and are called ***recombination centers.***

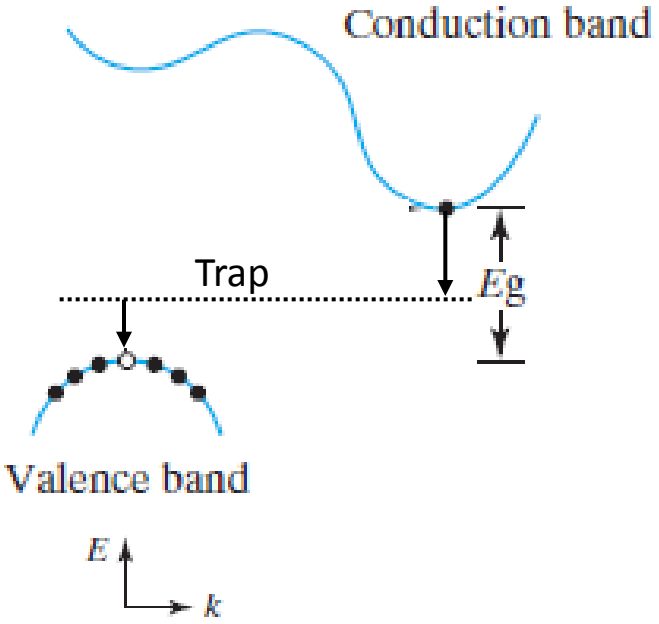


Direct and Indirect Band Gap



Direct band gap
Example: GaAs

Direct recombination is efficient as k conservation is satisfied.



Indirect band gap
Example: Si

Direct recombination is rare as k conservation is not satisfied

Rate of recombination ($s^{-1}cm^{-3}$)

$$\frac{dn'}{dt} = -\frac{n'}{\tau}$$

$$n' = p'$$

$$\frac{dn'}{dt} = -\frac{n'}{\tau} = -\frac{p'}{\tau} = \frac{dp'}{dt}$$

EXAMPLE: Photoconductors

A bar of Si is doped with boron at $10^{15}cm^{-3}$. It is exposed to light such that electron-hole pairs are generated throughout the volume of the bar at the rate of $10^{20}/s \cdot cm^3$. The recombination lifetime is $10\mu s$. What are (a) p_0 , (b) n_0 , (c) p', (d) n', (e) p , (f) n, and (g) the np product?

EXAMPLE: Photoconductors

Solution:

(a) What is p_0?

$p_0 = N_a = 10^{15}\,\text{cm}^{-3}$

(b) What is n_0 ?

$n_0 = n_i^2/p_0 = 10^5\,\text{cm}^{-3}$

(c) What is p' ?

In steady-state, the rate of generation is equal to the rate of recombination.

$$10^{20}/\text{s-cm}^3 = p'/\tau$$

$$\therefore p' = 10^{20}/\text{s-cm}^3 \cdot 10^{-5}\text{s} = 10^{15}\,\text{cm}^{-3}$$

EXAMPLE: Photoconductors

(d) What is n'?

$n' = p' = 10^{15}\,\text{cm}^{-3}$

(e) What is p?

$p = p_0 + p' = 10^{15}\text{cm}^{-3} + 10^{15}\text{cm}^{-3} = 2\times10^{15}\text{cm}^{-3}$

(f) What is n?

$n = n_0 + n' = 10^{5}\text{cm}^{-3} + 10^{15}\text{cm}^{-3} \sim 10^{15}\text{cm}^{-3}$ *since* $n_0 << n'$

(g) What is np?

$np \sim 2\times10^{15}\text{cm}^{-3} \cdot 10^{15}\text{cm}^{-3} = 2\times10^{30}\,\text{cm}^{-6} >> n_i^2 = 10^{20}\,\text{cm}^{-6}$.

The np product can be very different from n_i^2.

2.7 Thermal Generation

If *n'* is negative, there are fewer electrons than the equilibrium value.

As a result, there is a net rate of ***thermal generation*** at the rate of $|n'|/\tau$.

2.8 Quasi-equilibrium and Quasi-Fermi Levels

• Whenever $n' = p' \neq 0$, $np \neq n_i^2$. We would like to preserve and use the simple relations:

$$n = N_c e^{-(E_c - E_f)/kT}$$

$$p = N_v e^{-(E_f - E_v)/kT}$$

• But these equations lead to $np = n_i^2$. The solution is to introduce two ***quasi-Fermi levels*** E_{fn} and E_{fp} such that

$$n = N_c e^{-(E_c - E_{fn})/kT}$$

$$p = N_v e^{-(E_{fp} - E_v)/kT}$$

Even when electrons and holes are not at equilibrium, *within each group* the carriers can be at equilibrium. Electrons are closely linked to other electrons but only loosely to holes.

EXAMPLE: Quasi-Fermi Levels and Low-Level Injection

Consider a Si sample with $N_d = 10^{17} cm^{-3}$ *and* $n' = p' = 10^{15} cm^{-3}$.

(a) Find E_f.

$$n = N_d = 10^{17}\ \text{cm}^{-3} = N_c \exp[-(E_c - E_f)/kT]$$

$$\therefore\ E_c - E_f = 0.15\ \text{eV}.$$ *(E_f is below E_c by 0.15 eV.)*

Note: n' and p' are much less than the majority carrier concentration. This condition is called ***low-level injection.***

EXAMPLE: Quasi-Fermi Levels and Low-Level Injection

Now assume $n' = p' = 10^{15}\ cm^{-3}$.

(b) Find E_{fn} *and* E_{fp} .

$$n = 1.01\times10^{17}\text{cm}^{-3} = N_c e^{-(E_c - E_{fn})/kT}$$

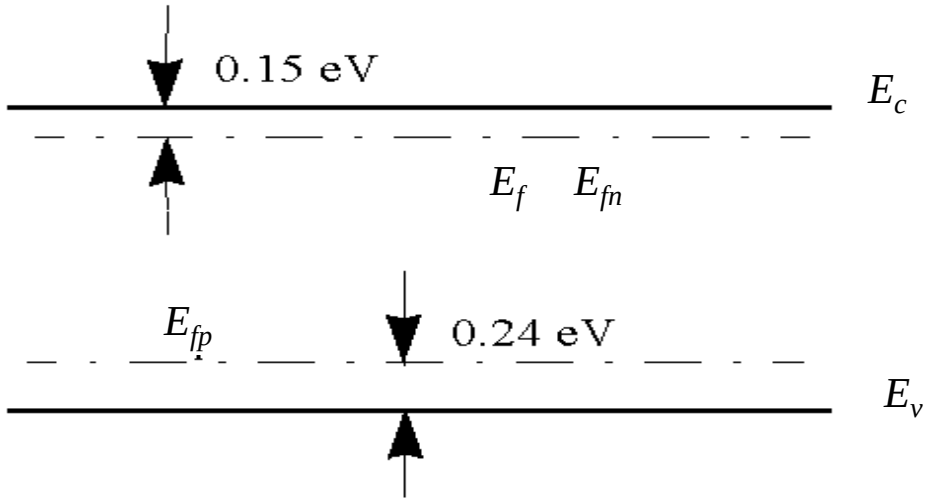
$$\therefore\ E_c - E_{fn} = kT \times \ln(N_c/1.01\times10^{17}\text{cm}^{-3})$$
$$= 26\ \text{meV} \times \ln(2.8\times10^{19}\text{cm}^{-3}/1.01\times10^{17}\text{cm}^{-3})$$
$$= 0.15\ \text{eV}$$

E_{fn} is nearly identical to E_f because $n \approx n_0$.

EXAMPLE: Quasi-Fermi Levels

$$p = 10^{15}\,\text{cm}^{-3} = N_v e^{-(E_{fp}-E_v)/kT}$$

$$\therefore\ E_{fp}-E_v = kT \times \ln(N_v/10^{15}\text{cm}^{-3})$$
$$= 26\ \text{meV} \times \ln(1.04\times10^{19}\text{cm}^{-3}/10^{15}\text{cm}^{-3})$$
$$= 0.24\ \text{eV}$$



2.9 Chapter Summary

$$v_p = \mu_p \mathbf{E}$$

$$v_n = -\mu_n \mathbf{E}$$

$$J_{p,drift} = qp\mu_p \mathbf{E}$$

$$J_{n,drift} = qn\mu_n \mathbf{E}$$

$$J_{n,diffusion} = qD_n \frac{dn}{dx}$$

$$J_{p,diffusion} = -qD_p \frac{dp}{dx}$$

$$D_n = \frac{kT}{q}\mu_n$$

$$D_p = \frac{kT}{q}\mu_p$$

2.9 Chapter Summary

τ is the recombination lifetime.

n' and *p'* are the ***excess carrier concentrations***.

$$n = n_0 + n'$$
$$p = p_0 + p'$$

Charge neutrality requires $n' = p'$.

$$\textit{rate of recombination} = n'/\tau = p'/\tau$$

E_{fn} and E_{fp} are the quasi-Fermi levels of electrons and holes.

$$n = N_c e^{-(E_c - E_{fn})/kT}$$
$$p = N_v e^{-(E_{fp} - E_v)/kT}$$

Chapter 3
Device Fabrication Technology

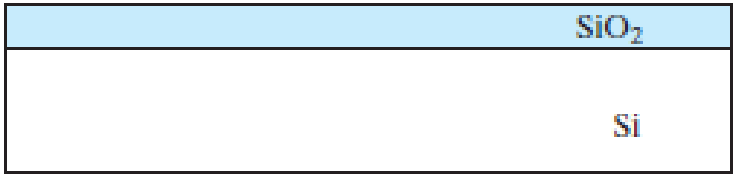
About 10^{20} transistors (or 10 billion for every person in the world) are manufactured every year.

VLSI (Very Large Scale Integration)
ULSI (Ultra Large Scale Integration)
GSI (Giga-Scale Integration)

Variations of this versatile technology are used for flat-panel displays, micro-electro-mechanical systems (***MEMS***), and chips for DNA screening...

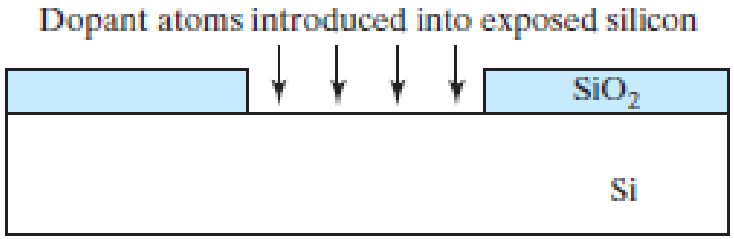
3.1 Introduction to Device Fabrication

Oxidation



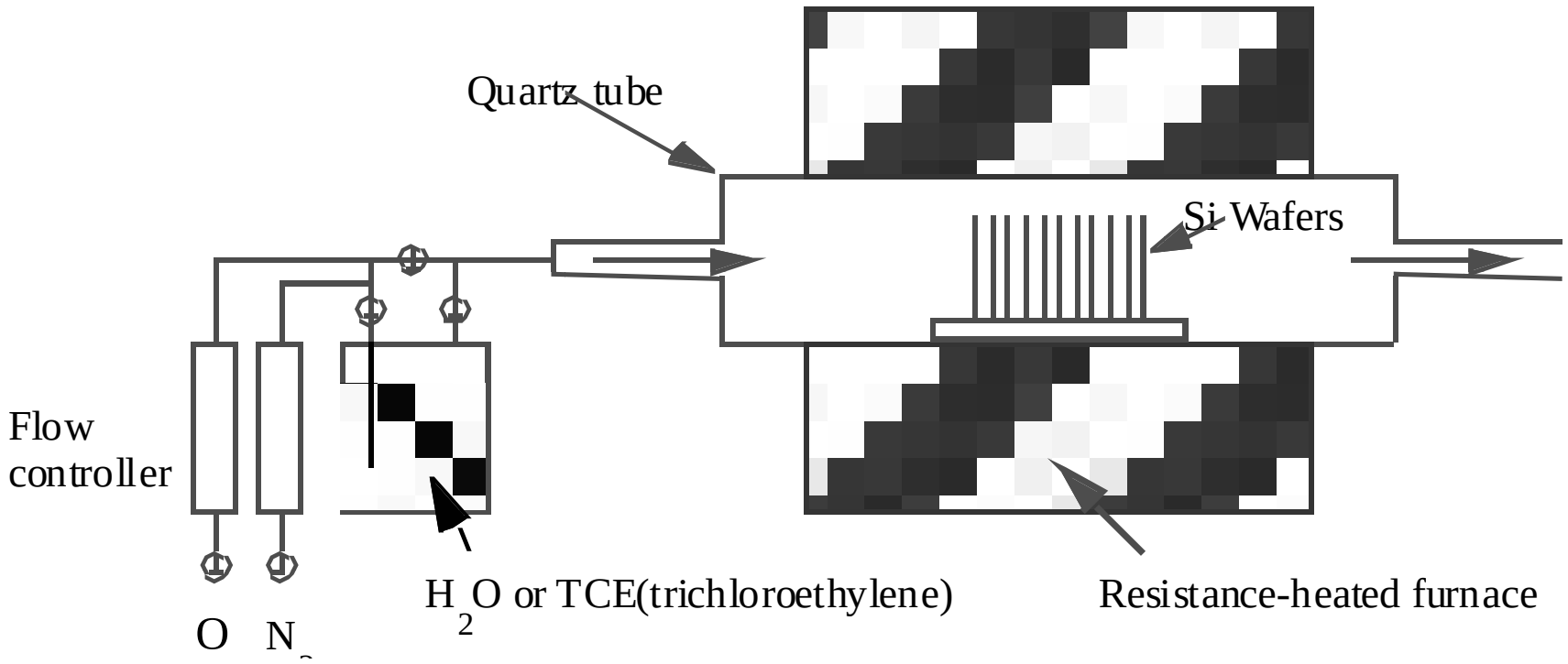
Lithography & Etching

Ion Implantation



Annealing & Diffusion

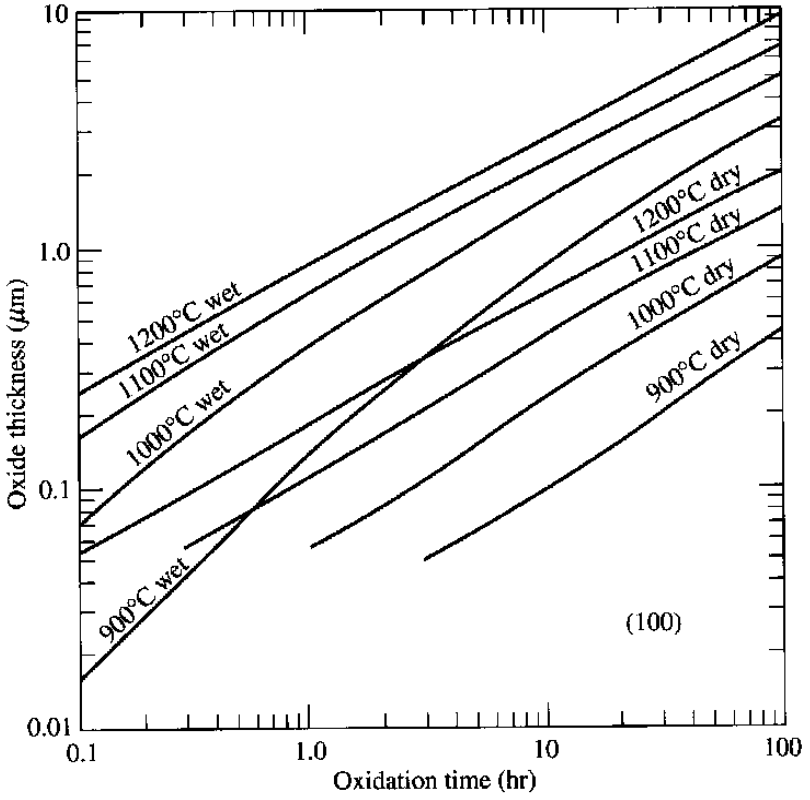
3.2 Oxidation of Silicon



3.2 Oxidation of Silicon

Dry Oxidation : $Si + O_2 \rightarrow SiO_2$

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



3.2 Oxidation of Silicon

EXAMPLE : Two-step Oxidation

(a) How long does it take to grow 0.1μm of dry oxide at 1000 ^{o}C ?

(b) After step (a), how long will it take to grow an additional 0.2μm of oxide at 900 ^{o}C in a wet ambient ?

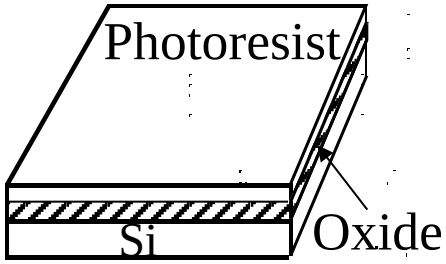
Solution:

(a) From the "1000^{o}C dry" curve in Slide 3-3, it takes 2.5 hr to grow 0.1μm of oxide.

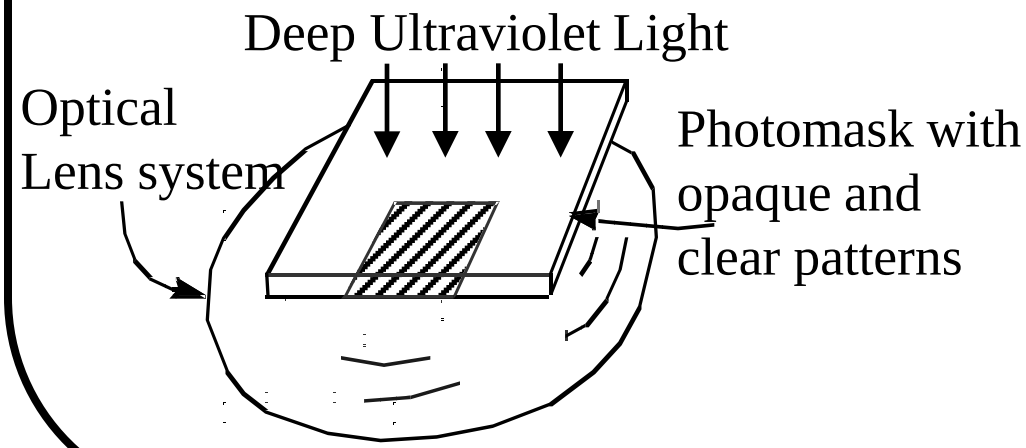
(b) Use the "900^{o}C wet" curve only. It would have taken 0.7hr to grow the 0.1 μm oxide and 2.4hr to grow 0.3 μm oxide from bare silicon. The answer is 2.4hr–0.7hr = 1.7hr.

3.3 Lithography

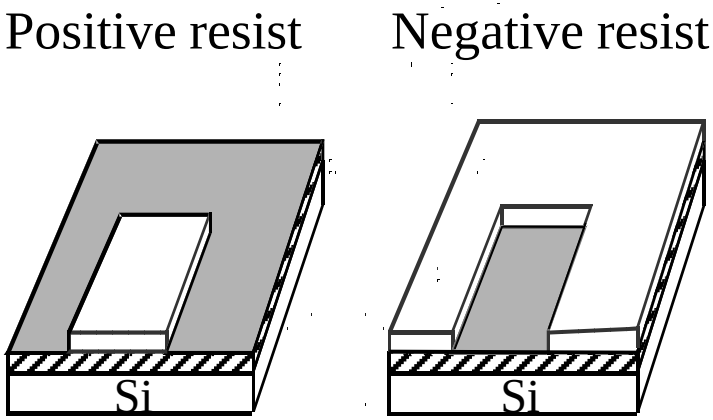
(a) Resist Coating



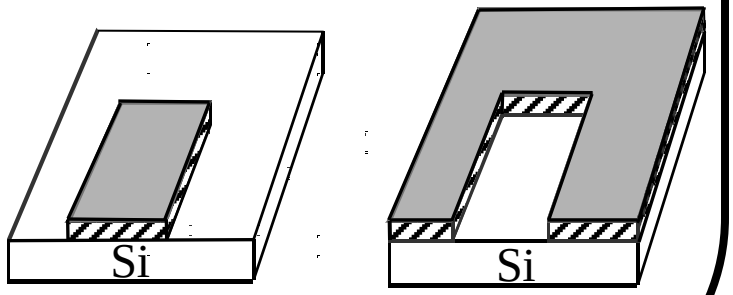
(b) Exposure



(c) Development



(d) Etching and Resist Strip



3.3 Lithography

Photolithography Resolution Limit, *R*

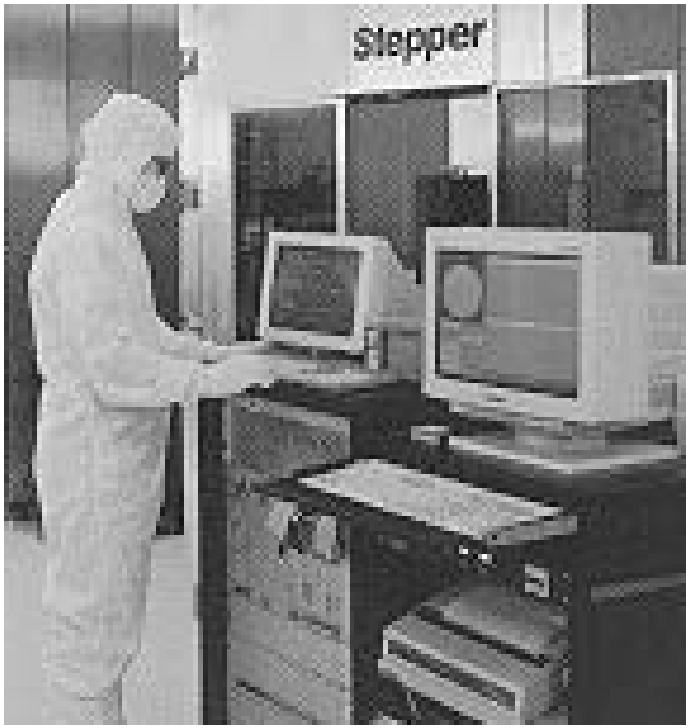
- $R \geq k\lambda$ due to optical diffraction
- Wavelength λ needs to be minimized. (248 nm, 193 nm, 157 nm?)
- k (<1) can be reduced will

  - Large aperture, high quality lens
  - Small exposure field, step-and-repeat using “stepper”
  - Optical proximity correction
  - Phase-shift mask, etc.

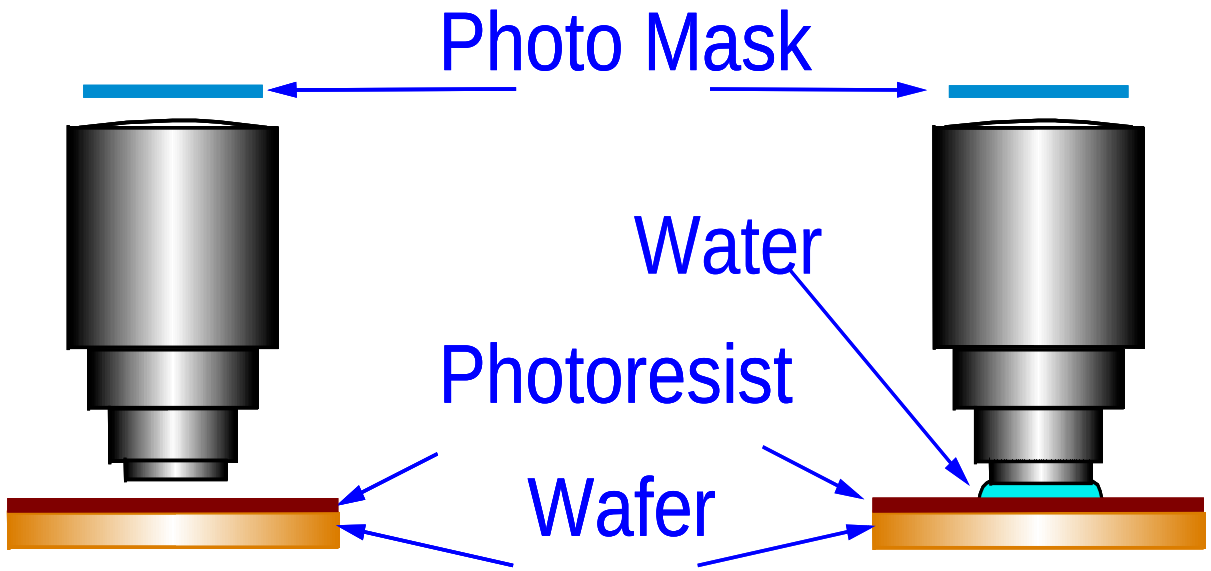
- Lithography is difficult and expensive. There can be 40 lithography steps in an IC process.

3.3 Lithography

Wafers are being loaded into a stepper in a clean room.



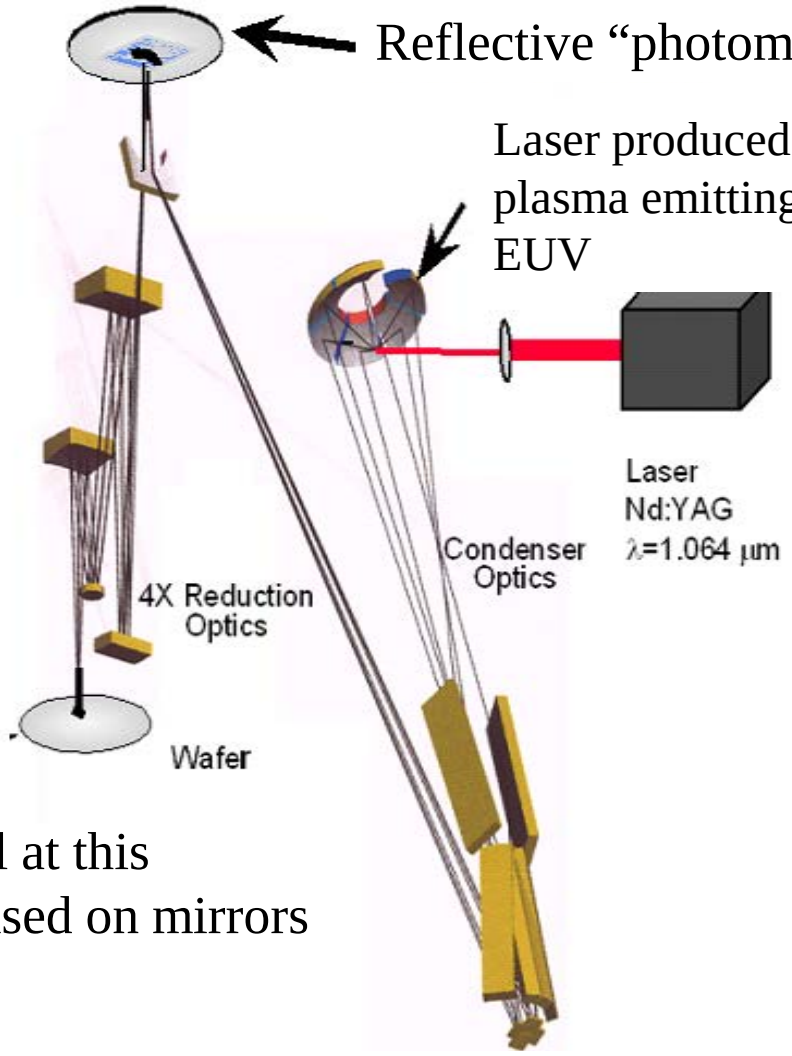
3.3.1 Wet Lithography



conventional dry lithography

wet or immersion lithography

Extreme UV Lithography (13nm wavelength)



No suitable lens material at this wavelength. Optics is based on mirrors with nm flatness.

Beyond Optical Lithography

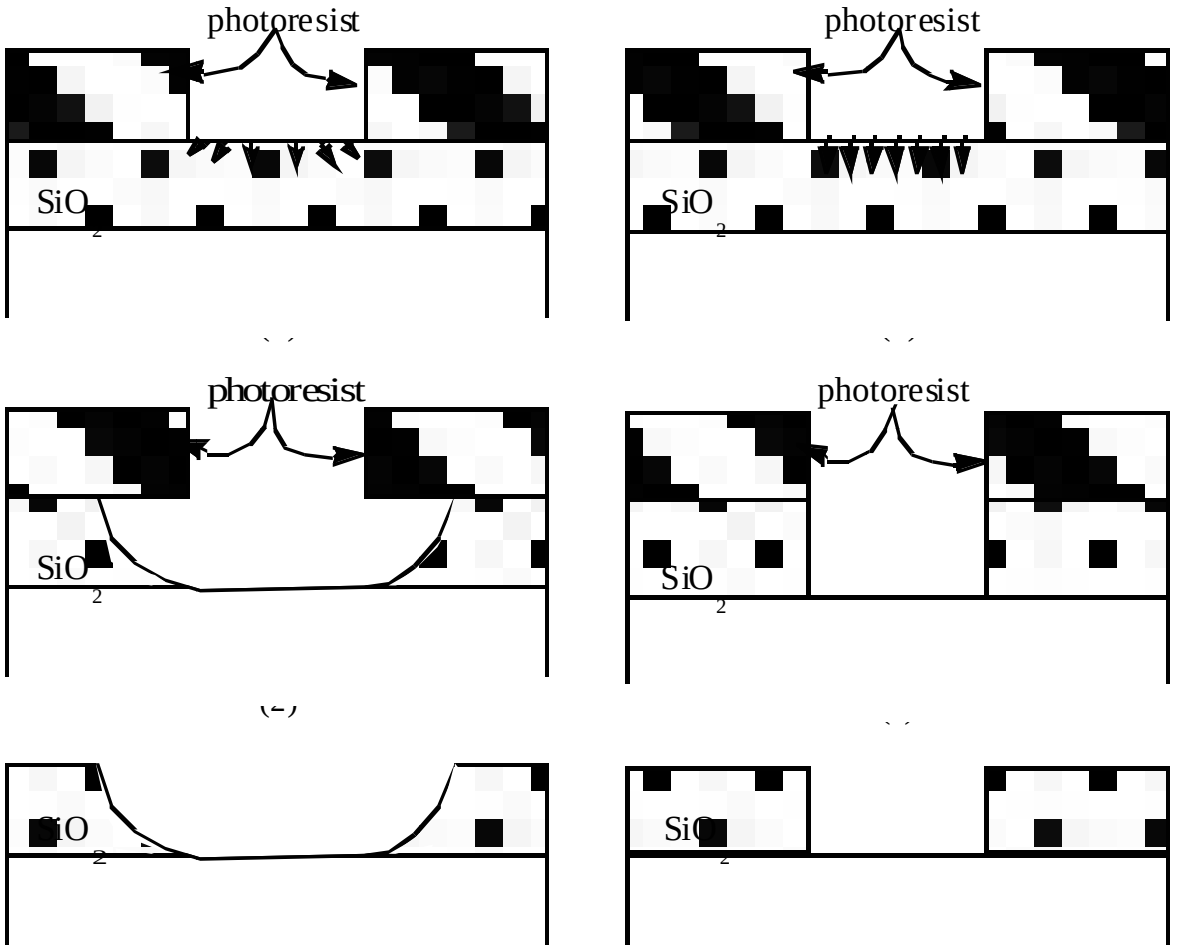
- ***Electron Beam Writing*** : Electron beam(s) scans and exposed electron resist on wafer. Ready technology with relatively low throughput.

- ***Electron Projection Lithography*** : Exposes a complex pattern using mask and electron lens similar to optical lithography.

- ***Nano-imprint*** : Patterns are etched into a durable material to make a "stamp." This stamp is pressed into a liquid film over the wafer surface. Liquid is hardened with UV to create an imprint of the fine patterns.

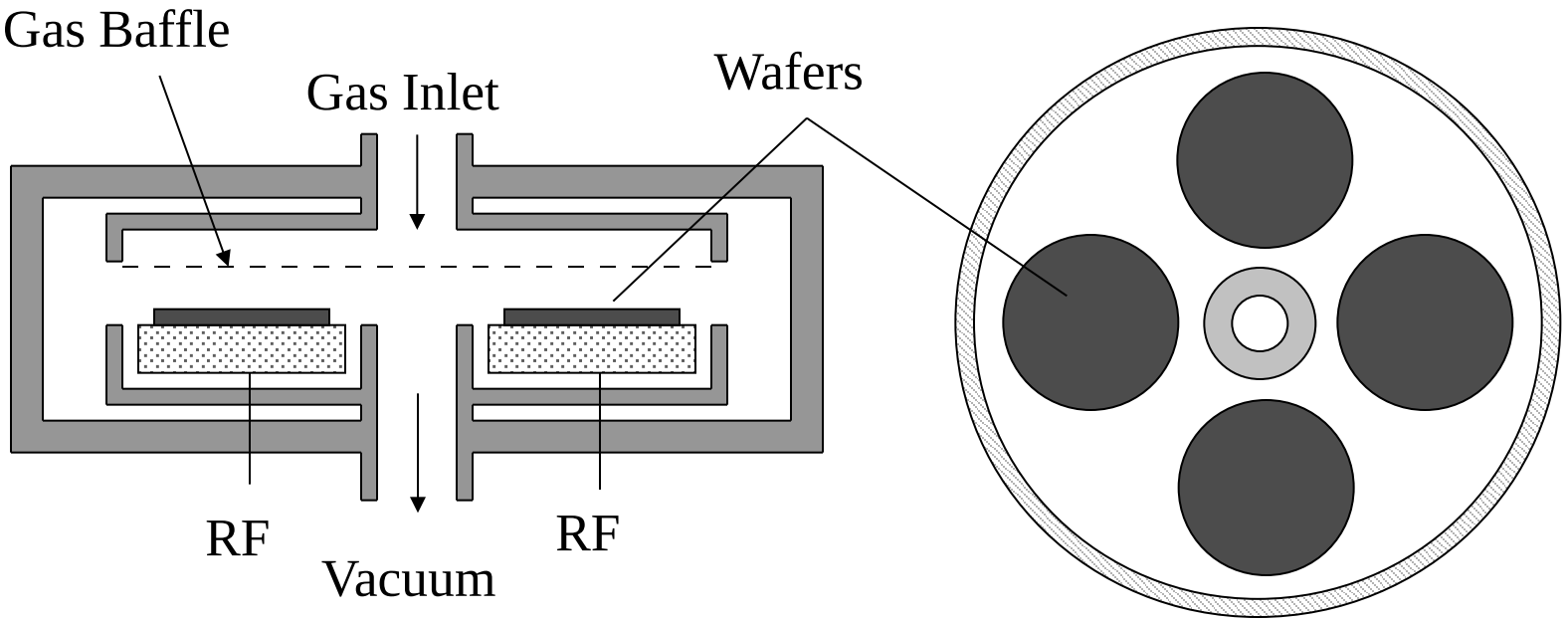
3.4 Pattern Transfer–Etching

Isotropic etching **Anisotropic etching**



3.4 Pattern Transfer–Etching

Reactive-Ion Etching Systems



Cross-section View

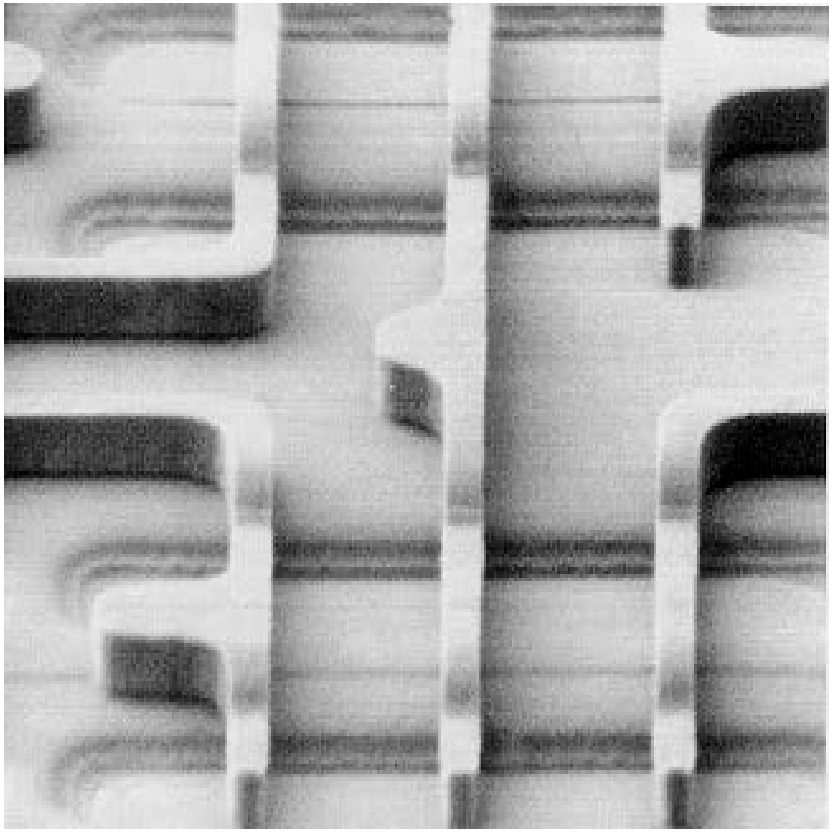
Top View

3.4 Pattern Transfer–Etching

Dry Etching (also known as Plasma Etching, or Reactive-Ion Etching) is anisotropic.

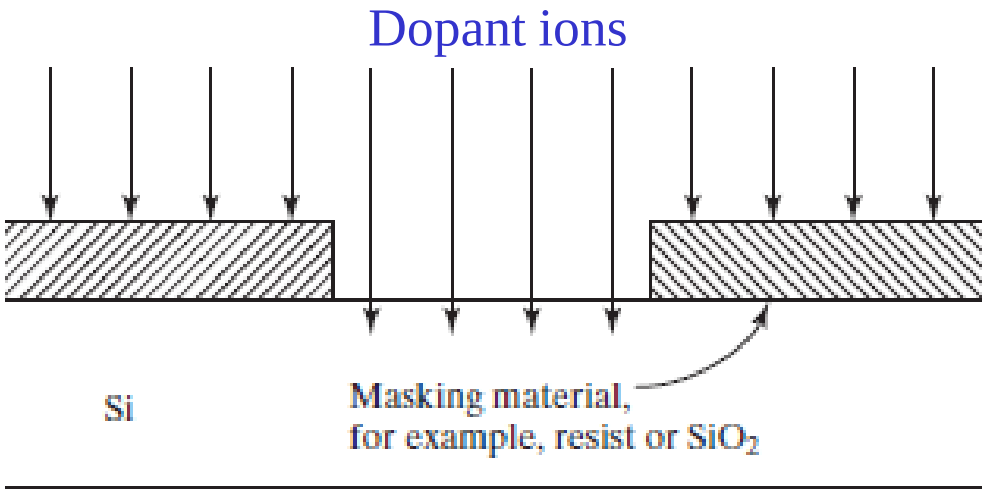
- Silicon and its compounds can be etched by plasmas containing F.
- Aluminum can be etched by Cl.
- Some concerns :
  - Selectivity and End-Point Detection
  - Plasma Process-Induced Damage or Wafer Charging Damage and Antenna Effect

Scanning electron microscope view of a plasma-etched 0.16 μm pattern in polycrystalline silicon film.



3.5 Doping

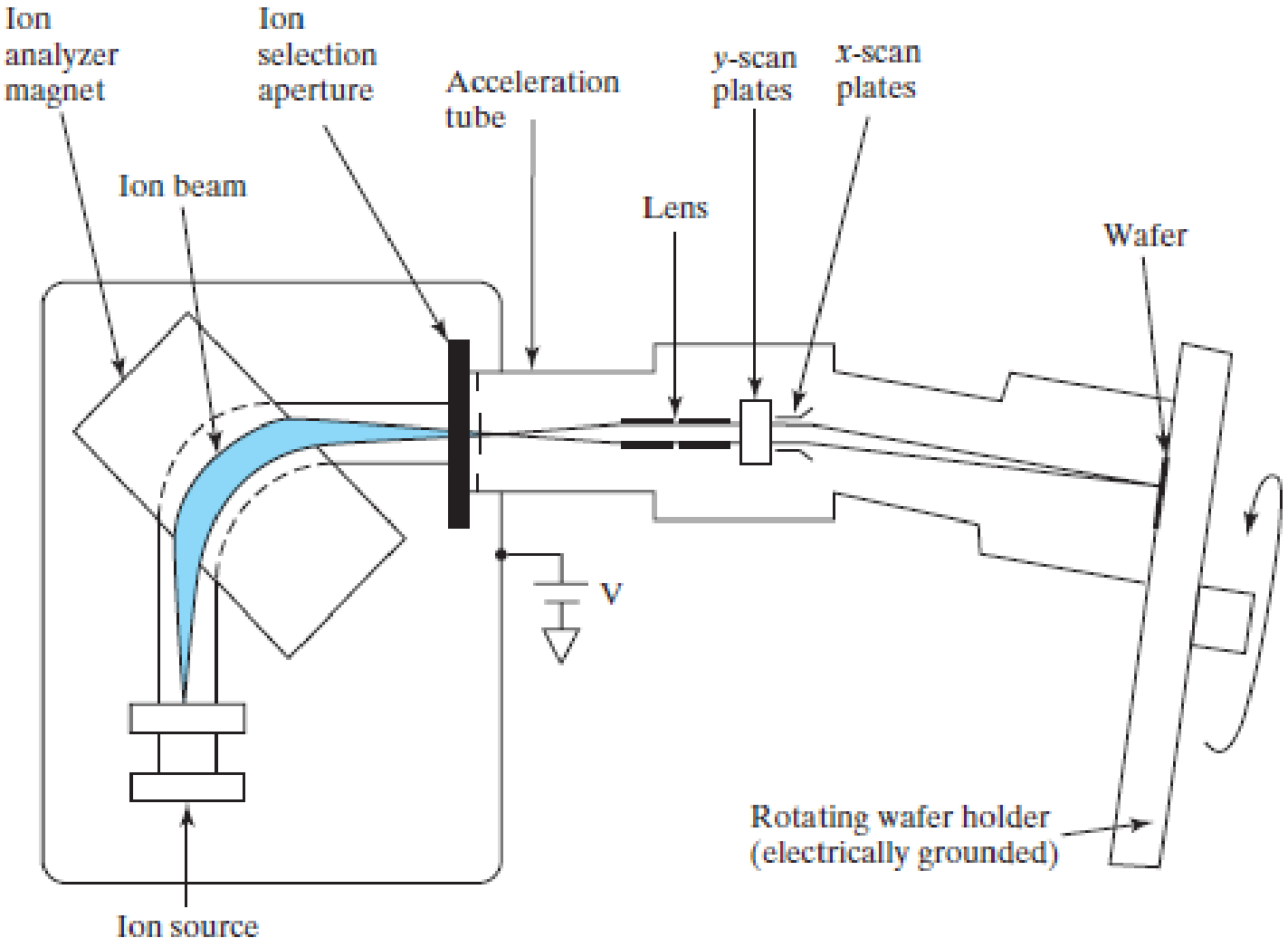
3.5.1 Ion Implantation



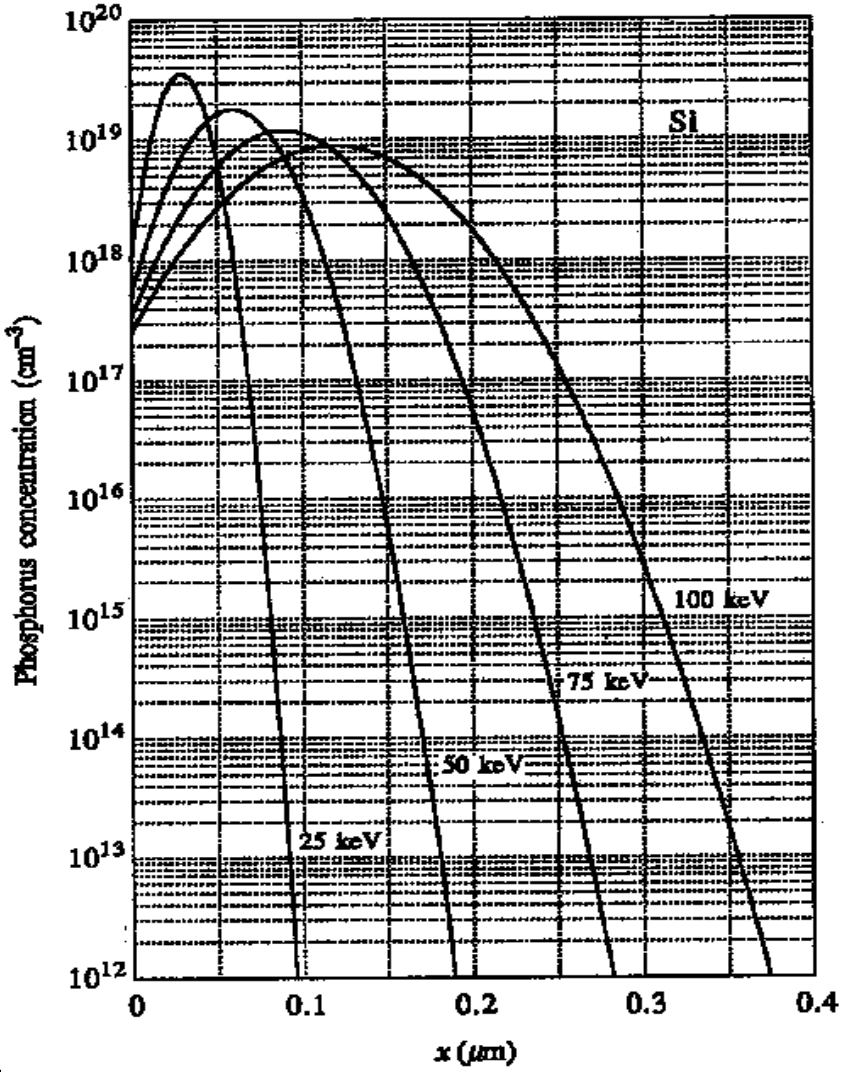
- The dominant doping method
- Excellent control of **dose** (cm^{-2})
- Good control of implant depth with energy (KeV to MeV)
- Repairing crystal damage and dopant activation requires annealing, which can cause dopant diffusion and loss of depth control.

3.5.1 Ion Implantation

Schematic of an Ion Implanter



3.5.1 Ion implantation



Phosphorous density profile after implantation

3.5.1 Ion Implantation

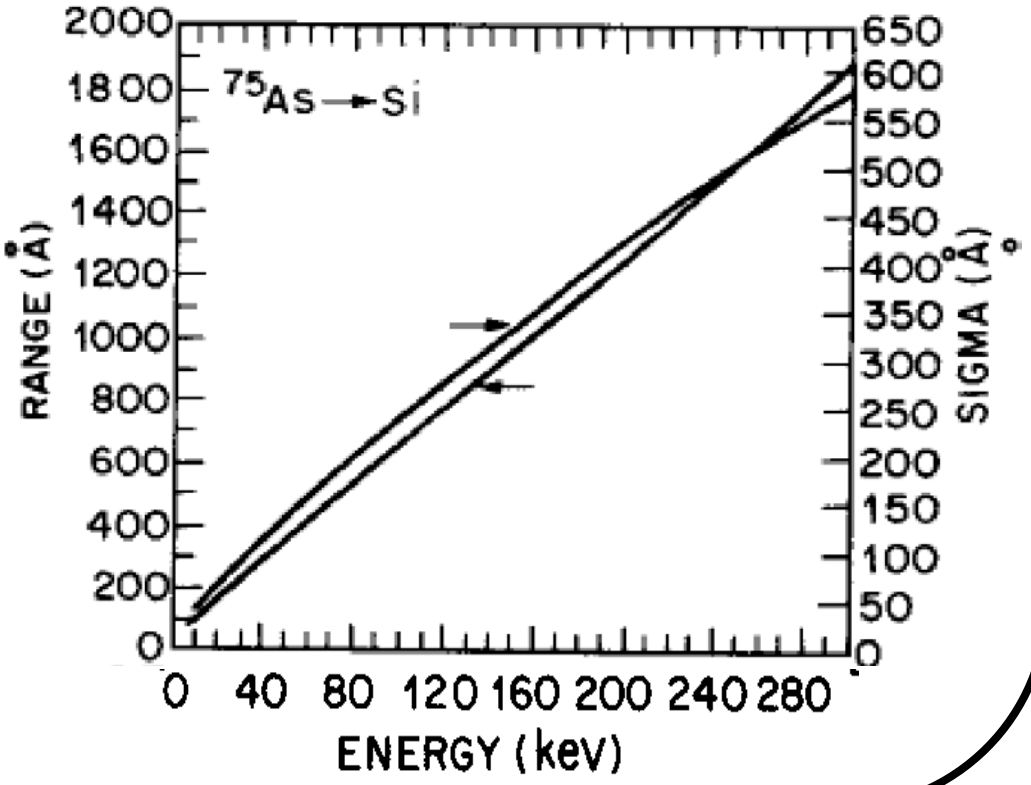
Model of Implantation Doping Profile (Gaussian)

$$N(x) = \frac{N_i}{\sqrt{2\pi}\cdot(\Delta R)} \cdot e^{-(x-R)^2/2\Delta R^2}$$

N_i : dose (cm^{-2})

R : range or depth

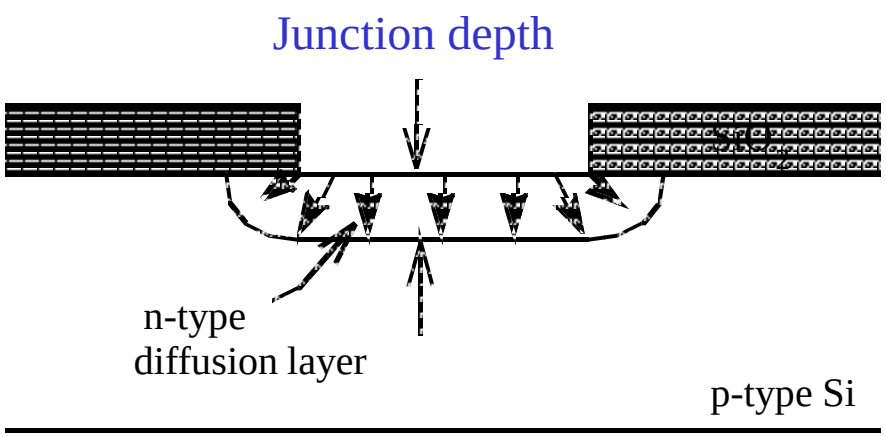
ΔR : spread or sigma



Other Doping Methods

- ***Gas-Source Doping*** : For example, dope Si with P using $POCl_3$.
- ***Solid-Source Doping*** : Dopant diffuses from a doped solid film (SiGe or oxide) into Si.
- ***In-Situ Doping*** : Dopant is introduced while a Si film is being deposited.

3.6 Dopant Diffusion



$$N(x,t) = \frac{N_o}{\sqrt{\pi \cdot Dt}} e^{-x^2/4Dt}$$

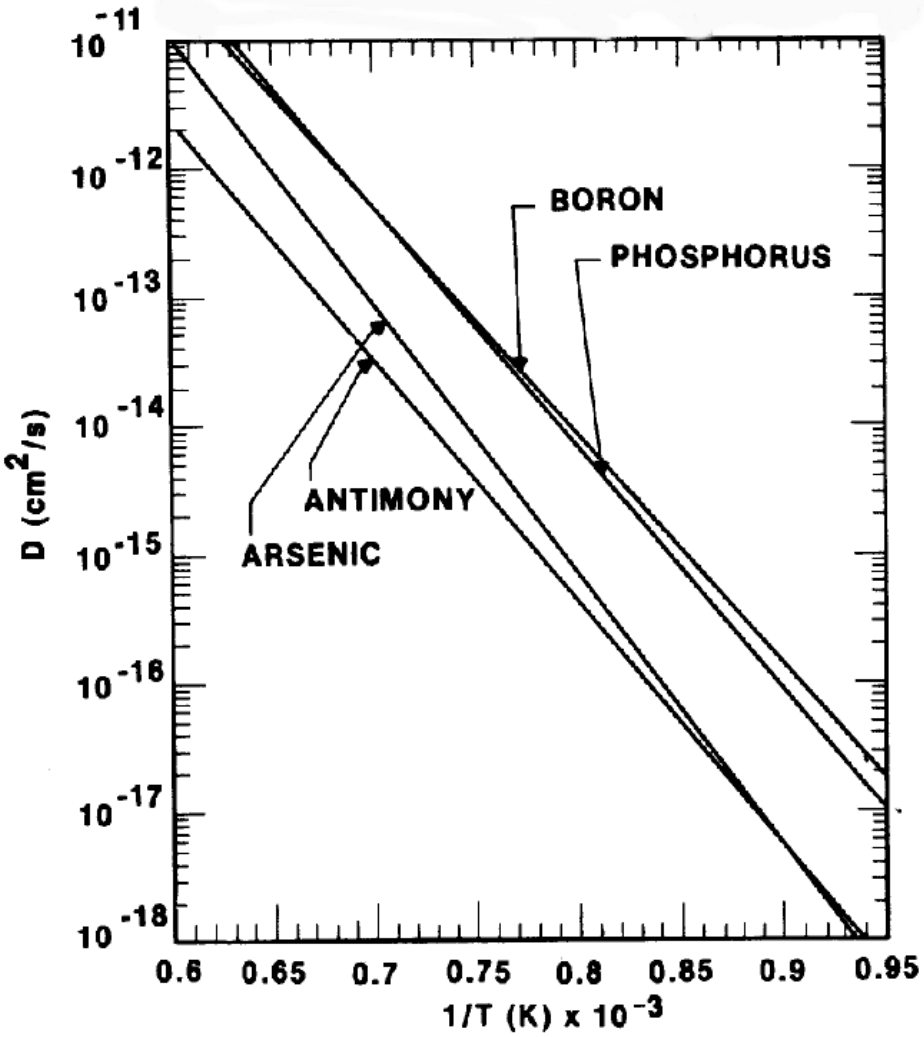
N : N_d or N_a (cm^{-3})

N_o : dopant atoms per cm^2

t : diffusion time

D : diffusivity, $\sqrt{Dt}$ is the approximate distance of dopant diffusion

3.6 Dopant Diffusion



- *D* increases with increasing temperature.

- Some applications need very deep junctions (high *T*, long *t*). Others need very shallow junctions (low *T*, short *t*).

3.6 Dopant Diffusion

Shallow Junction and Rapid Thermal Annealing

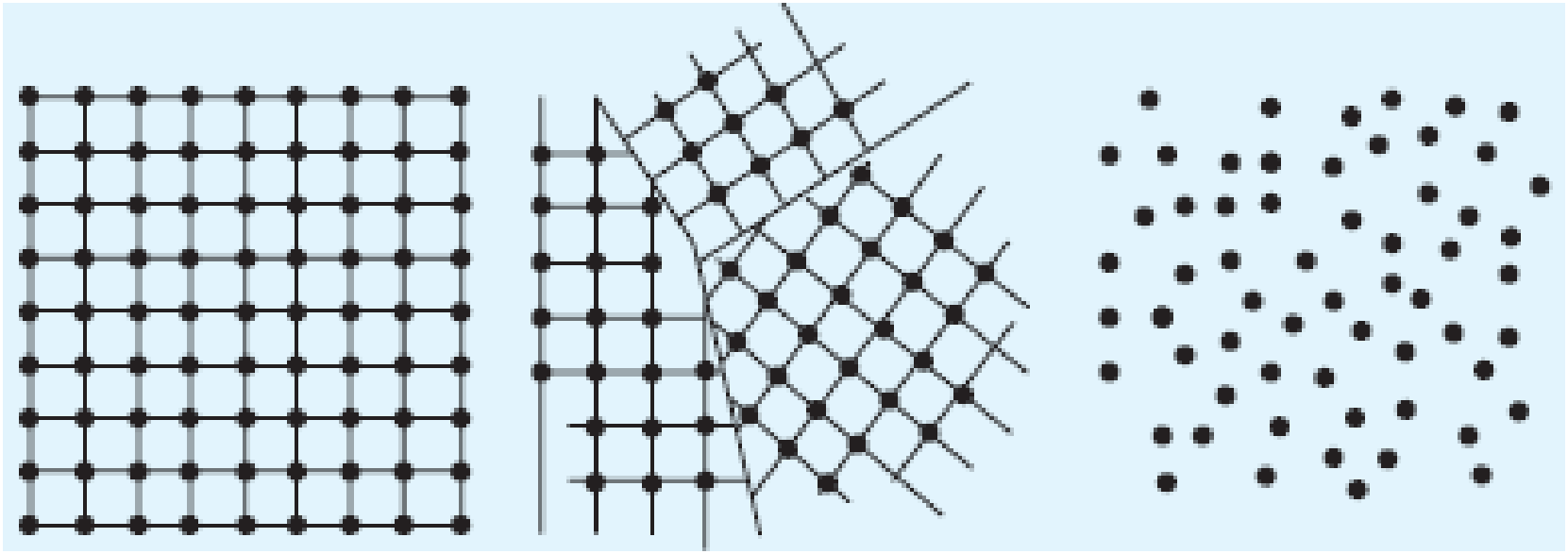
• After ion implantation, thermal annealing is required. Furnace annealing takes minutes and causes too much diffusion of dopants for some applications.

• In rapid thermal annealing (RTA), the wafer is heated to high temperature in seconds by a bank of heat lamps.

•In flash annealing (100mS) and laser annealing (<1uS), dopant ddiffusion is practically eliminated.

3.7 Thin-Film Deposition
Three Kinds of Solid

Crystalline | Polycrystalline | Amorphous



Example: Silicon wafer

Thin film of Si or metal.

Thin film of SiO_2 or Si_3N_4.

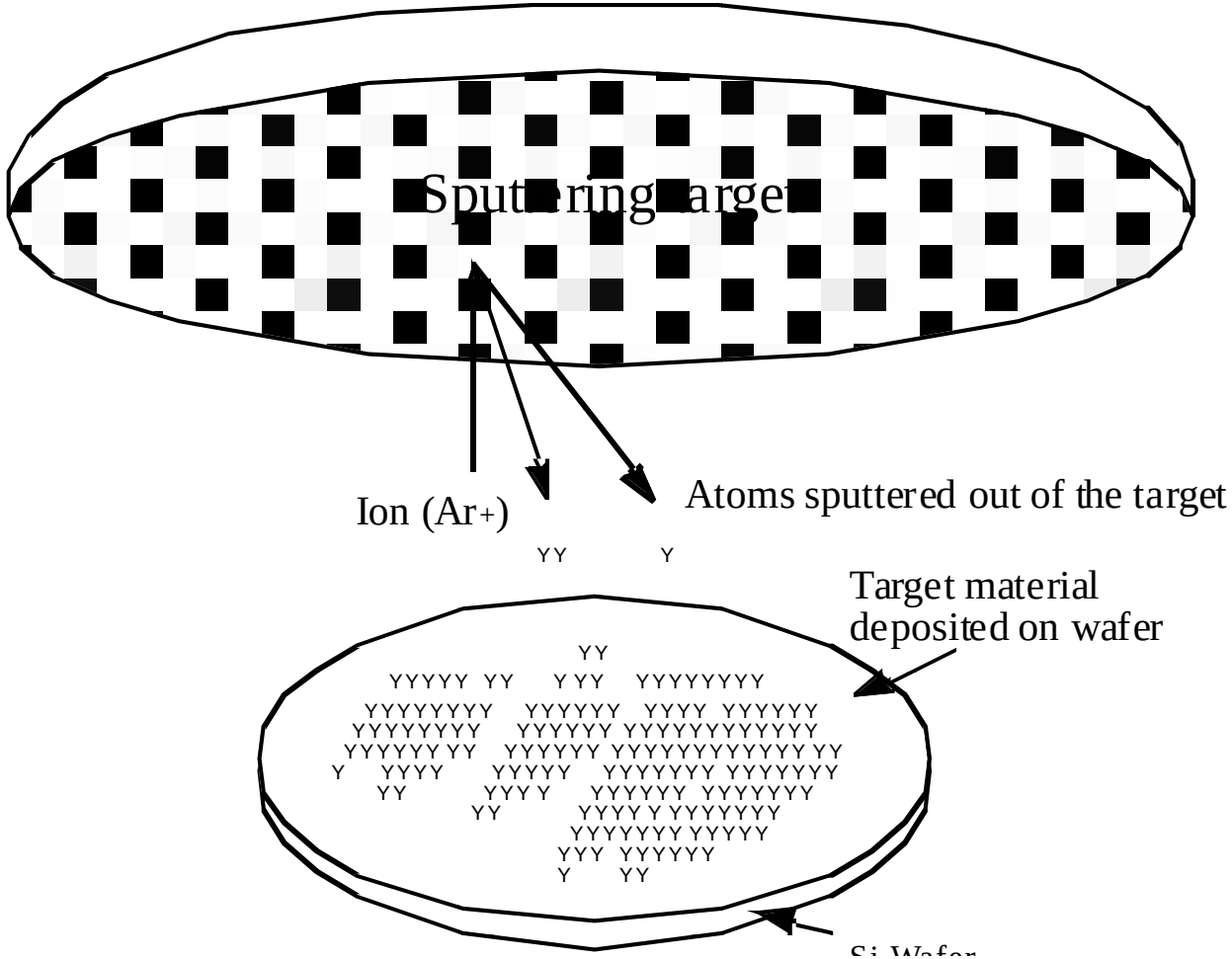
3.7 Thin-Film Deposition

Examples of thin films in integrated circuits

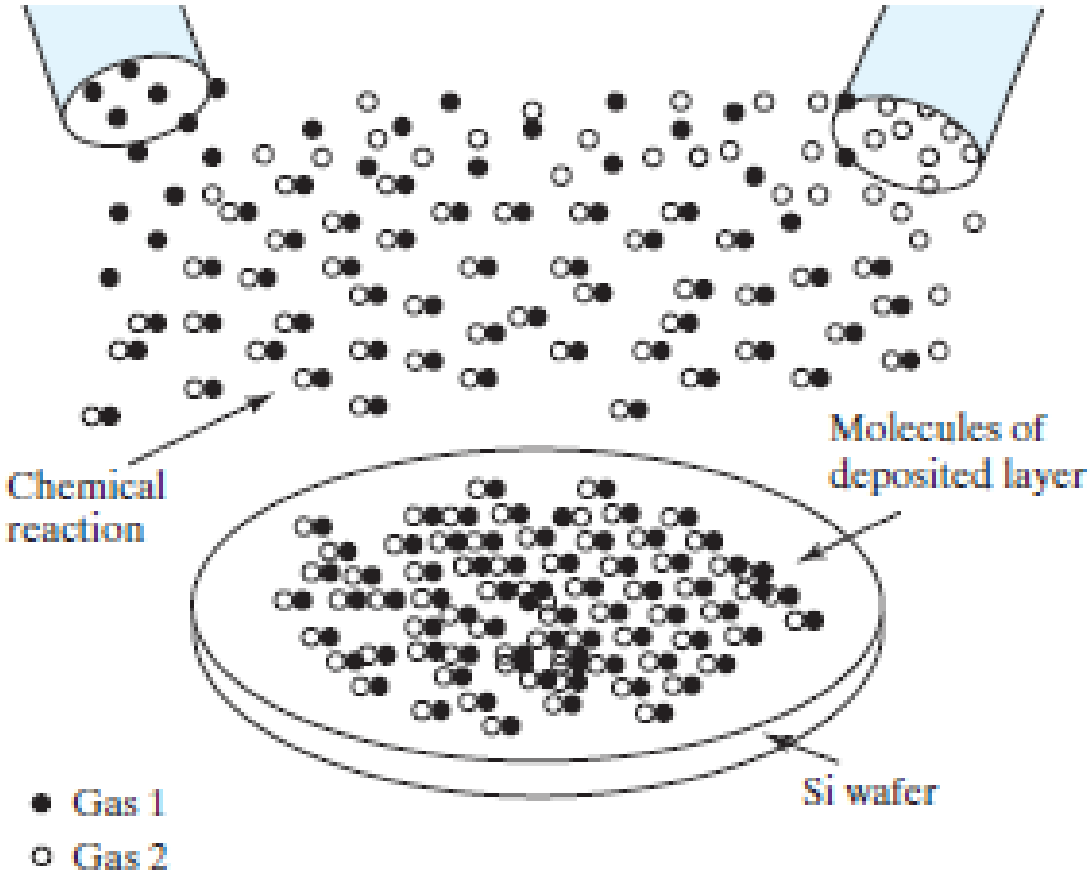
- Advanced MOSFET gate dielectric
- Poly-Si film for transistor gates
- Metal layers for interconnects
- Dielectric between metal layers
- Encapsulation of IC

3.7.1 Sputtering

Schematic Illustration of Sputtering Process



3.7.2 Chemical Vapor Deposition (CVD)



Thin film is formed from gas phase components.

Some Chemical Reactions of CVD

Poly-Si : $SiH_4\,(g) \longrightarrow Si\,(s) + 2H_2\,(g)$

Si3N4 : $3SiH_2Cl_2\,(g)+4NH_3\,(g) \longrightarrow Si_3N_4\,(s)+6HCl(g)+6H_2\,(g)$

SiO2 : $SiH_4\,(g) + O_2\,(g) \longrightarrow SiO_2\,(s) + 2H_2\,(g)$

or

$SiH_2Cl_2\,(g)+2N_2O\,(g) \longrightarrow SiO_2\,(s)+2HCl\,(g)+2N_2\,(g)$

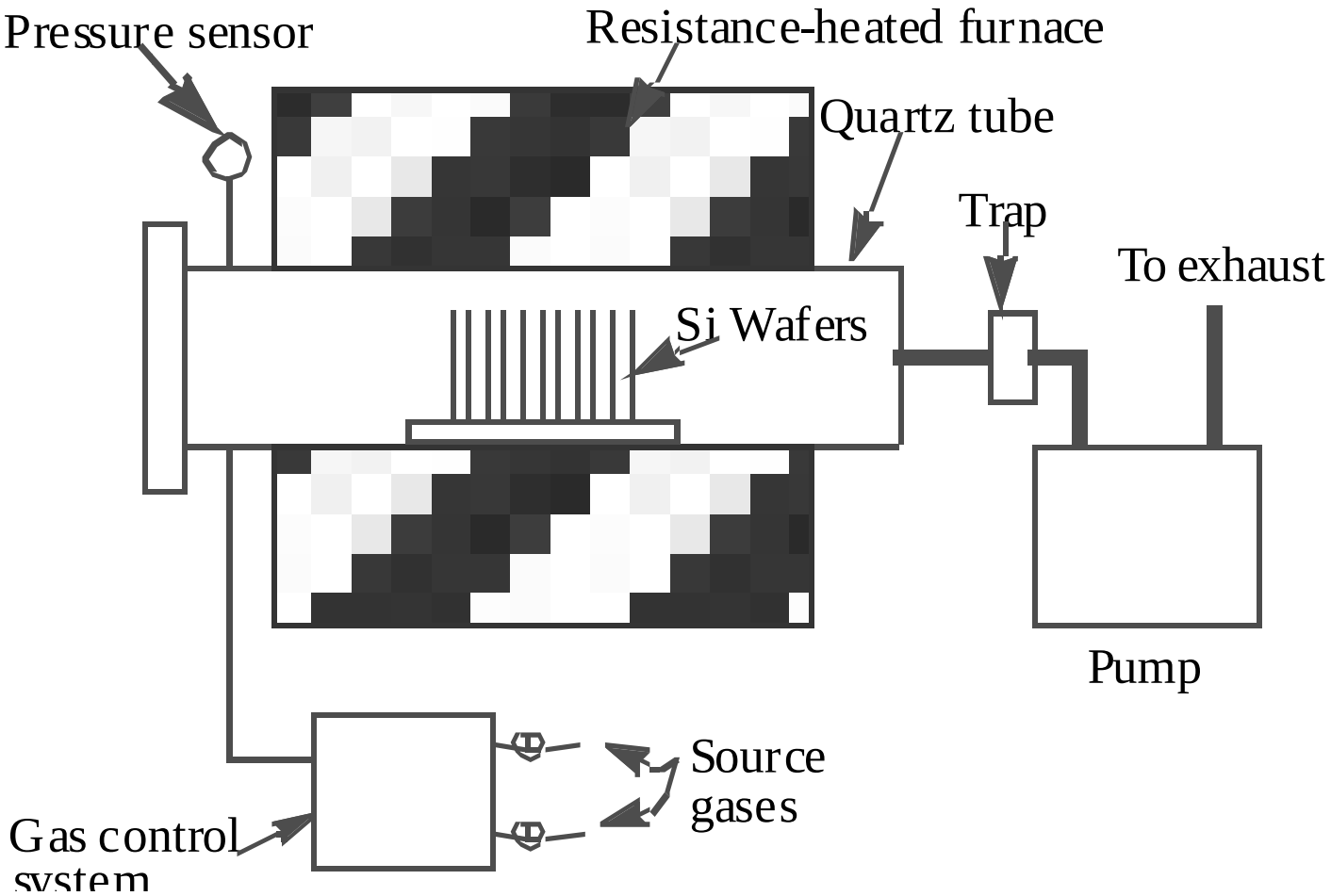
3.7.2 Chemical Vapor Deposition (CVD)

Two types of CVD equipment:

- **LPCVD (Low Pressure CVD)** : Good uniformity. Used for poly-Si, oxide, nitride.

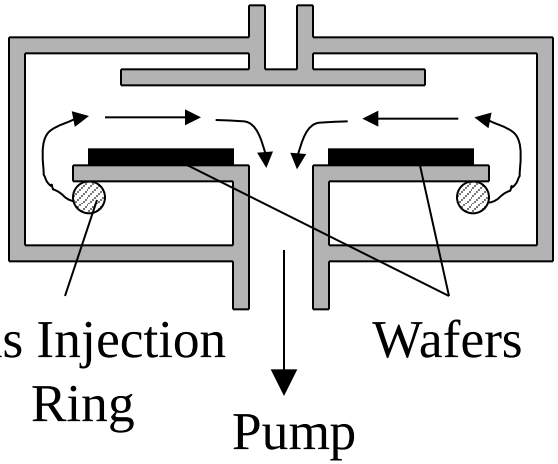
- **PECVD (Plasma Enhanced CVD)** : Low temperature process and high deposition rate. Used for oxide, nitride, etc.

3.7.2 Chemical Vapor Deposition (CVD)



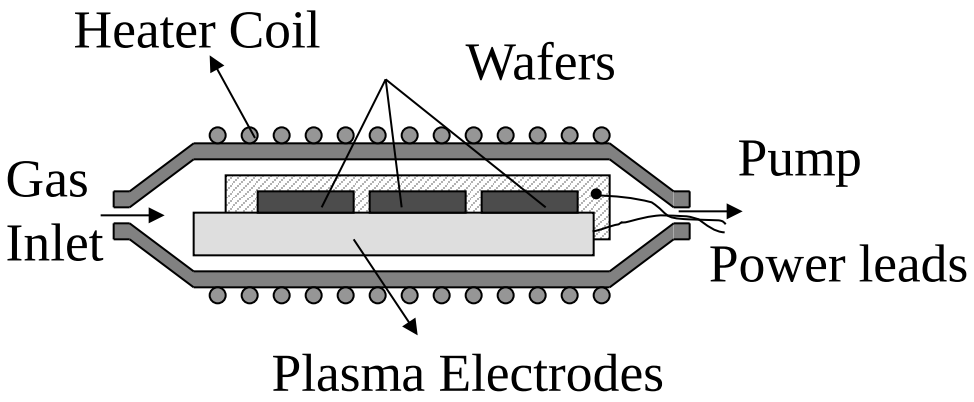
LPCVD Systems

3.7.2 Chemical Vapor Deposition (CVD)



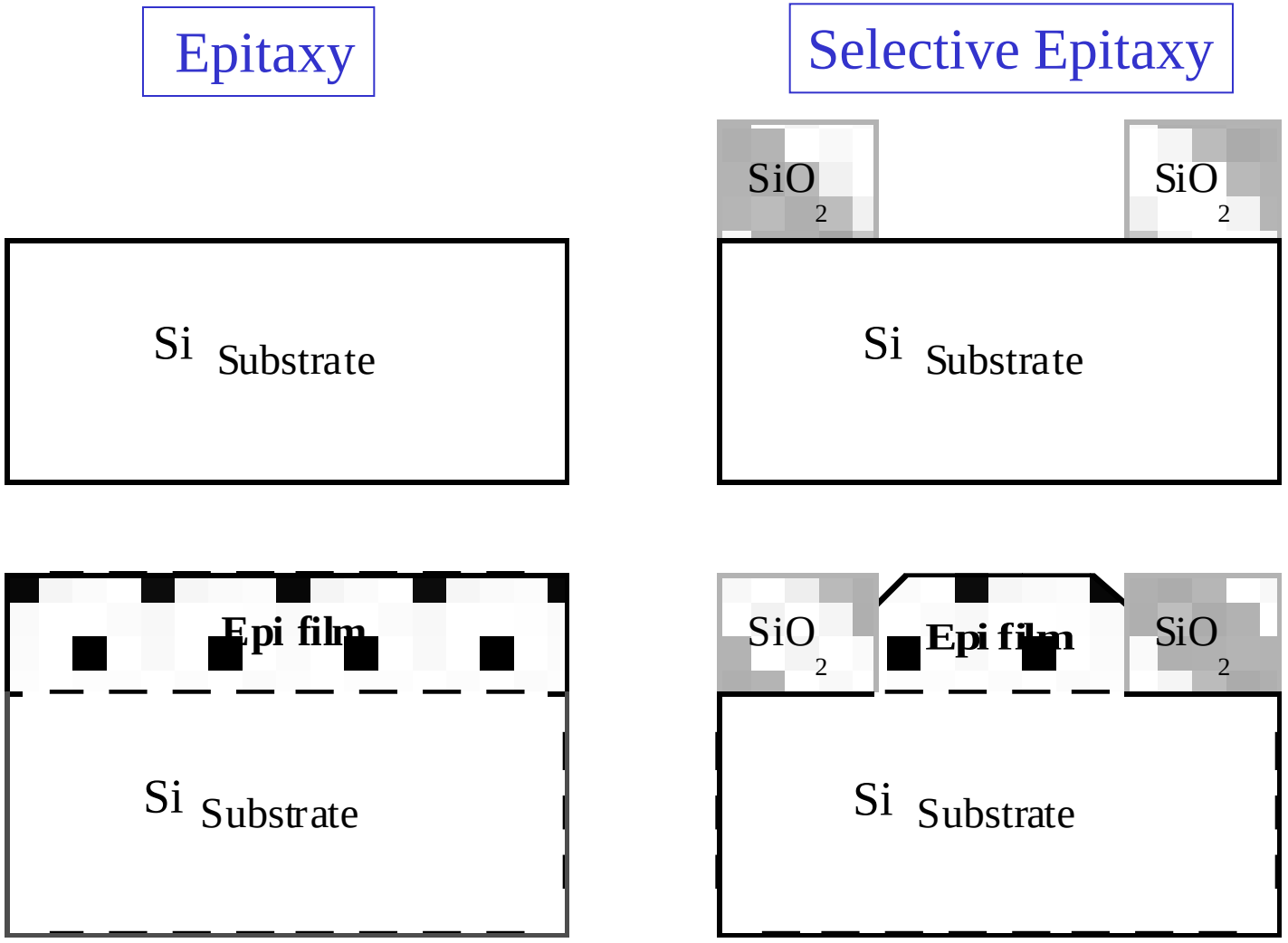
Cold Wall Parallel Plate

Hot Wall Parallel Plate

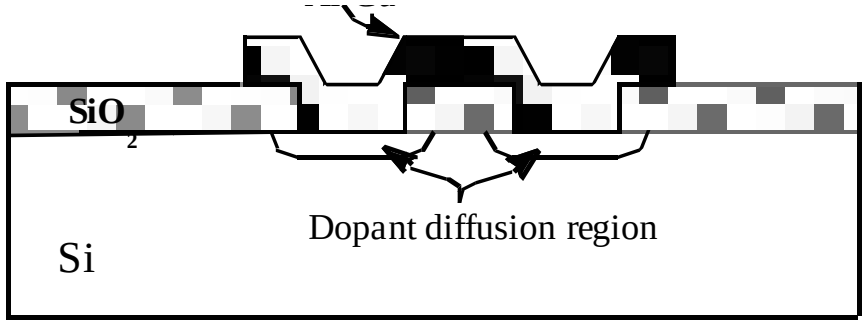


PECVD Systems

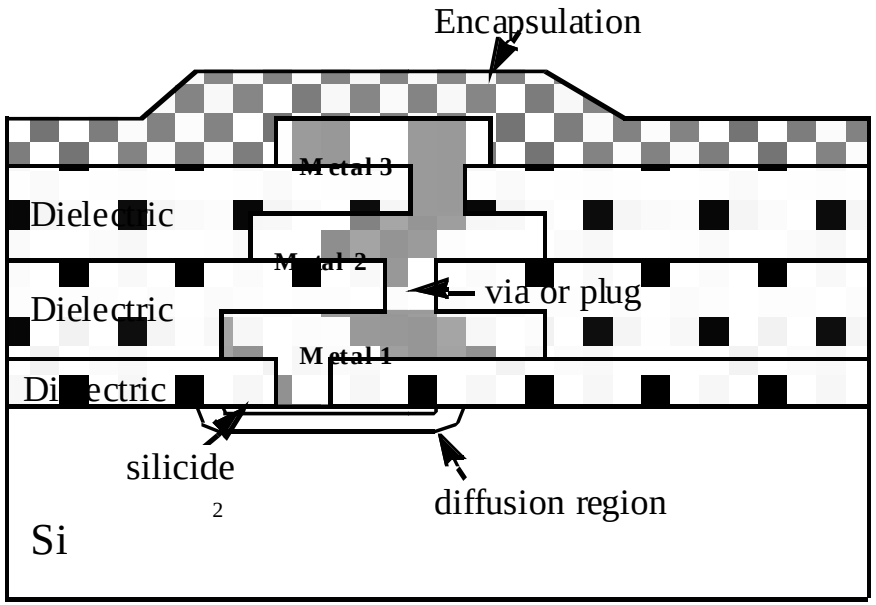
3.7.3 Epitaxy *(Deposition of Single-Crystalline Film)*



3.8 Interconnect – The Back-end Process



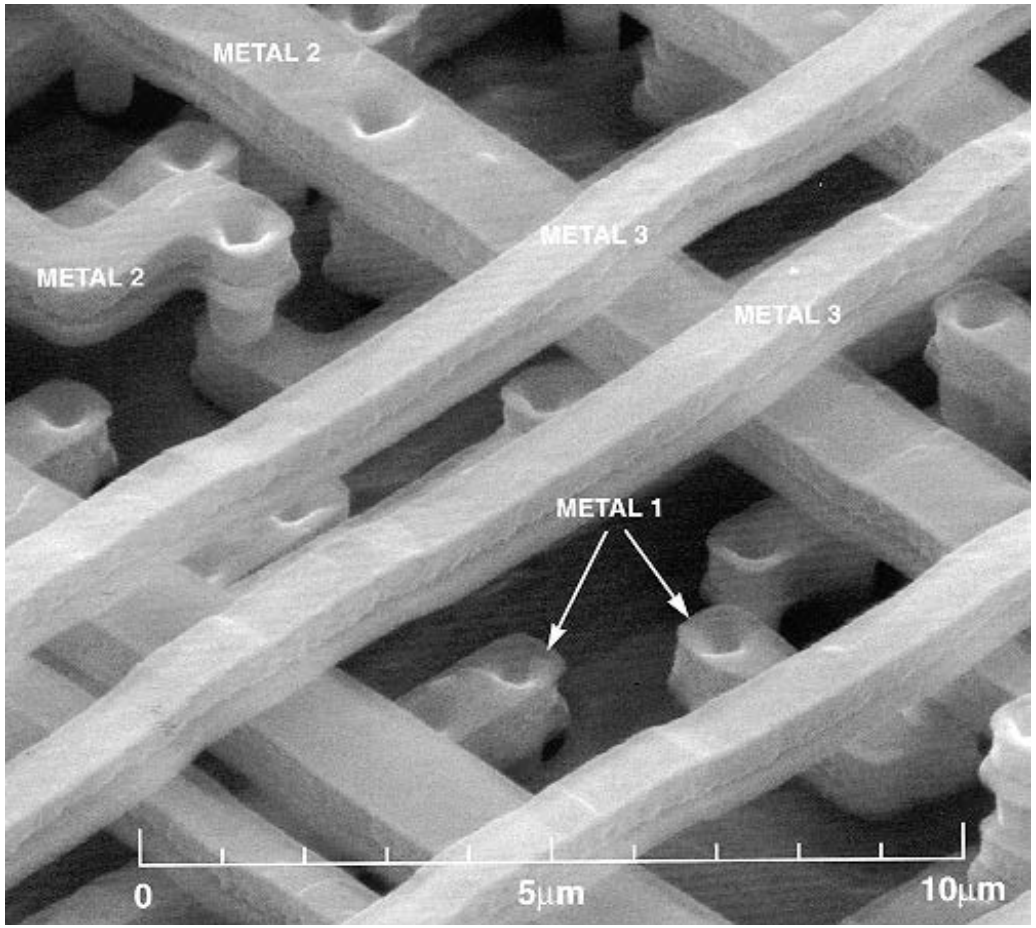
(a)



(b)

3.8 Interconnect – The Back-end Process

SEM: Multi-Level Interconnect (after removing the dielectric)



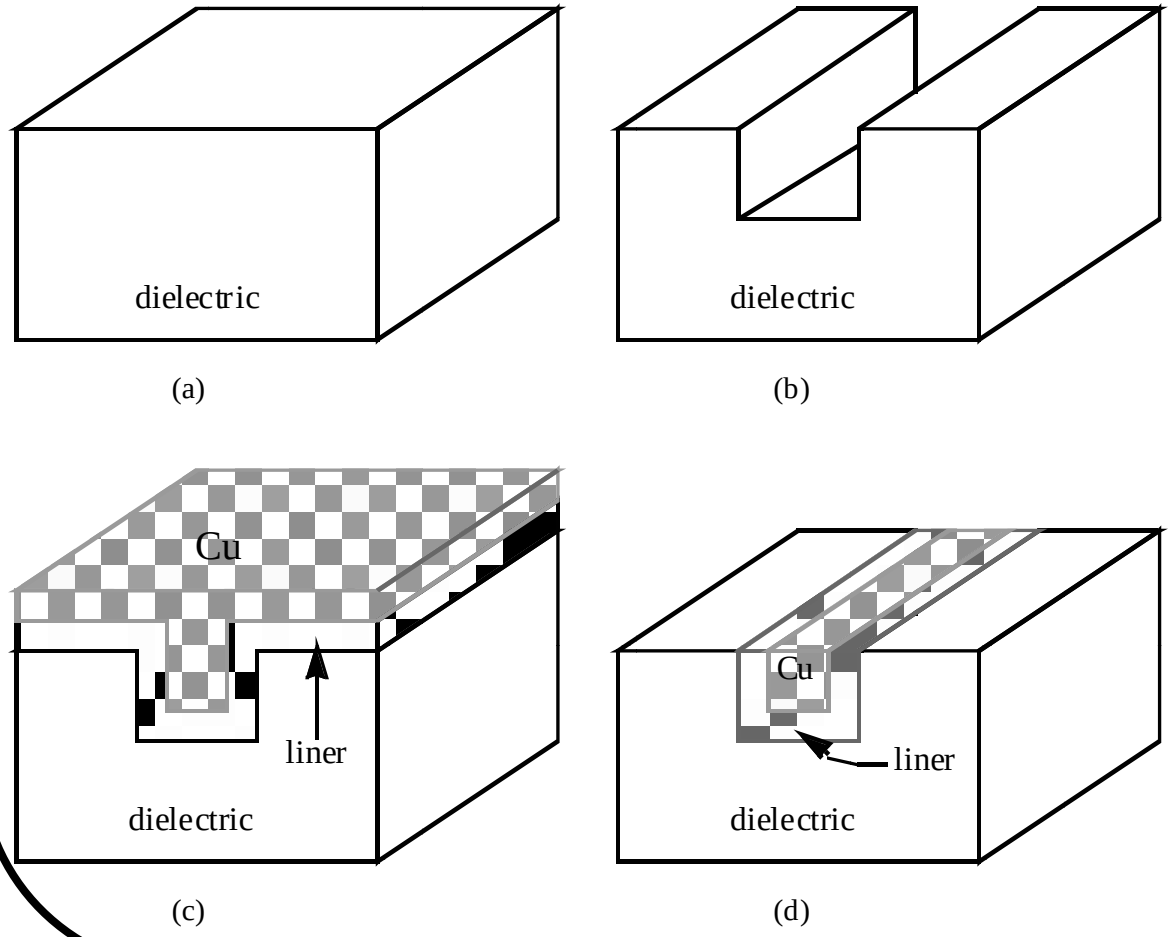
3.8 Interconnect – The Back-end Process

Copper Interconnect

- Al interconnect is prone to voids formation by electromigration.
- Cu has excellent electromigration reliability and 40% lower resistance than Al.
- Because dry etching of copper is difficult (copper etching products tend to be non-volatile), copper patterns are defined by a ***damascene*** process.

3.8 Interconnect – The Back-end Process

Copper Damascene Process



- Chemical-Mechanical Polishing (CMP) removes unwanted materials.
- Barrier liner prevents Cu diffusion.

3.8 Interconnect – The Back-end Process

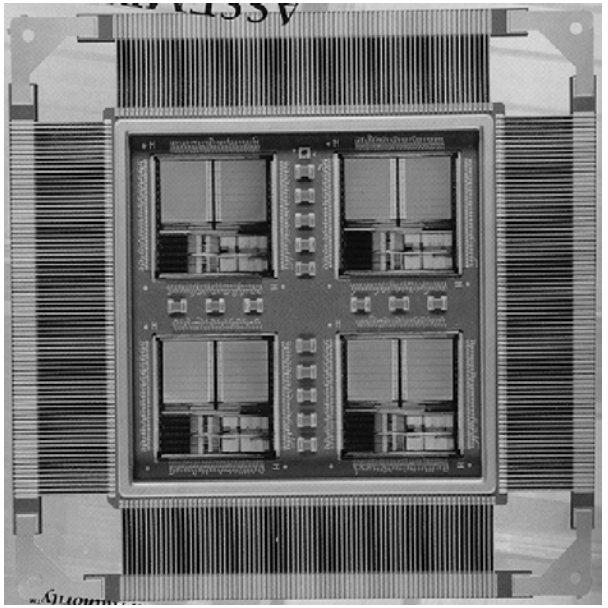
Planarization

- A flat surface is highly desirable for subsequent lithography and etching.

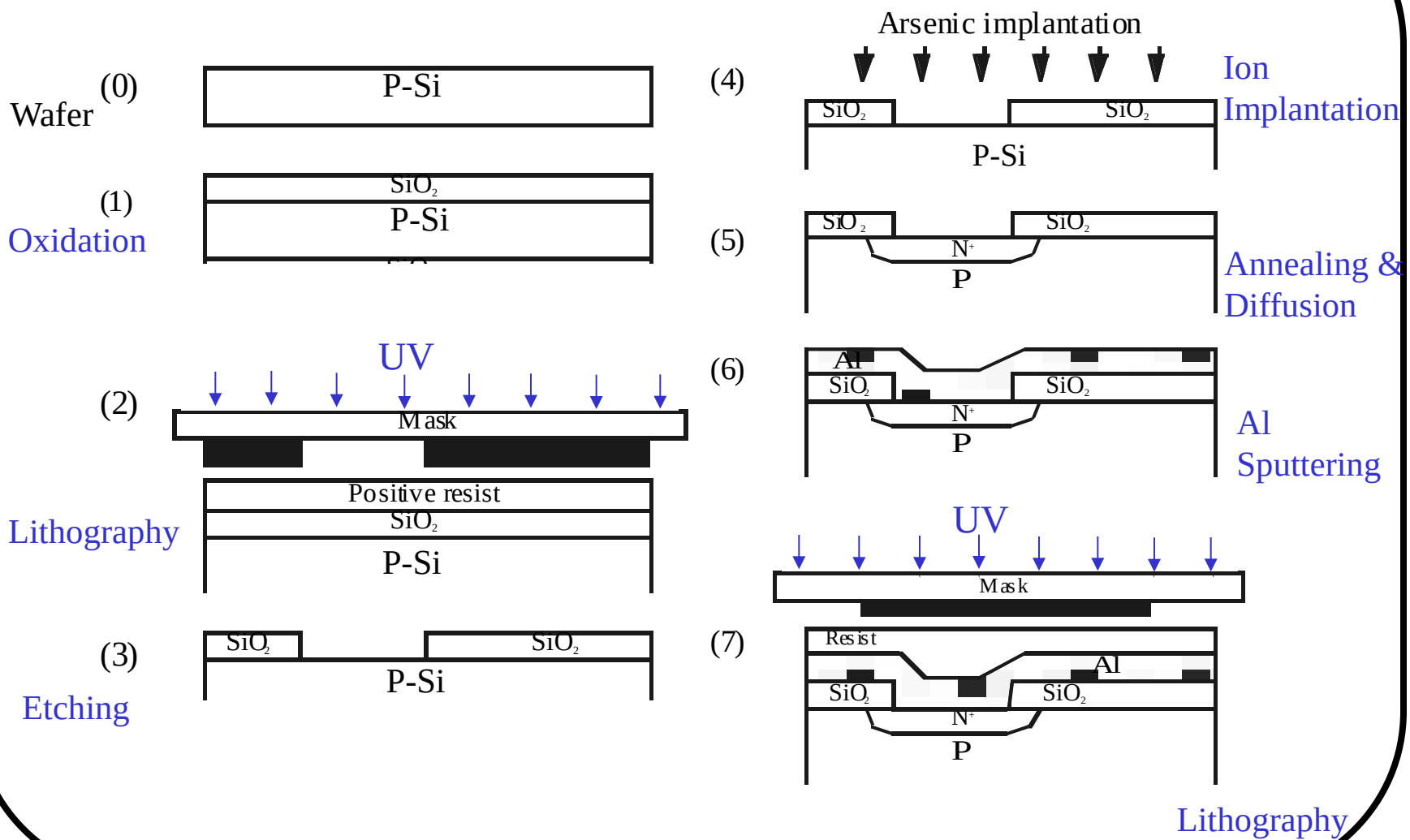
- CMP (Chemical-Mechanical Polishing) is used to planarize each layer of dielectric in the interconnect system. Also used in the front-end process.

3.9 Testing, Assembly, and Qualification

- Wafer acceptance test
- Die sorting
- Wafer sawing or laser cutting
- Packaging
- Flip-chip solder bump technology
- Multi-chip modules
- Burn-in
- Final test
- Qualification

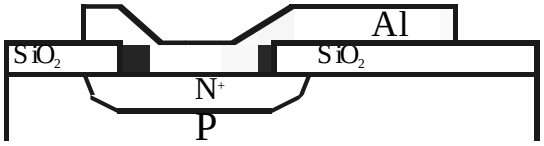


3.10 Chapter Summary–A Device Fabrication Example

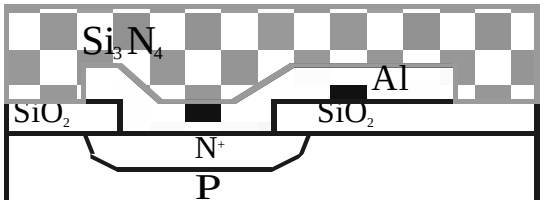


3.10 Chapter Summary–A Device Fabrication Example

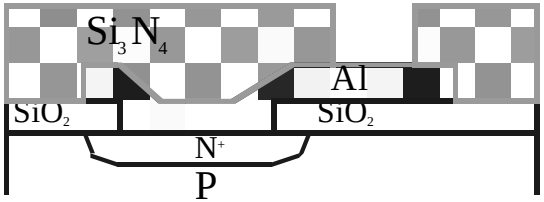
Metal etching (8)



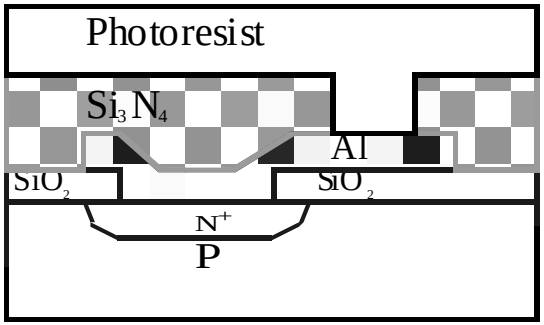
CVD nitride deposition (9)



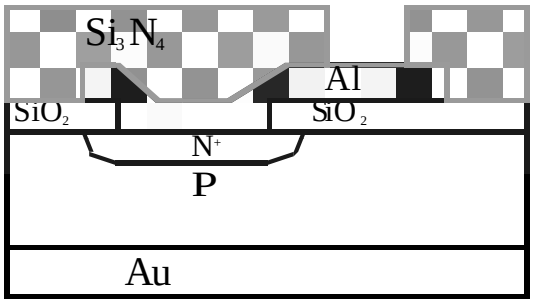
Lithography and etching (10)



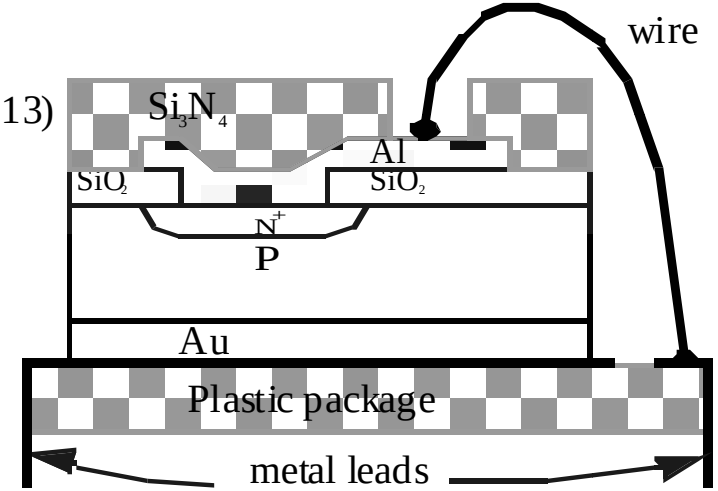
Back Side milling (11)



(12) Back side metallization



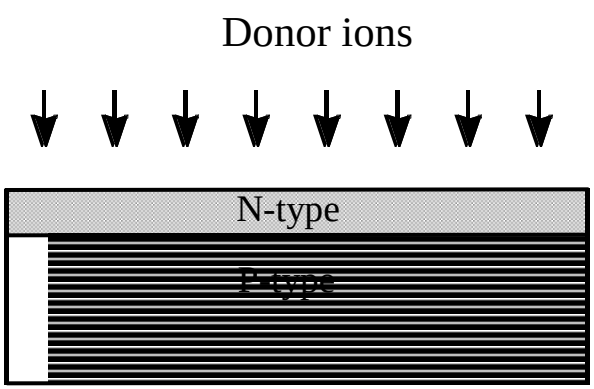
(13)

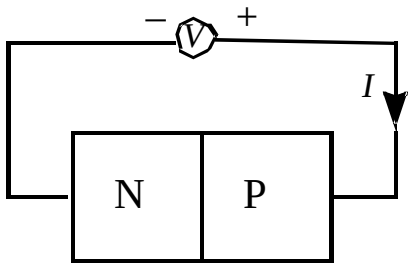


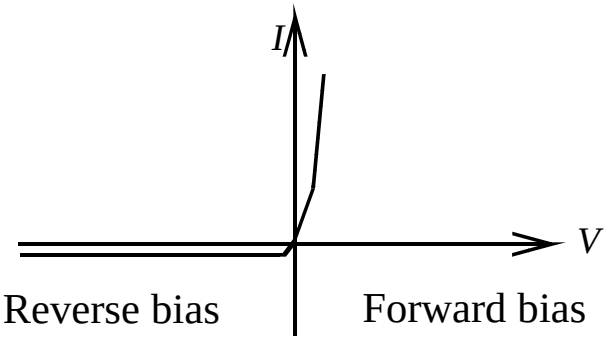
Dicing, wire bonding, and packaging

Chapter 4 PN and Metal-Semiconductor Junctions

4.1 Building Blocks of the PN Junction Theory



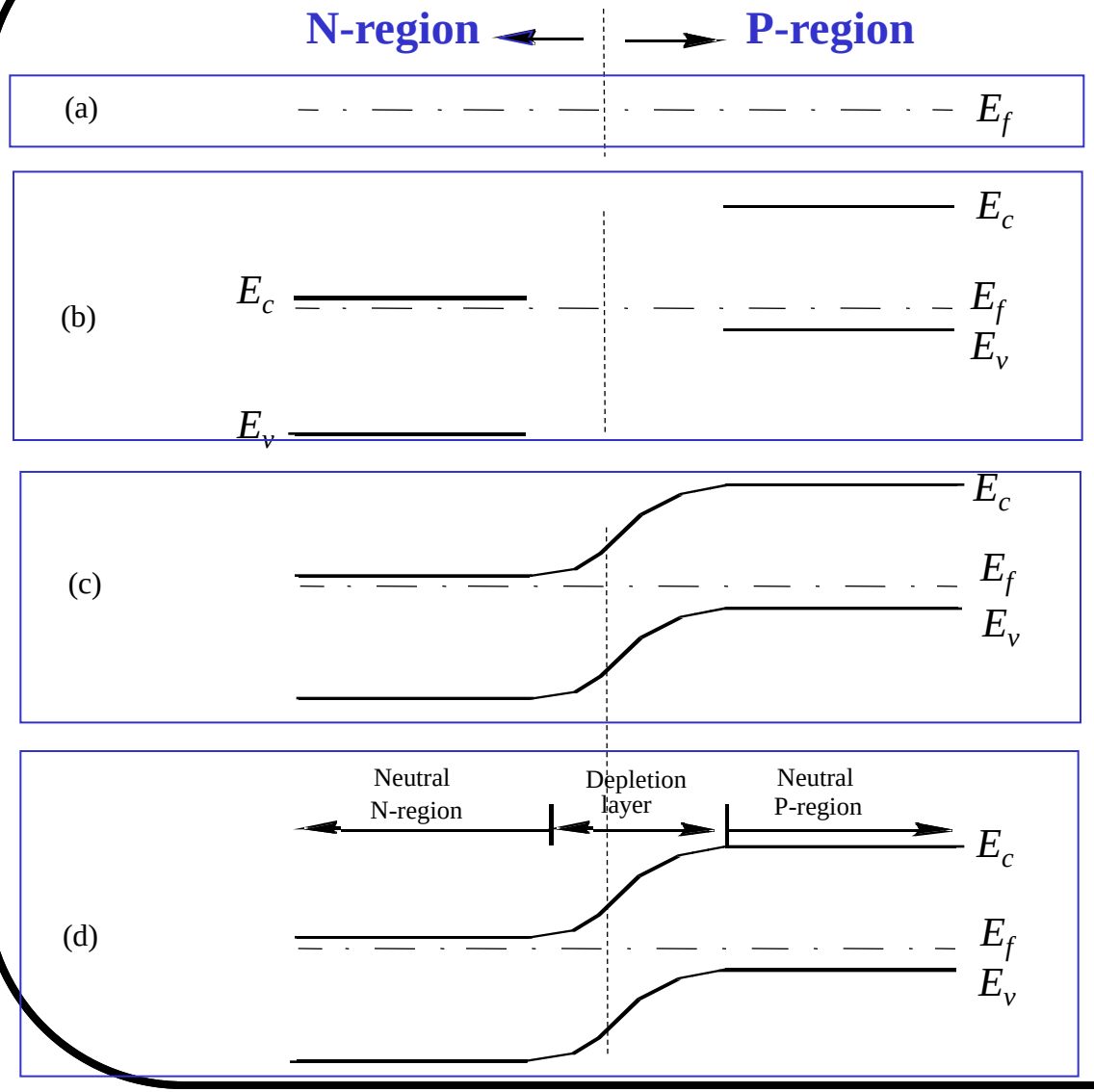




diode symbol

PN junction is present in perhaps every semiconductor device.

4.1.1 Energy Band Diagram of a PN Junction



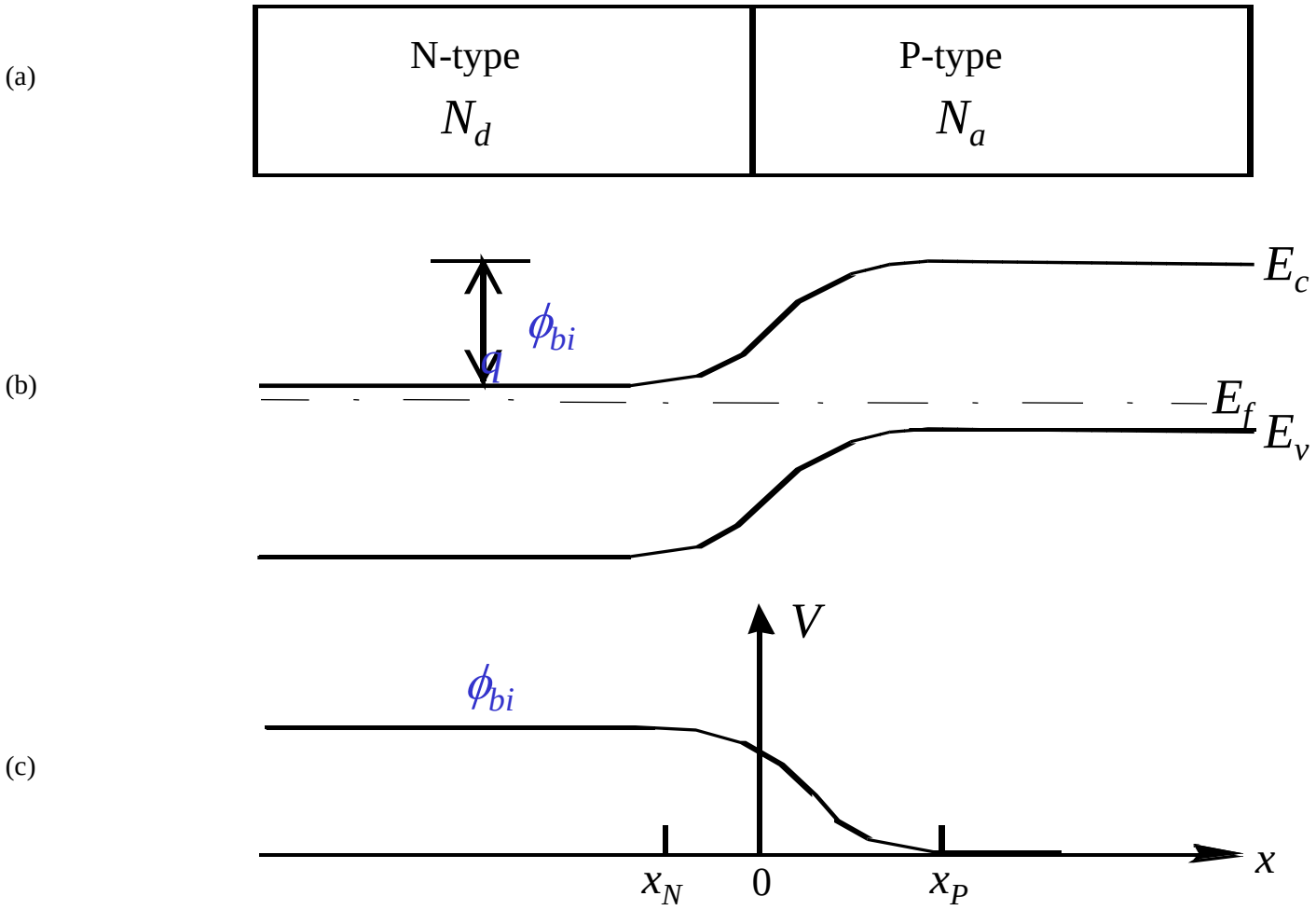
E_f is constant at equilibrium

E_c and E_v are known relative to E_f

E_c and E_v are smooth, the exact shape to be determined.

A depletion layer exists at the PN junction where $n \approx 0$ and $p \approx 0$.

4.1.2 Built-in Potential



Can the built-in potential be measured with a voltmeter?

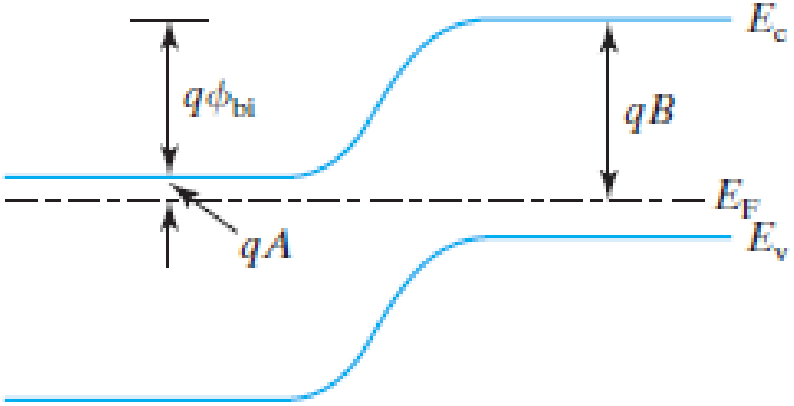
4.1.2 Built-in Potential

N-region $n = N_d = N_c e^{-qA/kT} \Rightarrow A = \frac{kT}{q} \ln \frac{N_c}{N_d}$

P-region $n = \frac{n_i^2}{N_a} = N_c e^{-qB/kT} \Rightarrow B = \frac{kT}{q} \ln \frac{N_c N_a}{n_i^2}$

$$\phi_{bi} = B - A = \frac{kT}{q}\left(\ln \frac{N_c N_a}{n_i^2} - \ln \frac{N_c}{N_d} \right)$$

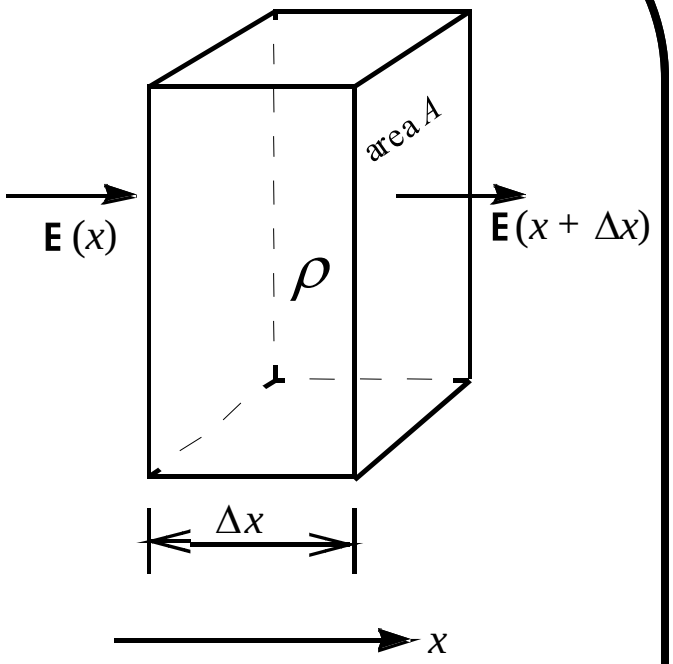
$$\phi_{bi} = \frac{kT}{q} \ln \frac{N_d N_a}{n_i^2}$$



4.1.3 Poisson's Equation

Gauss's Law:

$$\varepsilon_s \mathcal{E}(x+\Delta x)A - \varepsilon_s \mathcal{E}(x)A = \rho \Delta x A$$



ε_s: permittivity (~12ε_o for Si)

ρ: charge density (C/cm^3)

$$\frac{\mathcal{E}(x+\Delta x)-\mathcal{E}(x)}{\Delta x} = \frac{\rho}{\varepsilon_s}$$

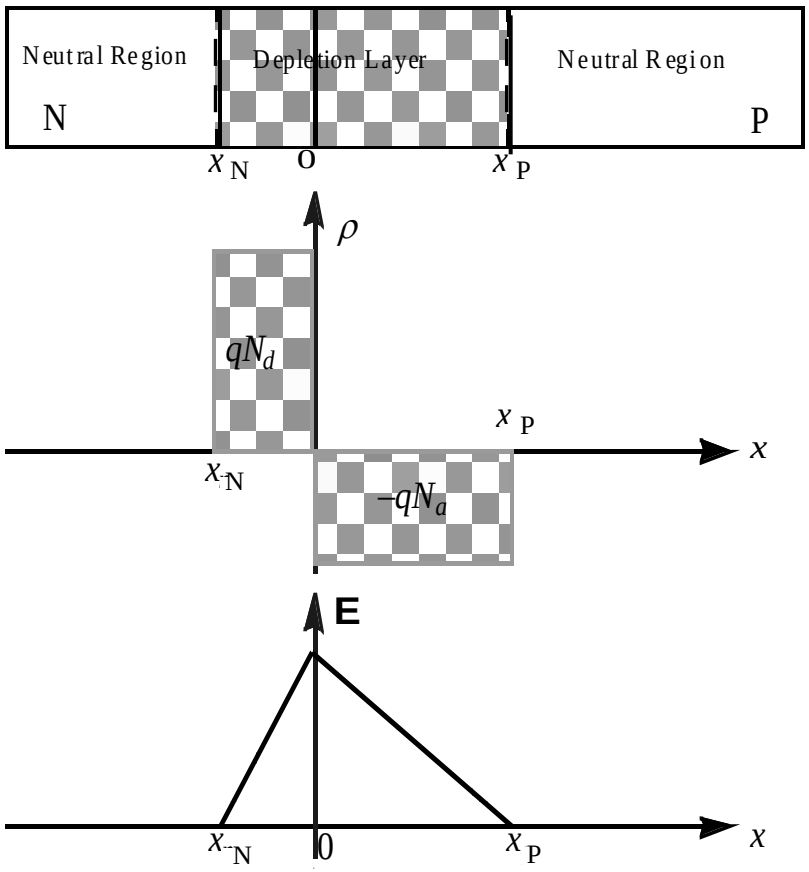
$$\frac{d\mathcal{E}}{dx} = \frac{\rho}{\varepsilon_s}$$

$$\frac{d^2V}{dx^2} = -\frac{d\mathcal{E}}{dx} = -\frac{\rho}{\varepsilon_s}$$

Poisson's equation

4.2 Depletion-Layer Model

4.2.1 Field and Potential in the Depletion Layer



On the ***P-side*** of the depletion layer, $\rho = -qN_a$

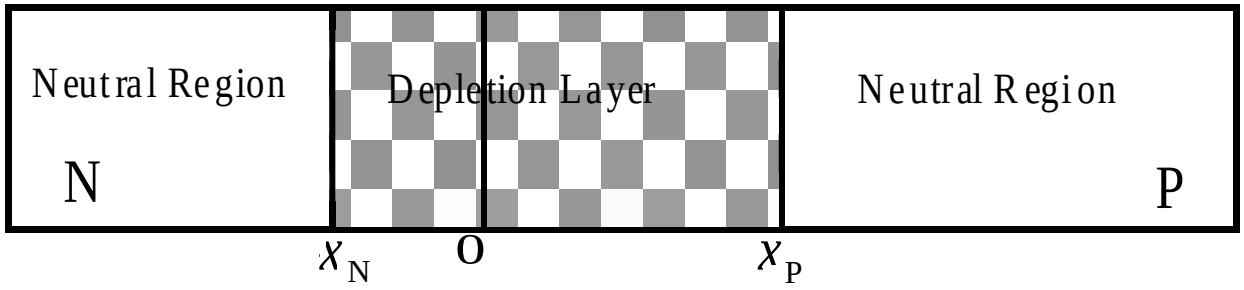
$$\frac{d\mathcal{E}}{dx} = -\frac{qN_a}{\varepsilon_s}$$

$$\mathcal{E}(x) = -\frac{qN_a}{\varepsilon_s}x + C_1 = \boxed{\frac{qN_a}{\varepsilon_s}(x_P - x)}$$

On the ***N-side***, $\rho = qN_d$

$$\boxed{\mathcal{E}(x) = \frac{qN_d}{\varepsilon_s}(x - x_N)}$$

4.2.1 Field and Potential in the Depletion Layer



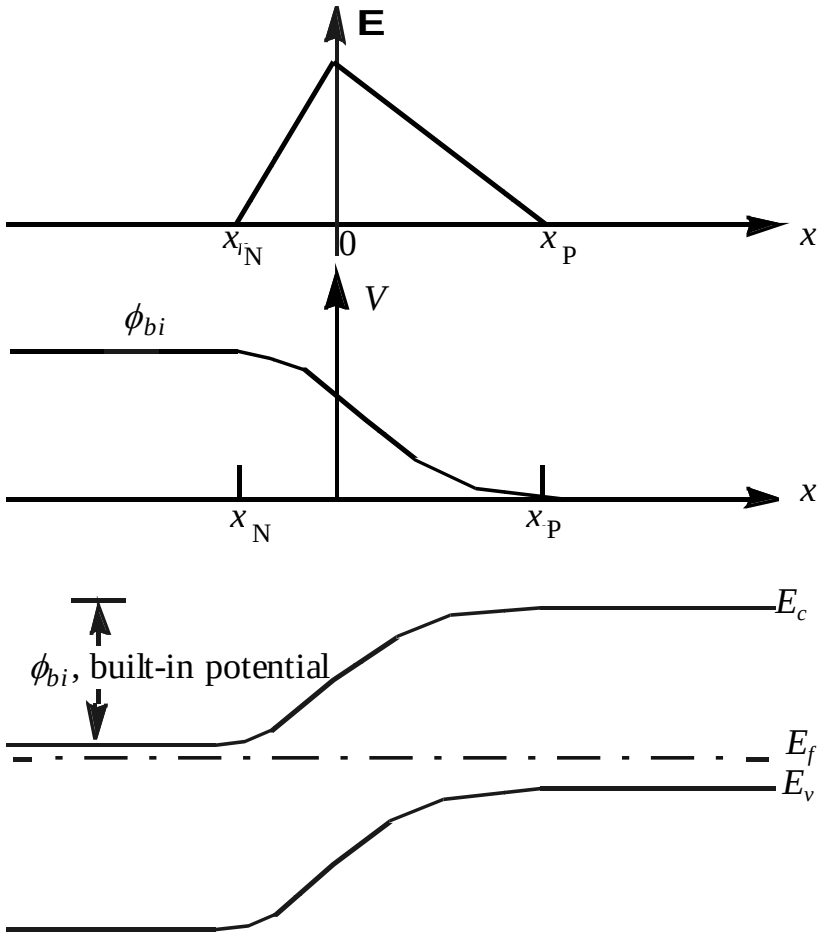
The electric field is continuous at $x = 0$.

$$N_a |x_P| = N_d |x_P|$$

Which side of the junction is depleted more?

A one-sided junction is called a ***N⁺P junction*** or ***P⁺N junction***

4.2.1 Field and Potential in the Depletion Layer



On the P-side,

$$V(x) = \frac{qN_a}{2\varepsilon_s}(x_P - x)^2$$

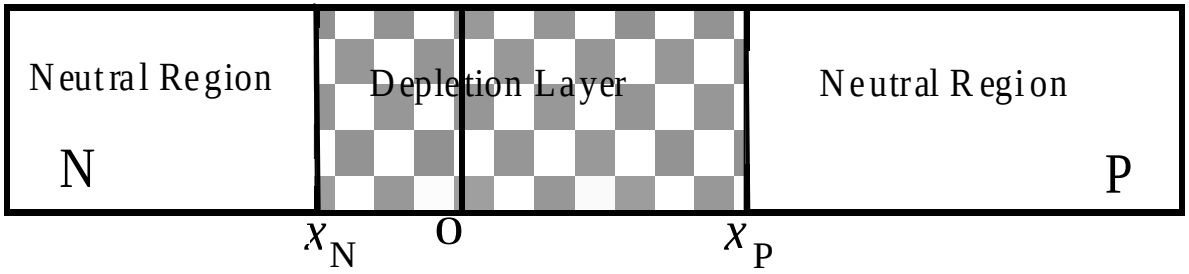
Arbitrarily choose the voltage at $x = x_P$ as $V = 0$.

On the N-side,

$$V(x) = D - \frac{qN_d}{2\varepsilon_s}(x - x_N)^2$$

$$= \phi_{bi} - \frac{qN_d}{2\varepsilon_s}(x - x_N)^2$$

4.2.2 Depletion-Layer Width



V is continuous at $x = 0$ ⟶

$$x_P - x_N = W_{dep} = \sqrt{\frac{2\varepsilon_s \phi_{bi}}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)}$$

If $N_a >> N_d$, as in a P^+N junction,

$$W_{dep} = \sqrt{\frac{2\varepsilon_s \phi_{bi}}{qN_d}} \approx |x_N|$$

$$|x_P| = |x_N| N_d / N_a \cong 0$$

What about a N^+P junction?

$$W_{dep} = \sqrt{2\varepsilon_s \phi_{bi} / qN} \quad \text{where} \quad \frac{1}{N} = \frac{1}{N_d} + \frac{1}{N_a} \approx \frac{1}{\textit{lighter dopant density}}$$

EXAMPLE: *A P^+N junction has $N_a=10^{20}\,cm^{-3}$ and $N_d=10^{17}cm^{-3}$. What is a) its built in potential, b)W_{dep} , c)x_N , and d) x_P ?*

Solution:

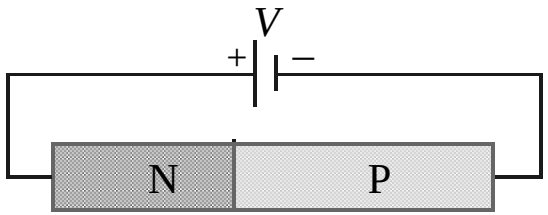
a) $$\phi_{bi} = \frac{kT}{q}\ln\frac{N_d N_a}{n_i^{\,2}} = 0.026\text{V}\ln\frac{10^{20}\times 10^{17}\,\text{cm}^{-6}}{10^{20}\,\text{cm}^{-6}} \approx 1\,\text{V}$$

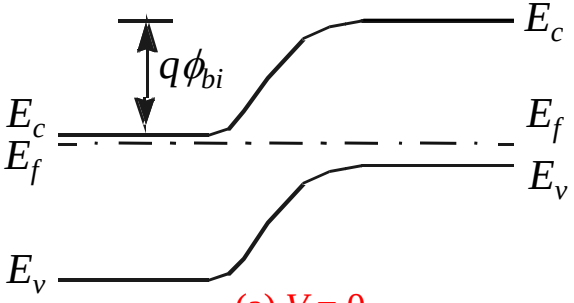
b) $$W_{dep} \approx \sqrt{\frac{2\varepsilon_s\phi_{bi}}{qN_d}} = \left(\frac{2\times 12\times 8.85\times 10^{-14}\times 1}{1.6\times 10^{-19}\times 10^{17}}\right)^{1/2} = 0.12\,\mu\text{m}$$

c) $$\left|x_N\right| \approx W_{dep} = 0.12\,\mu\text{m}$$

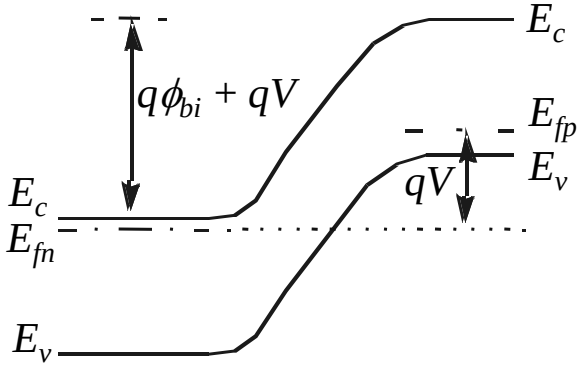
d) $$\left|x_P\right| = \left|x_N\right| N_d / N_a = 1.2\times 10^{-4}\mu\text{m} = 1.2\ \text{Å} \approx 0$$

4.3 Reverse-Biased PN Junction





(a) $V = 0$



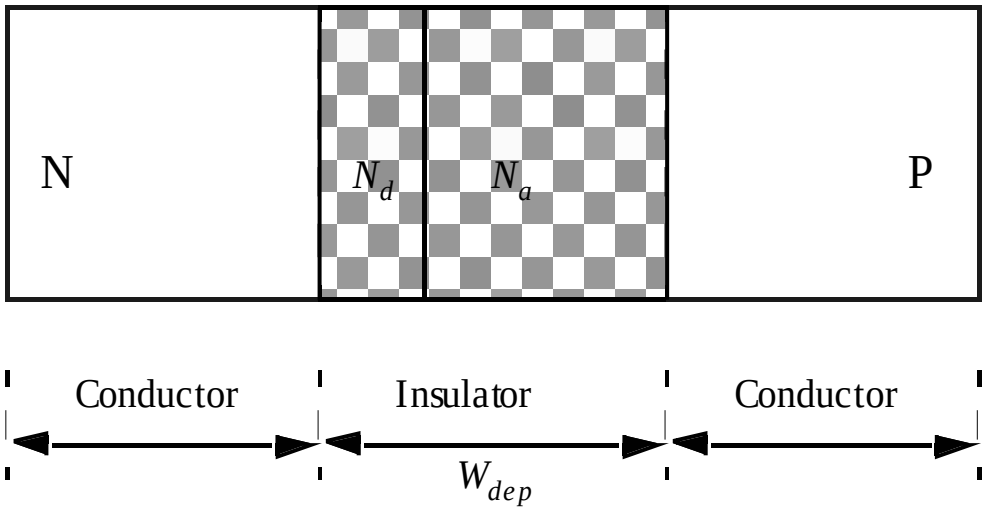
(b) reverse-biased

$$W_{dep} = \sqrt{\frac{2\varepsilon_s(\phi_{bi} + |V_r|)}{qN}} = \sqrt{\frac{2\varepsilon_s \cdot potential\ barrier}{qN}}$$

$$\frac{1}{N} = \frac{1}{N_d} + \frac{1}{N_a} \approx \frac{1}{lighter\ \ dopant\ \ density}$$

- ***Does the depletion layer widen or shrink with increasing reverse bias?***

4.4 Capacitance-Voltage Characteristics



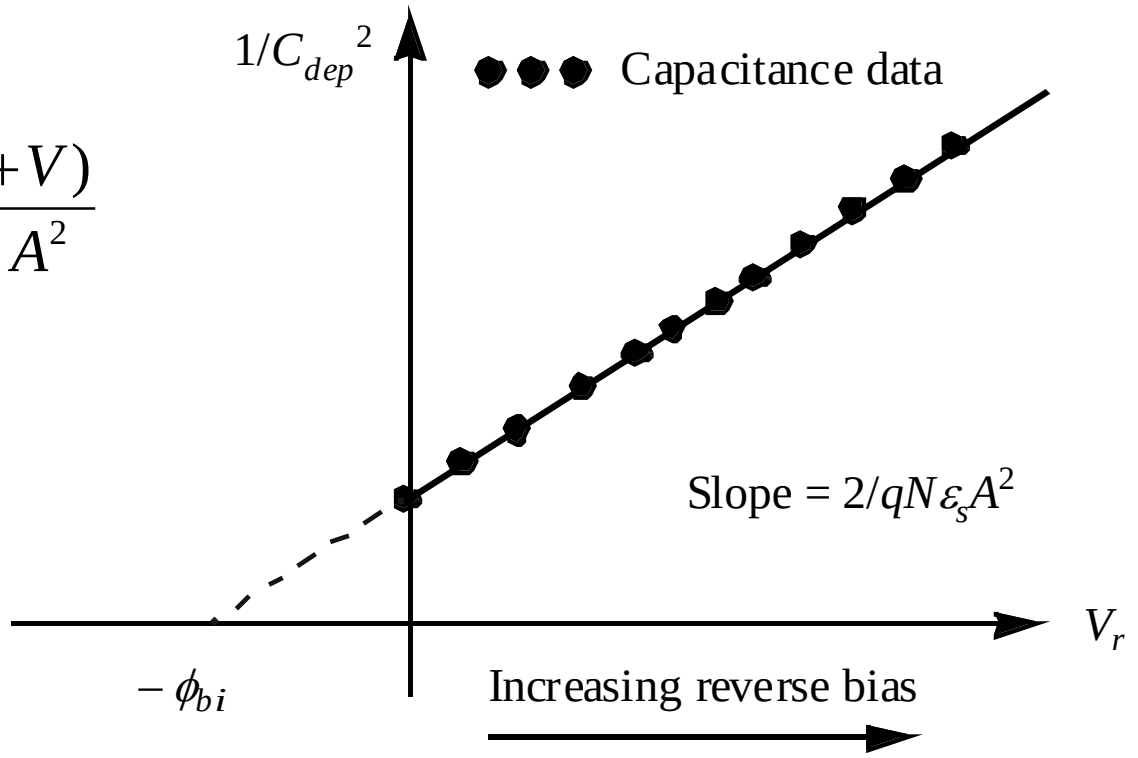
Reverse biased PN junction is a capacitor.

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}}$$

- Is C_{dep} a good thing?
- How to minimize junction capacitance?

4.4 Capacitance-Voltage Characteristics

$$\frac{1}{C_{dep}^{\ 2}} = \frac{W_{dep}^{\ 2}}{A^2 \varepsilon_s^{\ 2}} = \frac{2(\phi_{bi} + V)}{qN\varepsilon_S A^2}$$



- From this C-V data can N_a and N_d be determined?

EXAMPLE: *If the slope of the line in the previous slide is $2x10^{23}\ F^{-2}\ V^{-1}$, the intercept is 0.84V, and A is 1 μm^2, find the lighter and heavier doping concentrations N_l and N_h.*

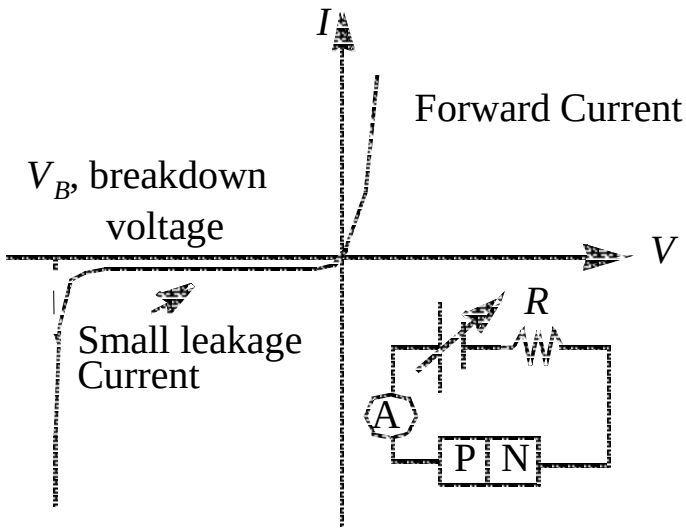
Solution:

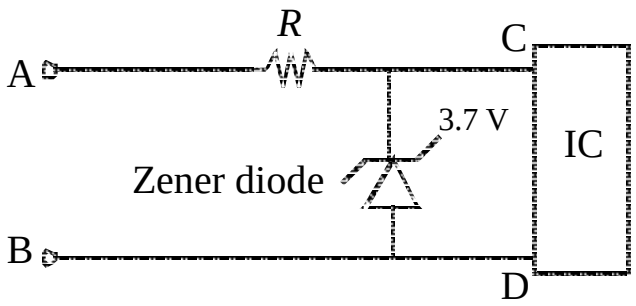
$$N_l = 2/(slope \times q\varepsilon_s A^2)$$
$$= 2/(2\times10^{23}\times1.6\times10^{-19}\times12\times8.85\times10^{-14}\times10^{-8}\,\text{cm}^2)$$
$$= 6\times10^{15}\ \text{cm}^{-3}$$

$$\phi_{bi} = \frac{kT}{q}\ln\frac{N_h N_l}{n_i^{\,2}} \quad\Rightarrow\quad N_h = \frac{n_i^{\,2}}{N_l}e^{\frac{q\phi_{bi}}{kT}} = \frac{10^{20}}{6\times10^{15}}e^{\frac{0.84}{0.026}} = 1.8\times10^{18}\ \text{cm}^{-3}$$

- *Is this an accurate way to determine N_l ? N_h ?*

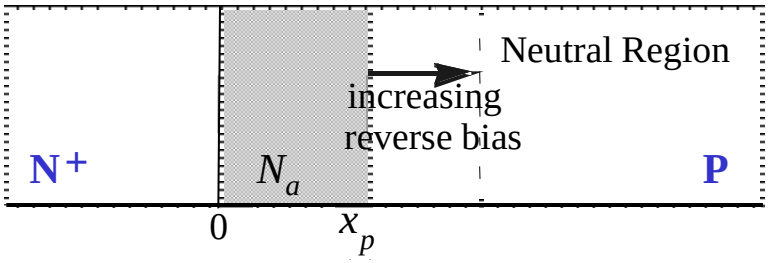
4.5 Junction Breakdown

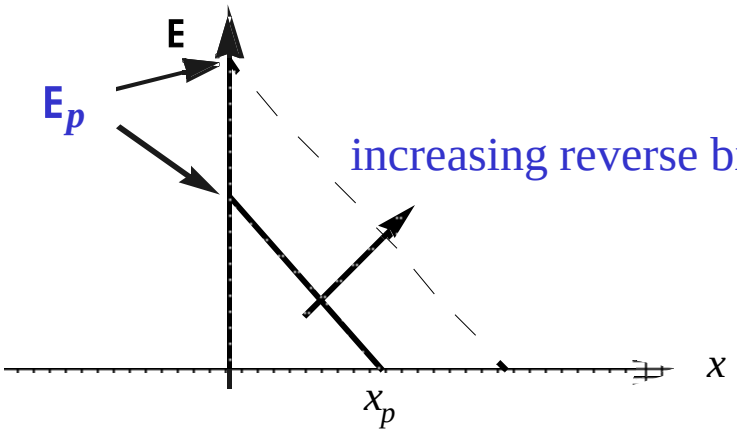




A ***Zener diode*** is designed to operate in the breakdown mode.

4.5.1 Peak Electric Field

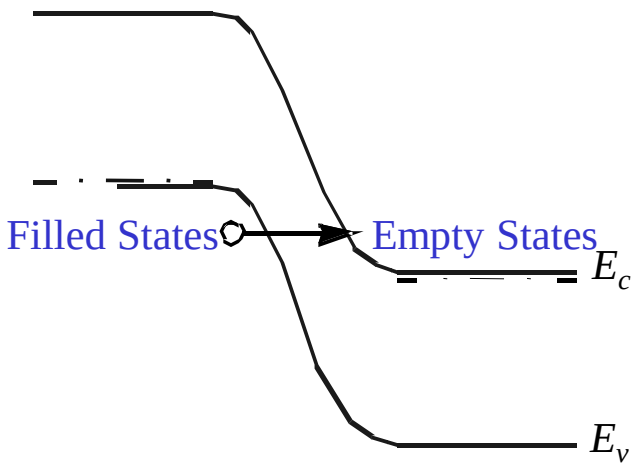


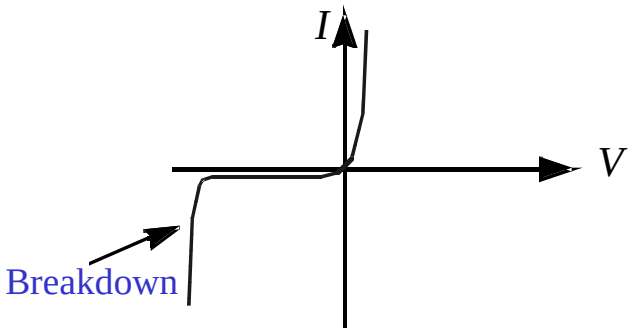


$$\mathbf{E}_p = \mathbf{E}(0) = \left[\frac{2qN}{\varepsilon_s} (\phi_{bi} + |V_r|) \right]^{1/2}$$

$$V_B = \frac{\varepsilon_s \mathbf{E}_{crit}^{\ 2}}{2qN} - \phi_{bi}$$

4.5.2 Tunneling Breakdown



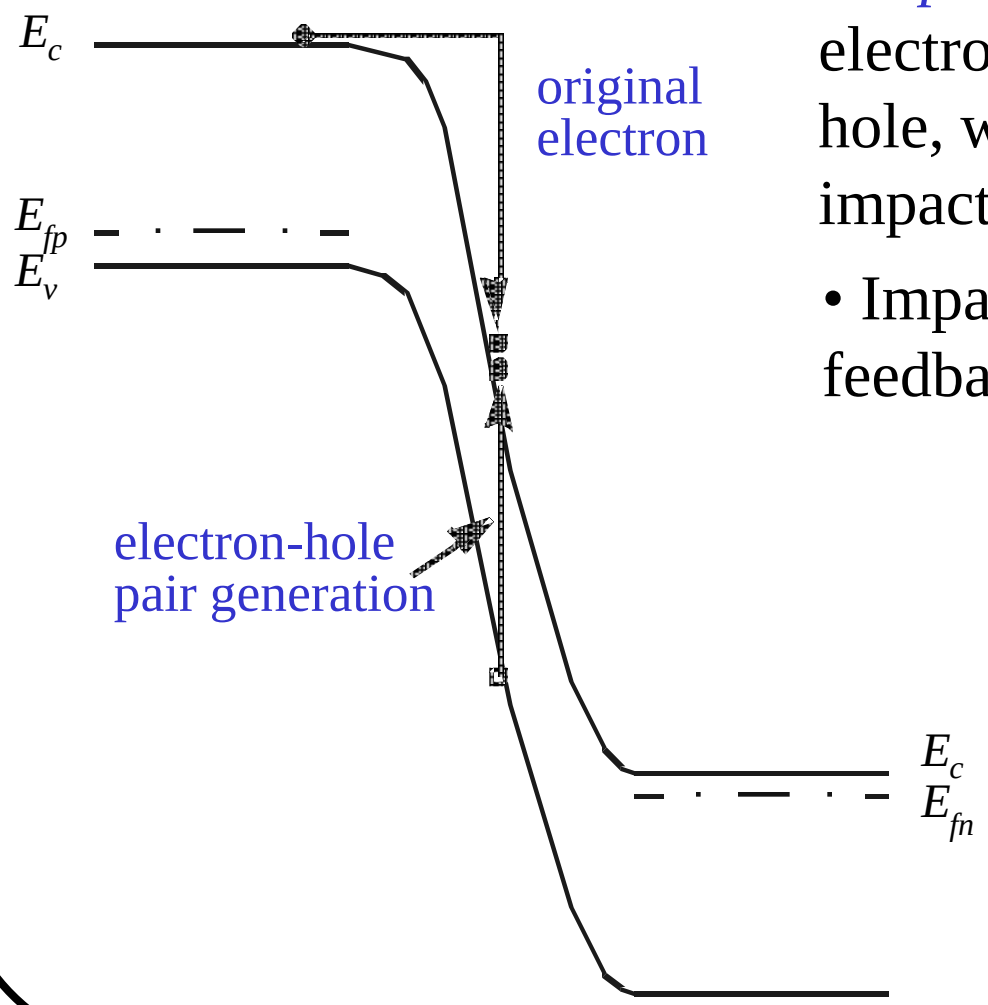


Dominant if both sides of a junction are very heavily doped.

$$J = G\, e^{-H/\mathcal{E}_p}$$

$$\mathcal{E}_p = \mathcal{E}_{crit} \approx 10^6 \text{ V/cm}$$

4.5.3 Avalanche Breakdown



- *impact ionization*: an energetic electron generating electron and hole, which can also cause impact ionization.
- Impact ionization + positive feedback→*avalanche breakdown*

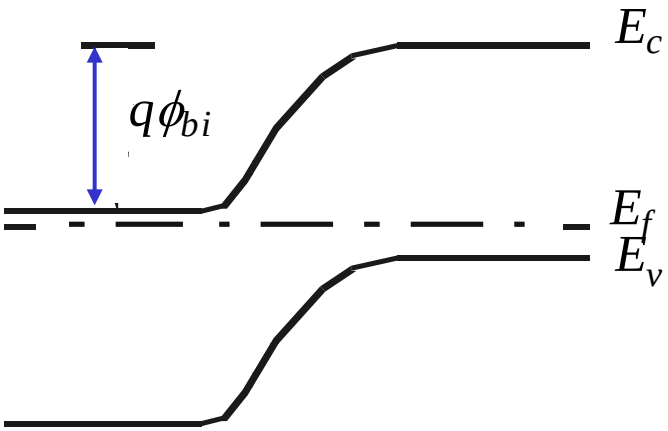
$$V_B = \frac{\varepsilon_s \mathcal{E}_{crit}^2}{2qN}$$

$$\boxed{V_B \propto \frac{1}{N}} = \frac{1}{N_a} + \frac{1}{N_d}$$

4.6 Forward Bias – Carrier Injection

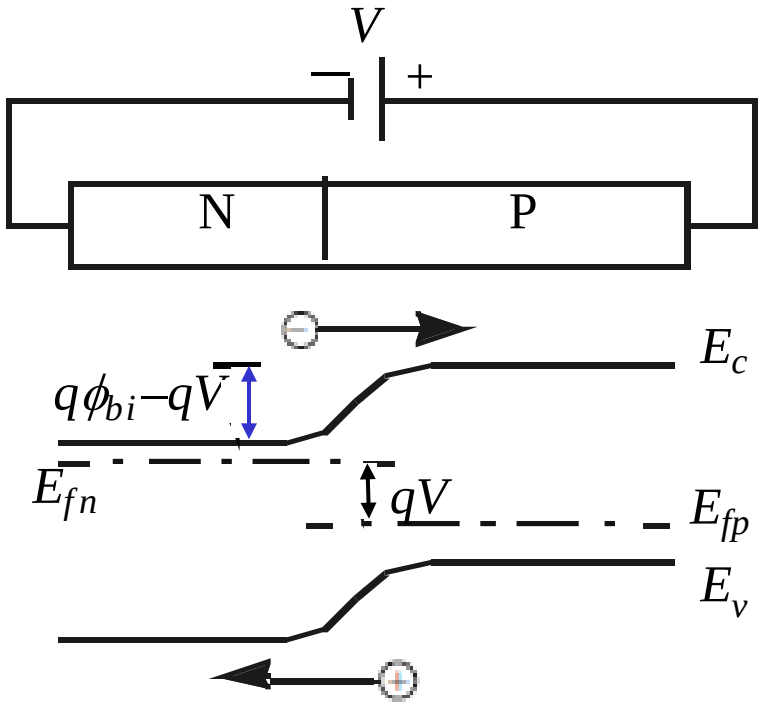
V=0

I=0



Drift and diffusion cancel out

Forward biased

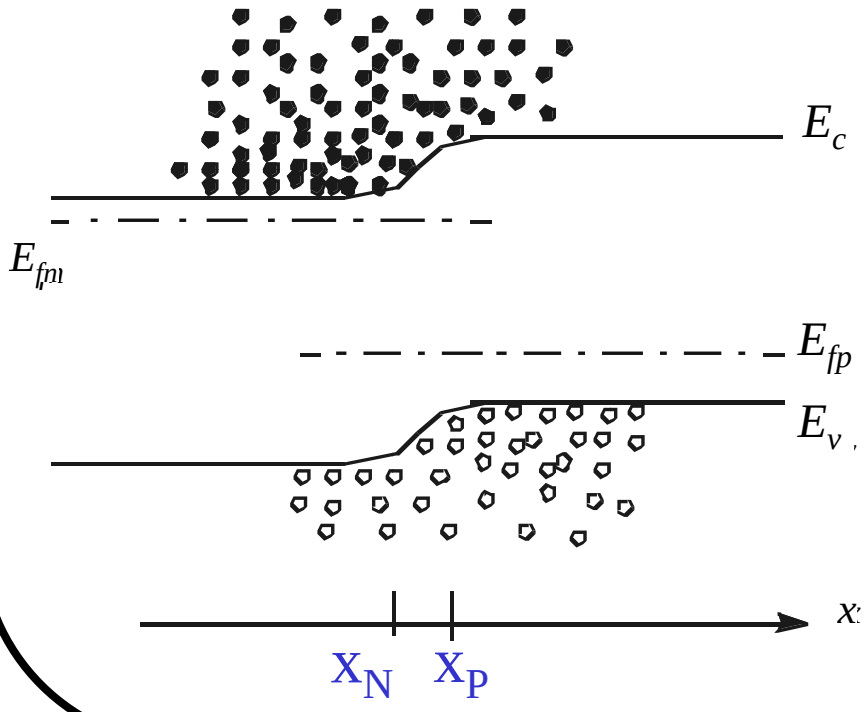


Minority carrier injection

4.6 Forward Bias – Quasi-equilibrium Boundary Condition

$$n(x_P) = N_c e^{-(E_c - E_{fn})/kT} = N_c e^{-(E_c - E_{fp})/kT} e^{(E_{fn} - E_{fp})/kT}$$

$$= n_{P0} e^{(E_{fn} - E_{fp})/kT} = n_{P0} e^{qV/kT}$$



- The minority carrier densities are raised by $e^{qV/kT}$
- Which side gets more carrier injection?

4.6 Carrier Injection Under Forward Bias– Quasi-equilibrium Boundary Condition

$$n(x_P) = n_{P0} e^{qV/kT} = \frac{n_i^2}{N_a} e^{qV/kT}$$

$$p(x_P) = p_{N0} e^{qV/kT} = \frac{n_i^2}{N_d} e^{qV/kT}$$

$$n'(x_P) \equiv n(x_P) - n_{P0} = n_{P0}(e^{qV/kT} - 1)$$

$$p'(x_N) \equiv p(x_N) - p_{N0} = p_{N0}(e^{qV/kT} - 1)$$

EXAMPLE: Carrier Injection

A PN junction has $N_a=10^{19}cm^{-3}$ and $N_d=10^{16}cm^{-3}$. The applied voltage is 0.6 V.

***Question**: What are the minority carrier concentrations at the depletion-region edges?*

***Solution**:* $n(x_P) = n_{P0}e^{qV/kT} = 10 \times e^{0.6/0.026} = 10^{11}\ \text{cm}^{-3}$

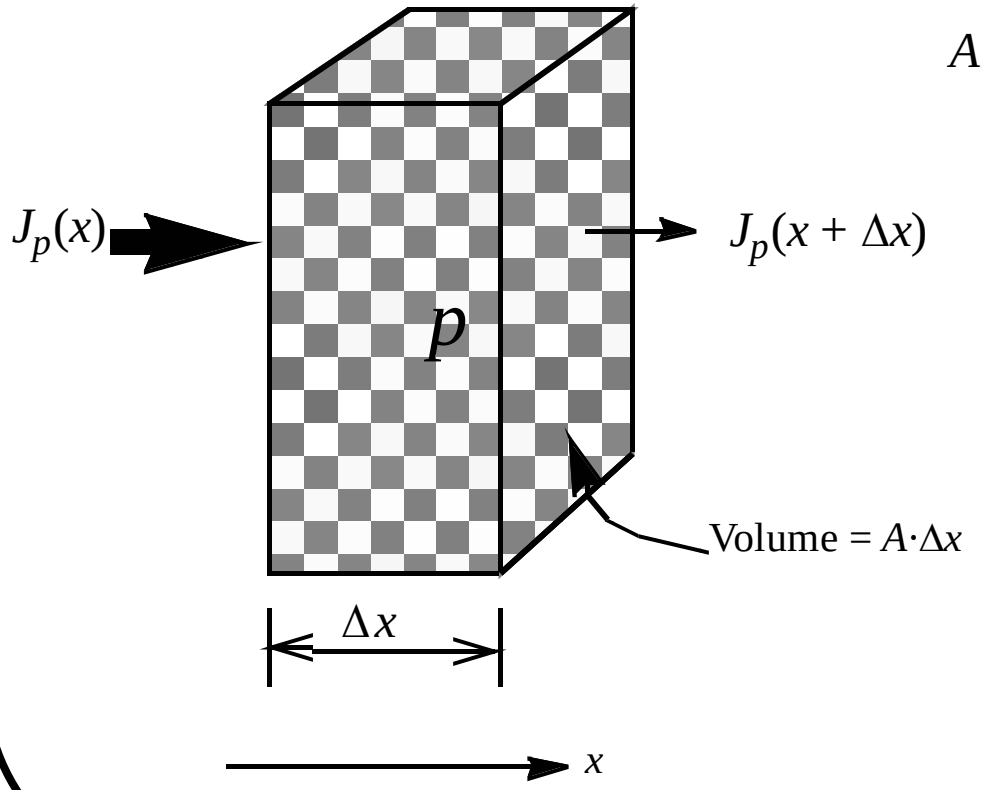
$p(x_N) = p_{N0}e^{qV/kT} = 10^4 \times e^{0.6/0.026} = 10^{14}\ \text{cm}^{-3}$

***Question**: What are the excess minority carrier concentrations?*

***Solution**:* $n'(x_P) = n(x_P) - n_{P0} = 10^{11} - 10 = 10^{11}\ \text{cm}^{-3}$

$p'(x_N) = p(x_N) - p_{N0} = 10^{14} - 10^4 = 10^{14}\ \text{cm}^{-3}$

4.7 Current Continuity Equation



$$A \cdot \frac{J_p(x)}{q} = A \cdot \frac{J_p(x+\Delta x)}{q} + A \cdot \Delta x \cdot \frac{p'}{\tau}$$

$$-\frac{J_p(x+\Delta x) - J_p(x)}{\Delta x} = q\frac{p'}{\tau}$$

$$-\frac{dJ_p}{dx} = q\frac{p'}{\tau}$$

4.7 Current Continuity Equation

$$-\frac{dJ_p}{dx} = q\frac{p'}{\tau}$$

Minority drift current is negligible;

$\therefore \quad J_p = -qD_p dp/dx$

$$qD_p \frac{d^2 p}{dx^2} = q\frac{p'}{\tau_p}$$

$$\frac{d^2 p'}{dx^2} = \frac{p'}{D_p \tau_p} = \frac{p'}{L_p^{\,2}}$$

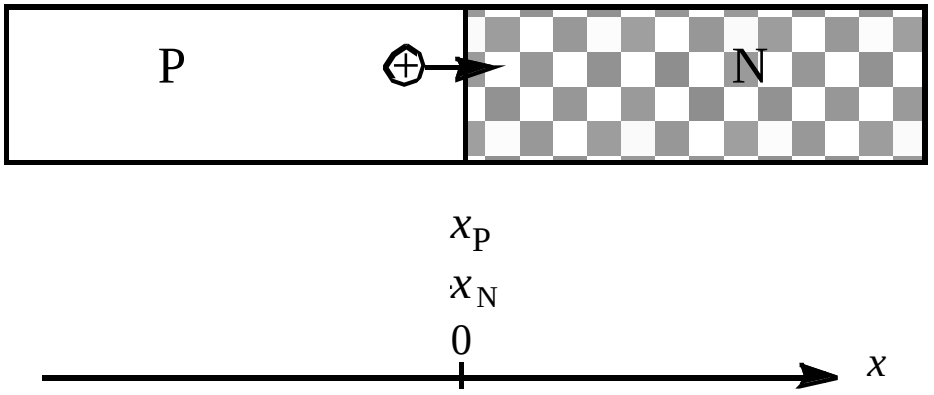
$$\frac{d^2 n'}{dx^2} = \frac{n'}{L_n^{\,2}}$$

L_p and L_n are the diffusion lengths

$$L_p \equiv \sqrt{D_p \tau_p}$$

$$L_n \equiv \sqrt{D_n \tau_n}$$

4.8 Forward Biased Junction-- Excess Carriers



$$\frac{d^2 p'}{dx^2} = \frac{p'}{L_p{}^2}$$

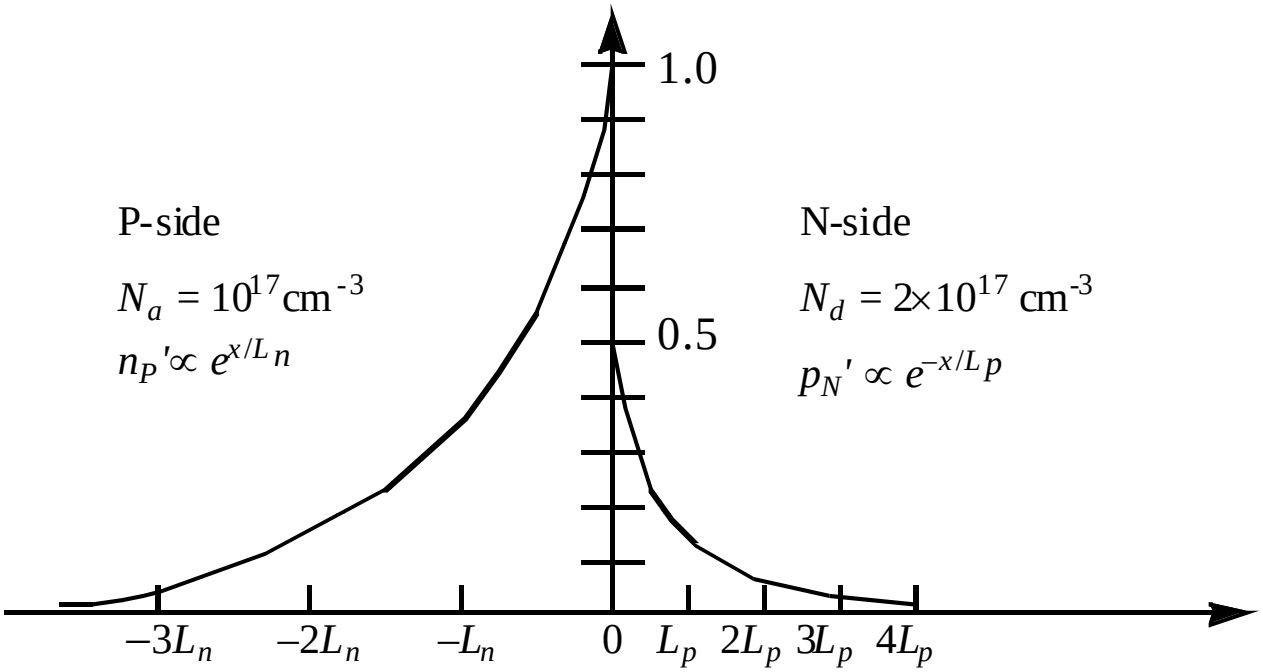
$$p'(\infty) = 0$$

$$p'(x_N) = p_{N0}(e^{qV/kT} - 1)$$

$$p'(x) = Ae^{x/L_p} + Be^{-x/L_p}$$

$$p'(x) = p_{N0}(e^{qV/kT} - 1)e^{-(x-x_N)/L_p}, \quad x > x_N$$

4.8 Excess Carrier Distributions



$$p'(x) = p_{N0}(e^{qV/kT} - 1)e^{-(x-x_N)/L_p}, \quad x > x_N$$

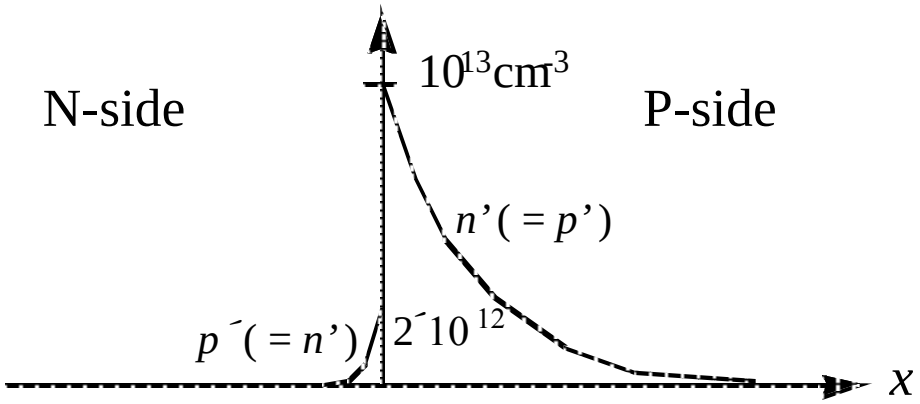
$$n'(x) = n_{P0}(e^{qV/kT} - 1)e^{(x-x_P)/L_n}, \quad x < x_P$$

EXAMPLE: Carrier Distribution in Forward-biased PN Diode

N-type	P-type
$N_d = 5\times10^{17}$ cm^{-3} $D_p = 12$ cm^2/s $\tau_p = 1$ μs	$N_a = 10^{17}$ cm^{-3} $D_n = 36.4$ cm^2/s $\tau_n = 2$ μs

- *Sketch n'(x) on the P-side.*

$$n'(x_P) = n_{P0}(e^{qV/kT} - 1) = \frac{n_i^2}{N_a}(e^{qV/kT} - 1) = \frac{10^{20}}{10^{17}} e^{0.6/0.026} = 10^{13} \text{ cm}^{-3}$$



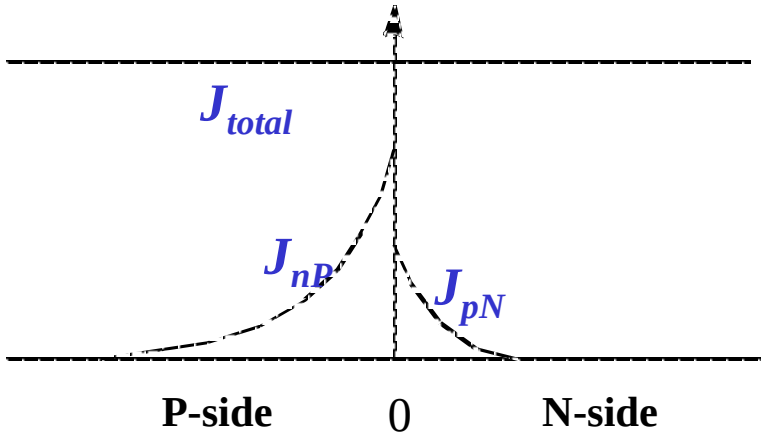
EXAMPLE: Carrier Distribution in Forward-biased PN Diode

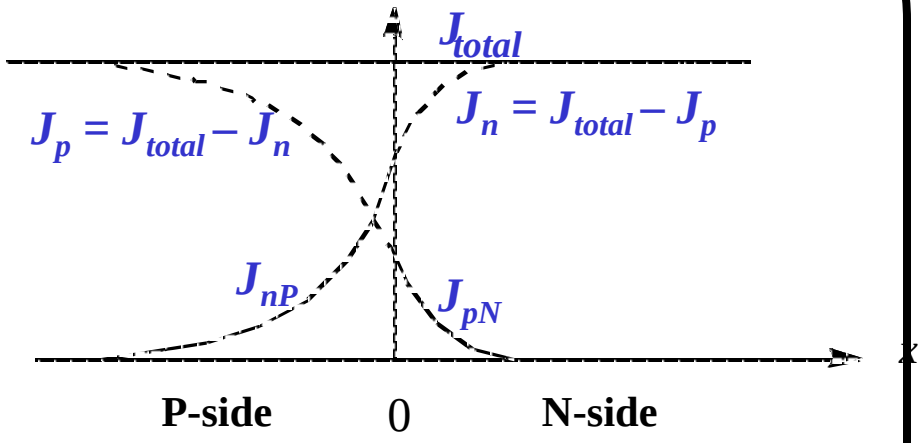
- *How does L_n compare with a typical device size?*

$$L_n = \sqrt{D_n \tau_n} = \sqrt{36 \times 2 \times 10^{-6}} = 85 \ \mu\text{m}$$

- *What is p'(x) on the P- side?*

4.9 PN Diode I-V Characteristics





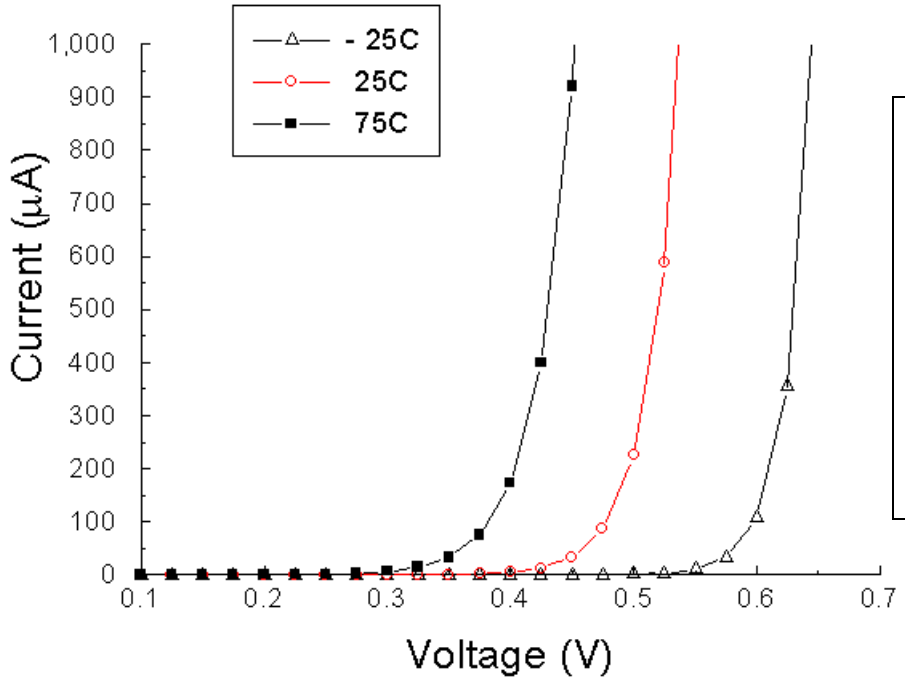
$$J_{pN} = -qD_p \frac{dp'(x)}{dx} = q\frac{D_p}{L_p} p_{N0}(e^{qV/kT} - 1)e^{-(x-x_N)/L_p}$$

$$J_{nP} = qD_n \frac{dn'(x)}{dx} = q\frac{D_n}{L_n} n_{P0}(e^{qV/kT} - 1)e^{(x-x_P)/L_n}$$

$$\text{Total current} = J_{pN}(x_N) + J_{nP}(x_P) = \left(q\frac{D_p}{L_p} p_{N0} + q\frac{D_n}{L_n} n_{P0} \right)(e^{qV/kT} - 1)$$

$$= J \text{ at all } x$$

The PN Junction as a Temperature Sensor

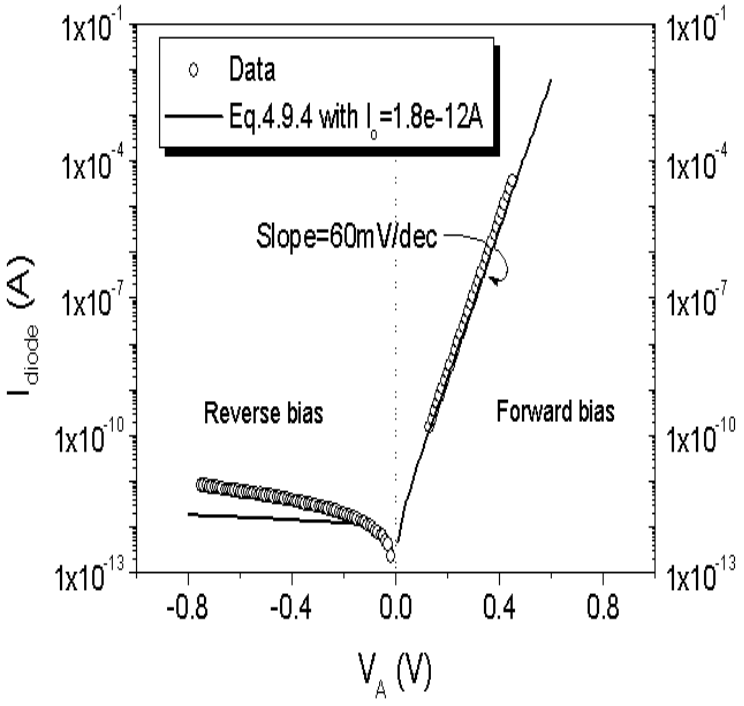


$$I = I_0(e^{qV/kT} - 1)$$

$$I_0 = Aqn_i^2\left(\frac{D_p}{L_pN_d} + \frac{D_n}{L_nN_a}\right)$$

What causes the IV curves to shift to lower V at higher T ?

4.9.1 Contributions from the Depletion Region



$$I_{leakage} = I_0 + A\frac{qn_iW_{dep}}{\tau_{dep}}$$

$$n \approx p \approx n_i e^{qV/2kT}$$

Net recombination (generation) rate :

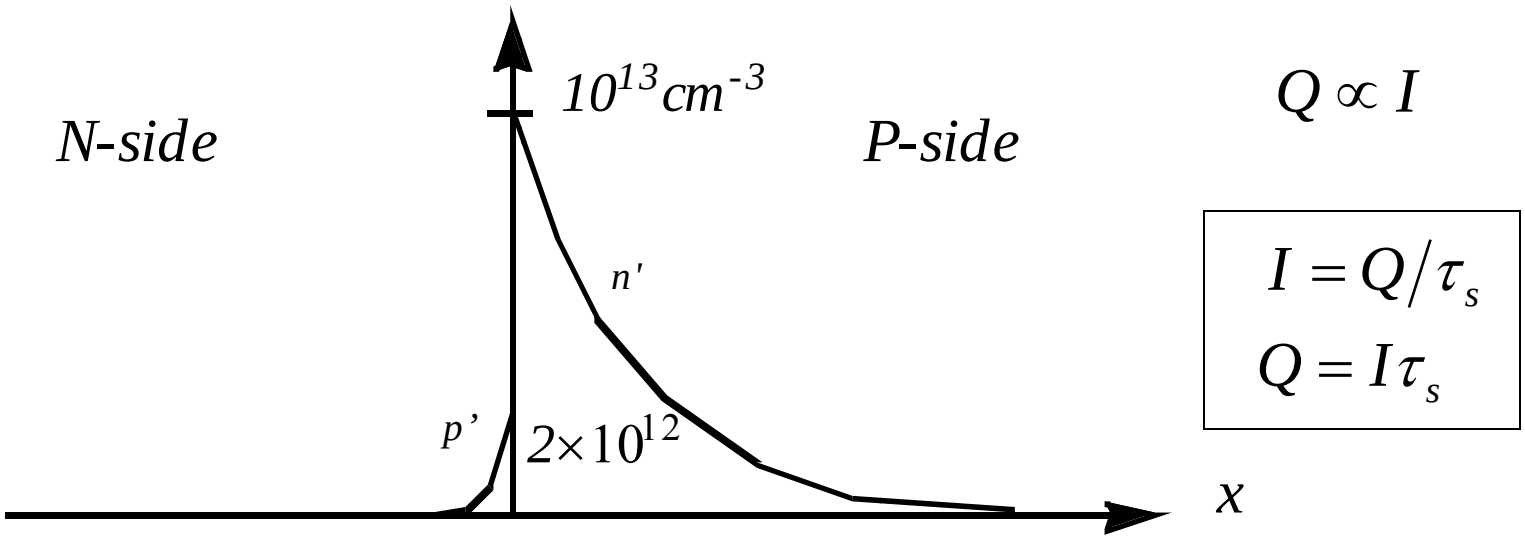
$$\frac{n_i}{\tau_{dep}}(e^{qV/2kT} - 1)$$

$$I = I_0(e^{qV/kT} - 1) + A\frac{qn_iW_{dep}}{\tau_{dep}}(e^{qV/2kT} - 1)$$

Space-Charge Region (SCR) current

Under forward bias, SCR current is an extra current with a slope 120mV/decade

4.10 Charge Storage

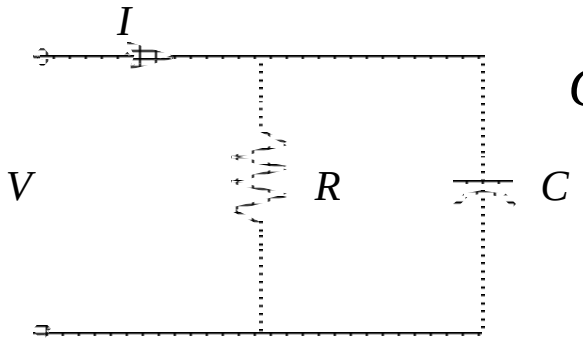


$$Q \propto I$$

$$I = Q/\tau_s$$
$$Q = I\tau_s$$

What is the relationship between τ_s (charge-storage time) and τ (carrier lifetime)?

4.11 Small-signal Model of the Diode



$$G \equiv \frac{1}{R} = \frac{dI}{dV} = \frac{d}{dV} I_0 (e^{qV/kT} - 1) \approx \frac{d}{dV} I_0 e^{qV/kT}$$

$$= \frac{q}{kT} I_0 (e^{qV/kT}) = I_{DC} / \frac{kT}{q}$$

What is G at 300K and I_{DC} = 1 mA?

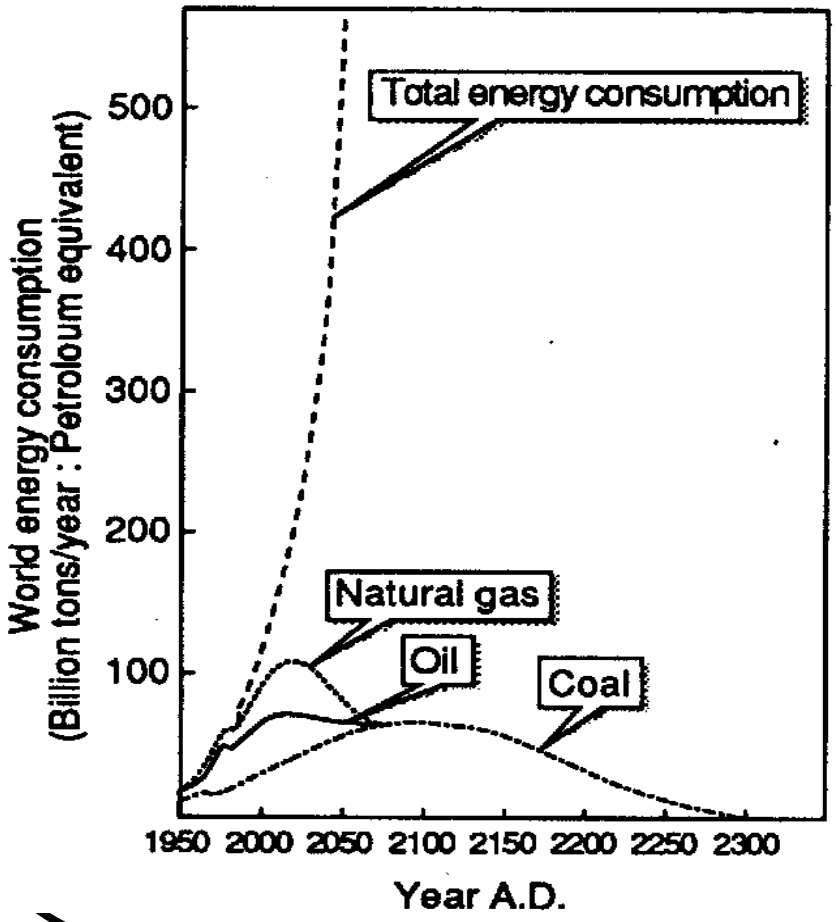
Diffusion Capacitance:

$$C = \frac{dQ}{dV} = \tau_s \frac{dI}{dV} = \tau_s G = \tau_s I_{DC} / \frac{kT}{q}$$

Which is larger, diffusion or depletion capacitance?

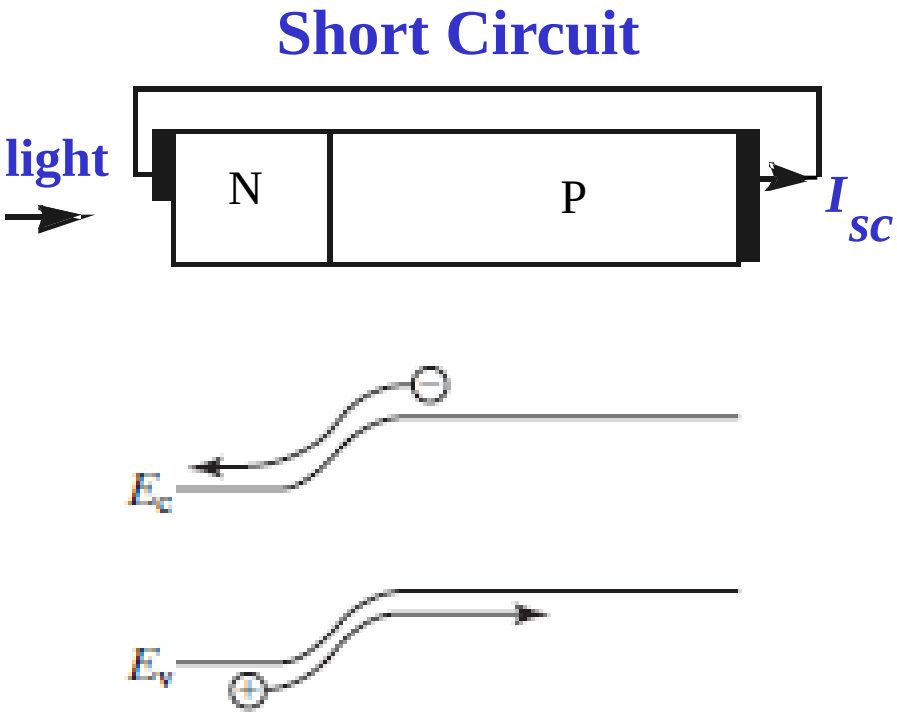
Part II: Application to Optoelectronic Devices

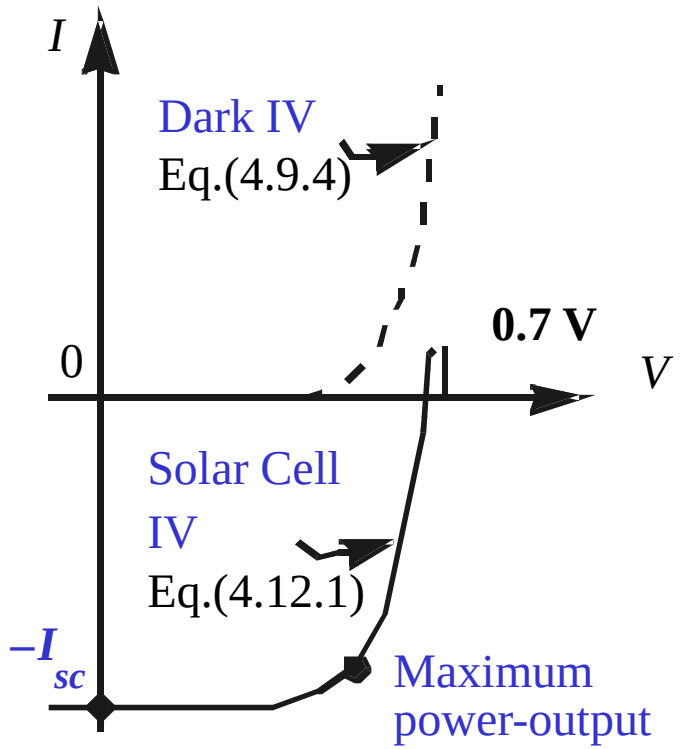
4.12 Solar Cells



- ***Solar Cells*** is also known as ***photovoltaic cells*.**
- Converts sunlight to electricity with 10-30% conversion efficiency.
- 1 m^2 solar cell generate about 150 W peak or 25 W continuous power.
- Low cost and high efficiency are needed for wide deployment.

4.12.1 Solar Cell Basics

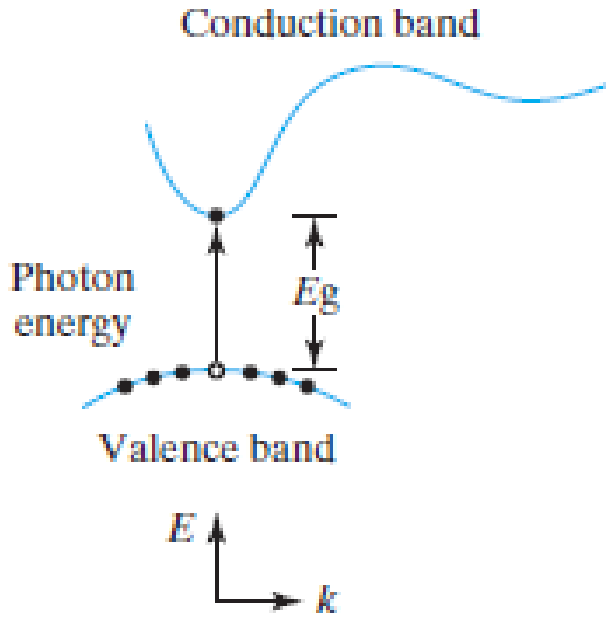




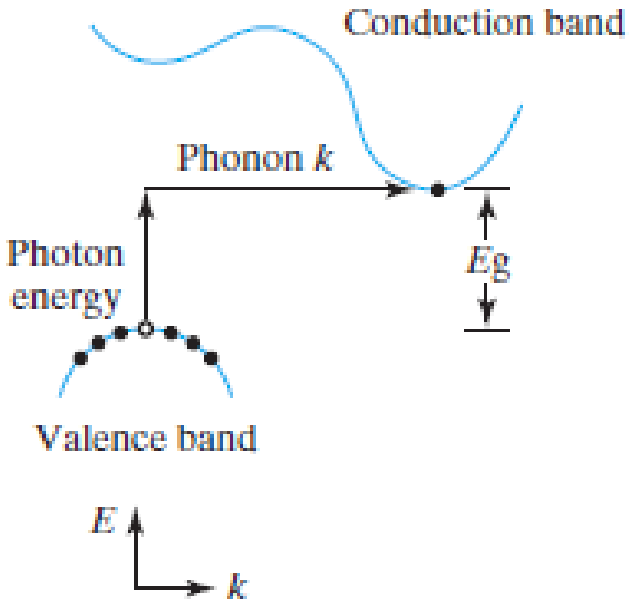
$$I = I_0(e^{qV/kT} - 1) - I_{sc}$$

Direct-Gap and Indirect-Gap Semiconductors

- Electrons have both particle and wave properties.
- An electron has energy E and wave vector k.

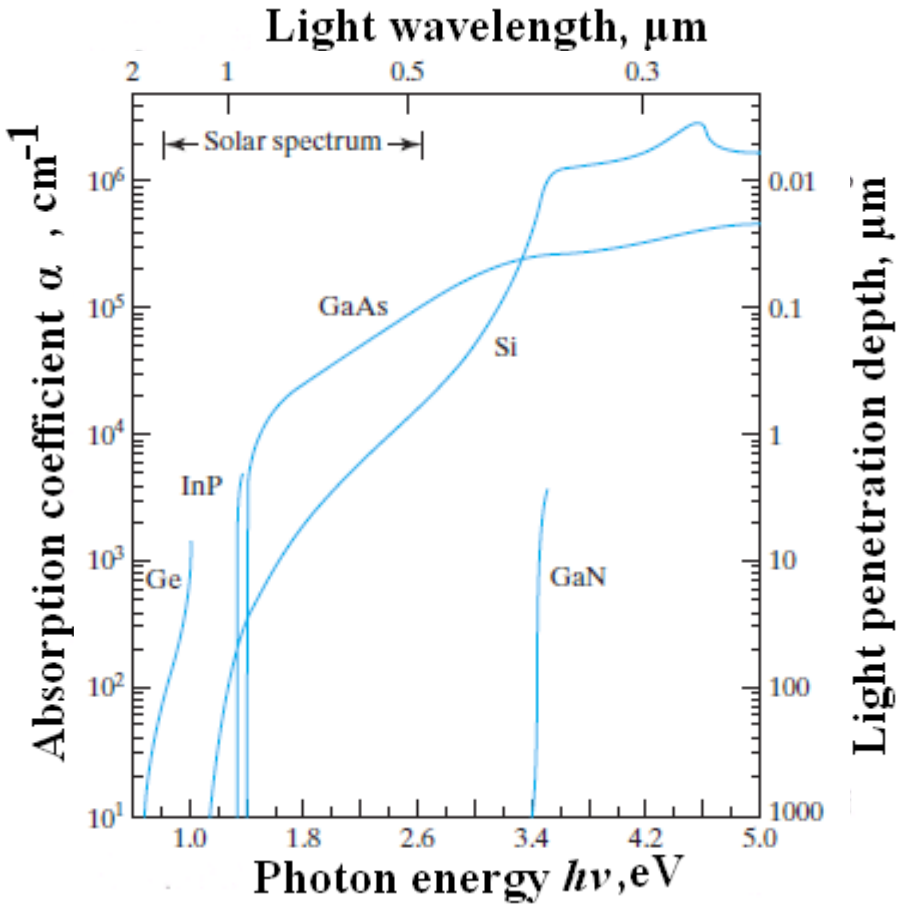


direct-gap semiconductor



indirect-gap semiconductor

4.12.2 Light Absorption



Light intensity (x) $\propto e^{-\alpha x}$

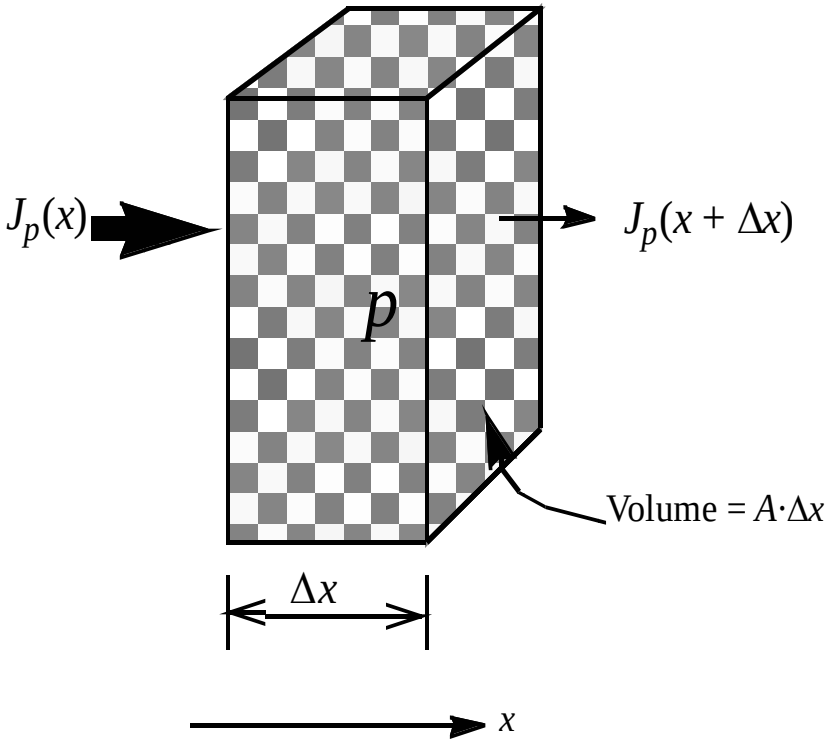
$\alpha(1/cm)$: absorption coefficient

$1/\alpha$: light penetration depth

$$\text{Photon Energy (eV)} = \frac{hc}{\lambda} = \frac{1.24}{\lambda}(\mu m)$$

A thinner layer of direct-gap semiconductor can absorb most of solar radiation than indirect-gap semiconductor. But Si…

4.12.3 Short-Circuit Current and Open-Circuit Voltage



If light shines on the N-type semiconductor and generates holes (and electrons) at the rate of G $s^{-1}cm^{-3}$,

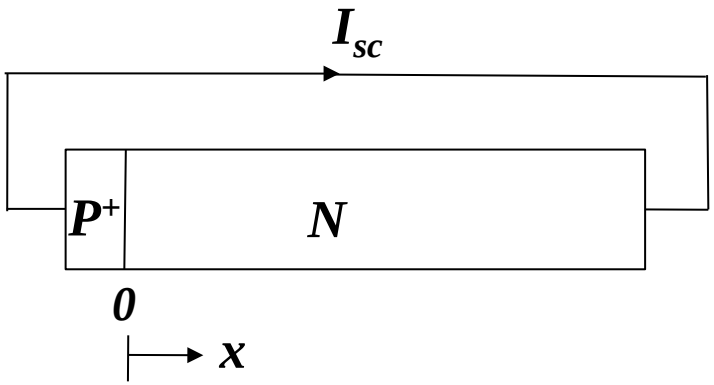
$$\frac{d^2 p'}{dx^2} = \frac{p'}{L_p^2} - \frac{G}{D_p}$$

If the sample is uniform (no PN junction),

$d^2p'/dx^2 = 0 \rightarrow p' = GL_p^2/D_p = G\tau_p$

Solar Cell Short-Circuit Current, I_{sc}

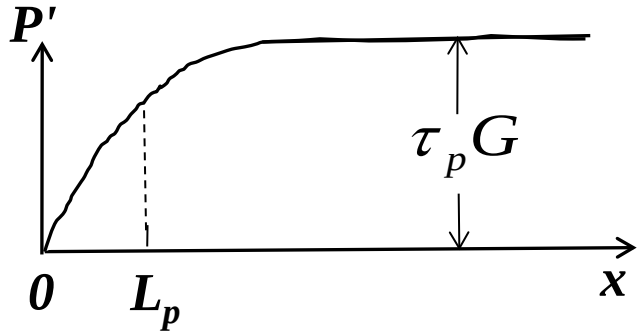
Assume very thin P+ layer and carrier generation in N region only.



$$p'(\infty) = L_p^2 \frac{G}{D_p} = \tau_p G$$

$$p'(0) = 0$$

$$p'(x) = \tau_p G(1 - e^{-x/L_p})$$



$$J_p = -qD_p \frac{dp'(x)}{dx} = q\frac{D_p}{L_p}\tau_p G e^{-x/L_p}$$

$$I_{sc} = AJ_p(0) = AqL_pG$$

G is really not uniform. L_p needs be larger than the light penetration depth to collect most of the generated carriers.

Open-Circuit Voltage

- Total current is I_{SC} plus the PV diode (dark) current:

$$I = Aq\frac{n_i^2}{N_d}\frac{D_p}{L_p}(e^{qV/kT} - 1) - AqL_pG$$

- Solve for the open-circuit voltage (V_{oc}) by setting $I=0$

(assuming $e^{qV_{oc}/kT} >> 1$)

$$0 = \frac{n_i^2}{N_d}\frac{D_p}{L_p}e^{qV_{oc}/kT} - L_pG$$

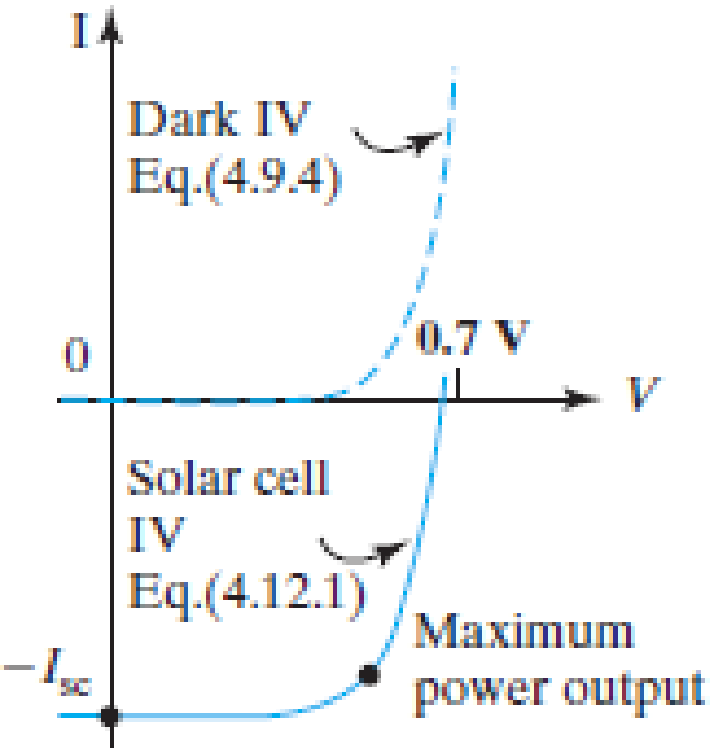
$$V_{oc} = \frac{kT}{q}\ln(\tau_p G N_d / n_i^2)$$

How to raise V_{oc} ?

4.12.4 Output Power

A particular operating point on the solar cell I-V curve maximizes the output power (I ×V).

Output Power $= I_{sc} \times V_{oc} \times FF$



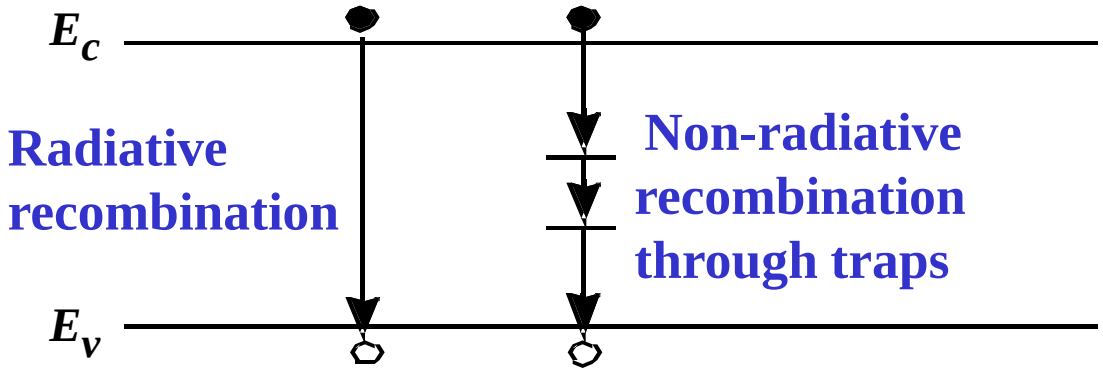
- Si solar cell with 15-20% efficiency dominates the market now

- Theoretically, the highest efficiency (~24%) can be obtained with 1.9eV >Eg>1.2eV. Larger Eg lead to too low Isc (low light absorption); smaller Eg leads to too low Voc.
- ***Tandem solar cells*** gets 35% efficiency using large ***and*** small Eg materials tailored to the short and long wavelength solar light.

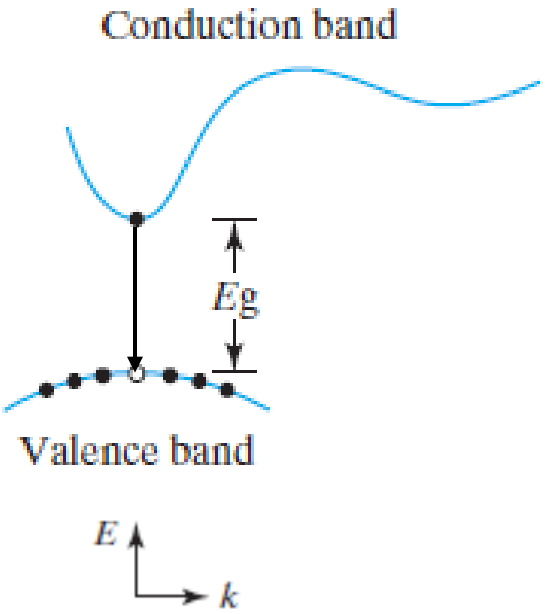
4.13 Light Emitting Diodes and Solid-State Lighting

Light emitting diodes (LEDs)

- LEDs are made of compound semiconductors such as InP and GaN.
- Light is emitted when electron and hole undergo ***radiative recombination*****.**

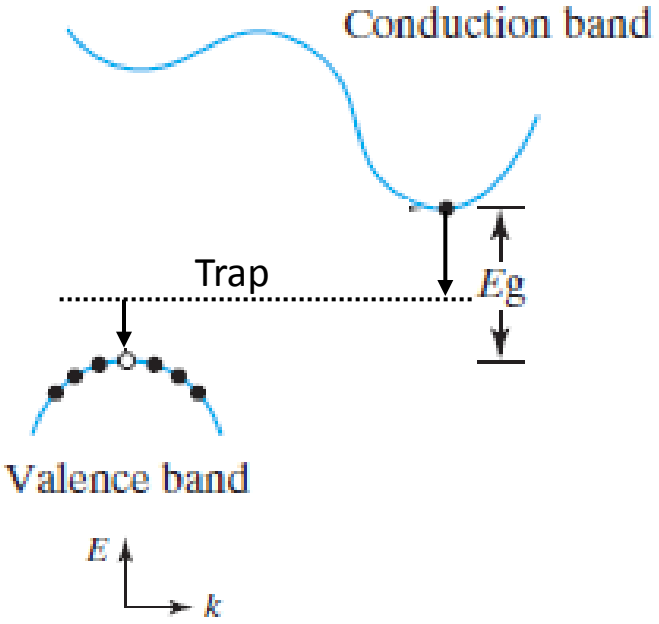


Direct and Indirect Band Gap



Direct band gap
Example: GaAs

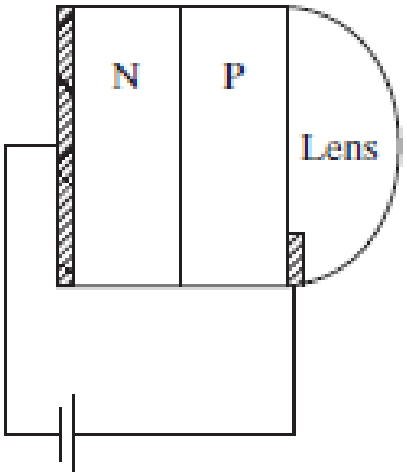
Direct recombination is efficient as k conservation is satisfied.

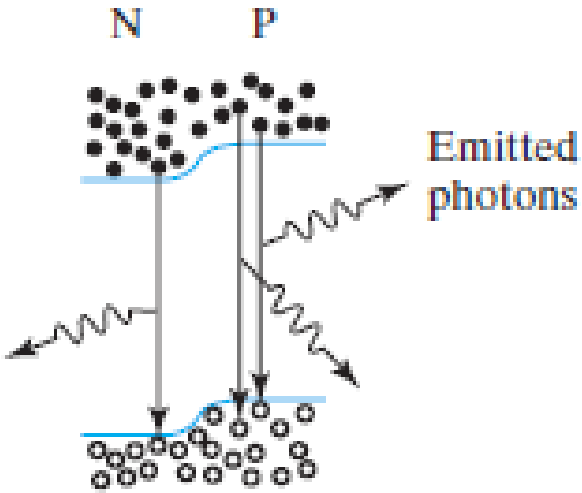


Indirect band gap
Example: Si

Direct recombination is rare as k conservation is not satisfied

4.13.1 LED Materials and Structure





$$\text{LED wavelength}\,(\mu\,\text{m}) = \frac{1.24}{\text{photon energy}} \approx \frac{1.24}{E_g\,(eV)}$$

4.13.1 LED Materials and Structure

	$E_g(eV)$	Wavelength (μm)	Color	Lattice constant (Å)
InAs	0.36	3.44		6.05
InN	0.65	1.91	infrared	3.45
InP	1.36	0.92		5.87
GaAs	1.42	0.87	Red	5.66
GaP	2.26	0.55	Yellow Green Blue	5.46
AlP	3.39	0.51		5.45
GaN	2.45	0.37		3.19
AlN	6.20	0.20	UV	3.11

Light-emitting diode materials

compound semiconductors

binary semiconductors:

- Ex: GaAs, efficient emitter

ternary semiconductor :

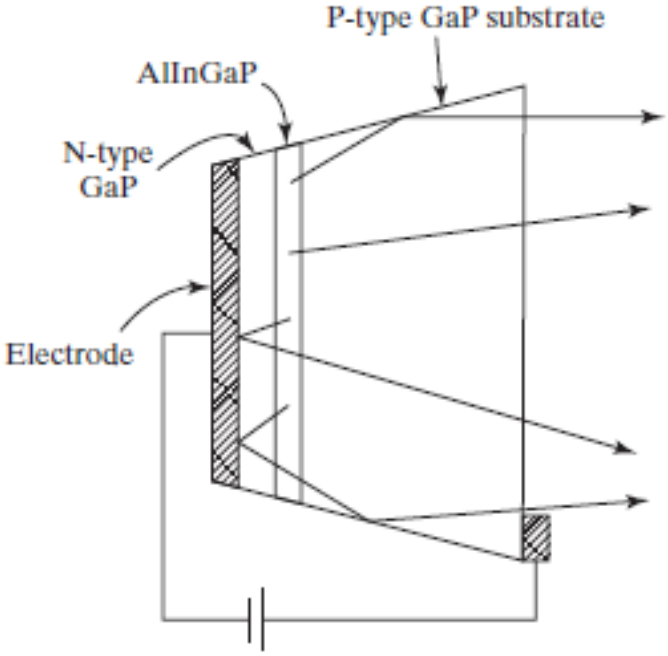
- Ex: $GaAs_{1-x}P_x$, tunable E_g (to vary the color)

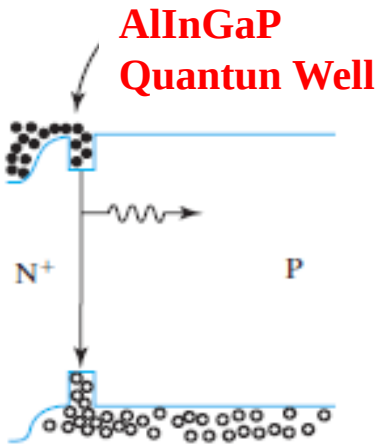
quaternary semiconductors:

- Ex: AlInGaP , tunable E_g and lattice constant (for growing high quality epitaxial films on inexpensive substrates)

Common LEDs

Spectral range	Material System	Substrate	Example Applications
Infrared	InGaAsP	InP	Optical communication
Infrared -Red	GaAsP	GaAs	Indicator lamps. Remote control
Red-Yellow	AlInGaP	GaA or GaP	Optical communication. High-brightness traffic signal lights
Green-Blue	InGaN	GaN or sapphire	High brightness signal lights. Video billboards
Blue-UV	AlInGaN	GaN or sapphire	Solid-state lighting
Red-Blue	Organic semicon-ductors	glass	Displays





4.13.2 Solid-State Lighting

luminosity (lumen, lm): a measure of visible light energy normalized to the sensitivity of the human eye at different wavelengths

Incandescent lamp	Compact fluorescent lamp	Tube fluorescent lamp	**White LED**	Theoretical limit at peak of eye sensitivity (λ=555nm)	Theoretical limit (white light)
17	**60**	**50-100**	**90-?**	**683**	**~340**

Luminous efficacy of lamps in lumen/watt

Organic Light Emitting Diodes (OLED) :
has lower efficacy than nitride or aluminide based compound semiconductor LEDs.

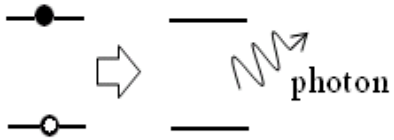
Terms: **luminosity** measured in **lumens**. **luminous efficacy,**

4.14 Diode Lasers

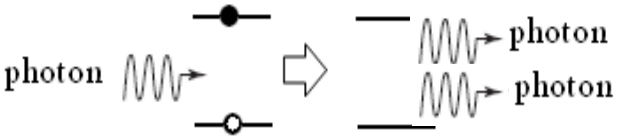
4.14.1 Light Amplification

(a) Absorption

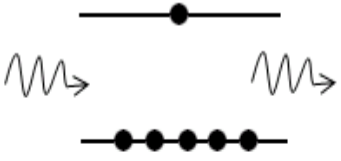
(b) Spontaneous Emission



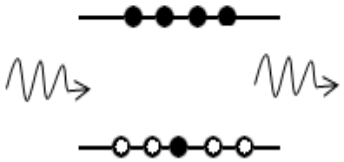
(c) Stimulated Emission



(d) Net Light Absorption



(e) Net Light Amplification



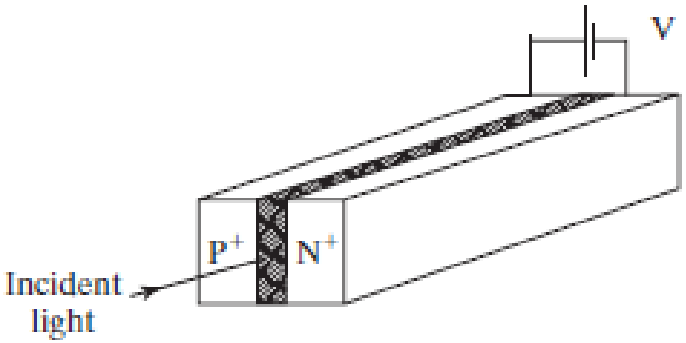
Light amplification **requires *population inversion*: electron occupation probability is larger for higher E states than lower E states.**

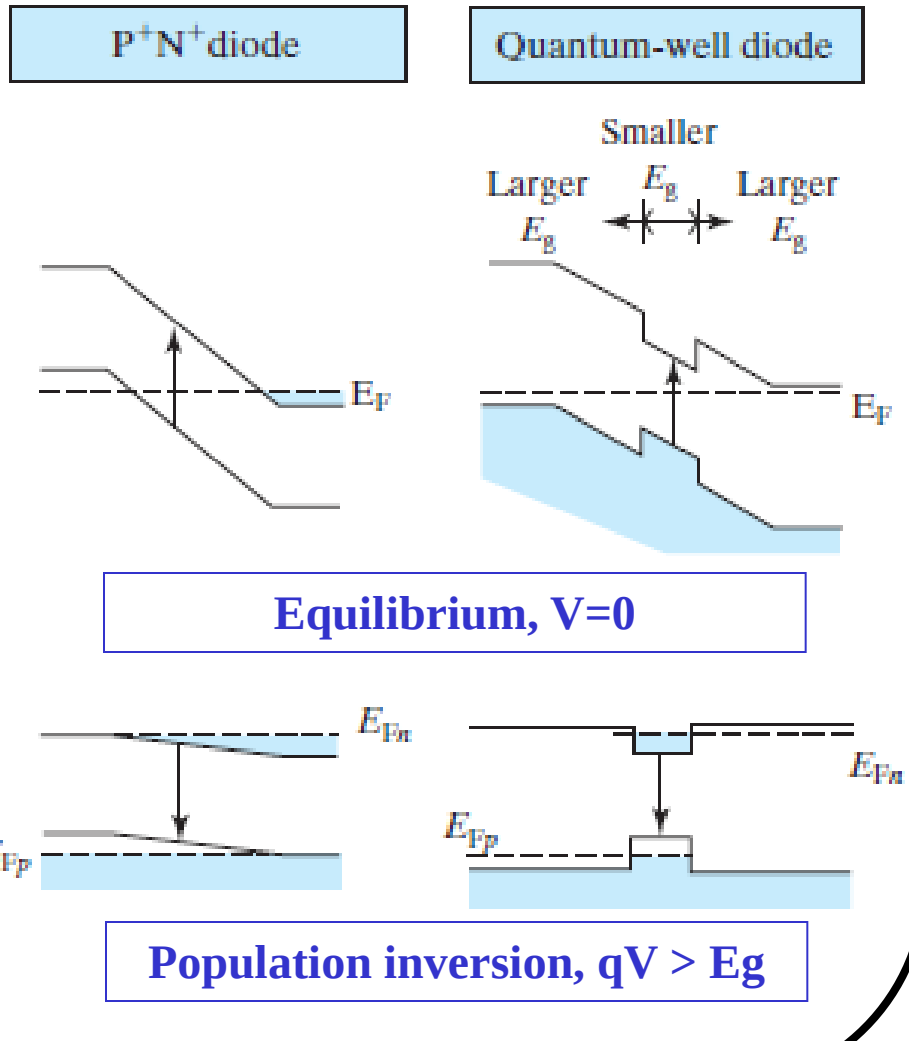
Stimulated emission: emitted photon has identical frequency and directionality as the stimulating photon; light wave is amplified.

4.14.1 Light Amplification in PN Diode

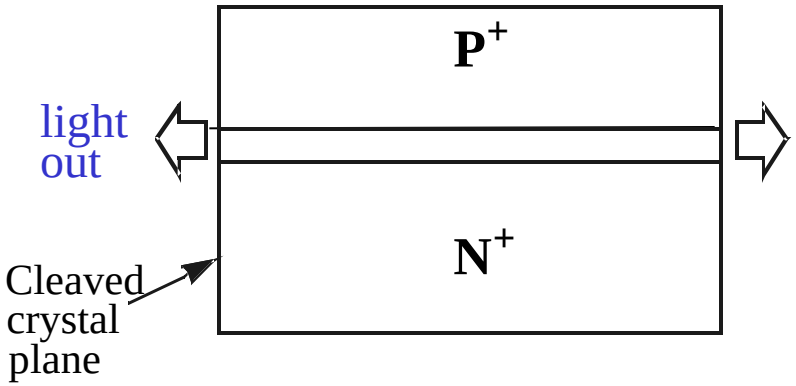
Population inversion is achieved when

$$qV = E_{fn} - E_{fp} > E_g$$





4.14.2 Optical Feedback and Laser



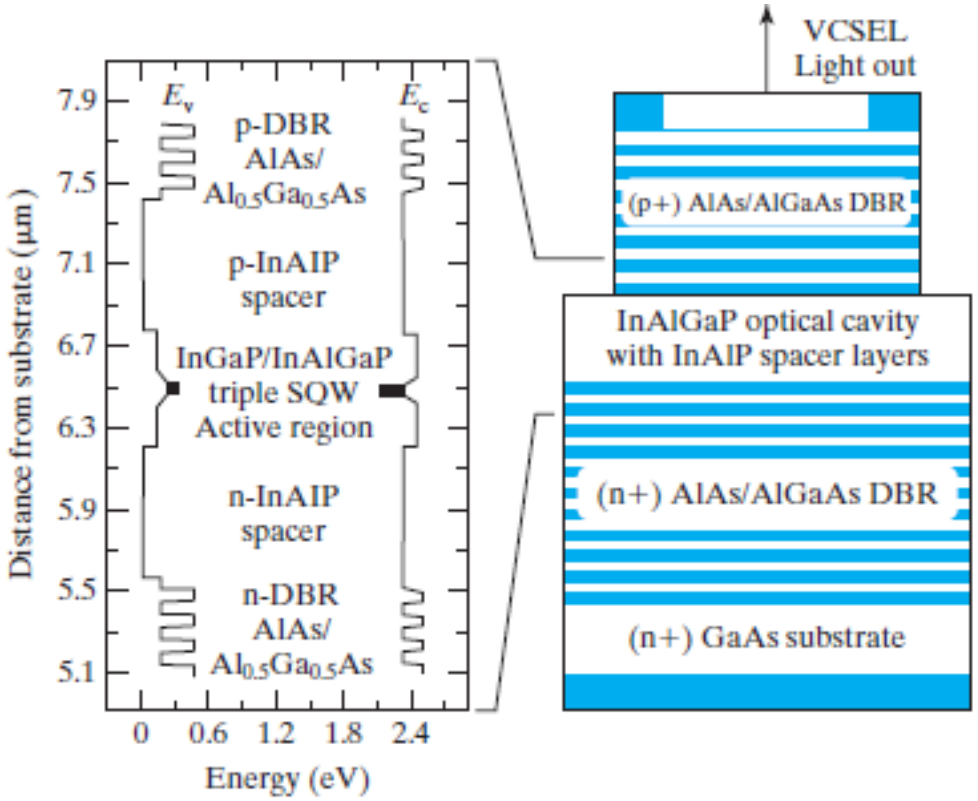
***Laser threshold* is reached (light intensity grows by feedback) when**

$$R_1 \times R_2 \times G \geq 1$$

- **R1, R2**: reflectivities of the two ends
- **G** : light amplification factor (gain) for a round-trip travel of the light through the diode

Light intensity grows until $R_1 \times R_2 \times G = 1$, when the light intensity is just large enough to stimulate carrier recombinations at the same rate the carriers are injected by the diode current.

4.14.2 Optical Feedback and Laser Diode



- ***Distributed Bragg reflector (DBR)*** **reflects light with multi-layers of semiconductors.**
- ***Vertical-cavity surface-emitting laser (VCSEL)*** **is shown on the left.**
- **Quantum-well laser has smaller threshold current because fewer carriers are needed to achieve population inversion in the small volume of the thin small-*Eg* well.**

4.14.3 Laser Applications

***Red* diode lasers**: *CD, DVD reader/writer*

***Blue* diode lasers**: *Blu-ray DVD (higher storage density)*

***1.55 μm infrared* diode lasers:** *Fiber-optic communication*

4.15 Photodiodes

Photodiodes: Reverse biased PN diode. Detects photo-generated current (similar to Isc of solar cell) for optical communication, DVD reader, etc.

Avalanche photodiodes: Photodiodes operating near avalanche breakdown amplifies photocurrent by impact ionization.

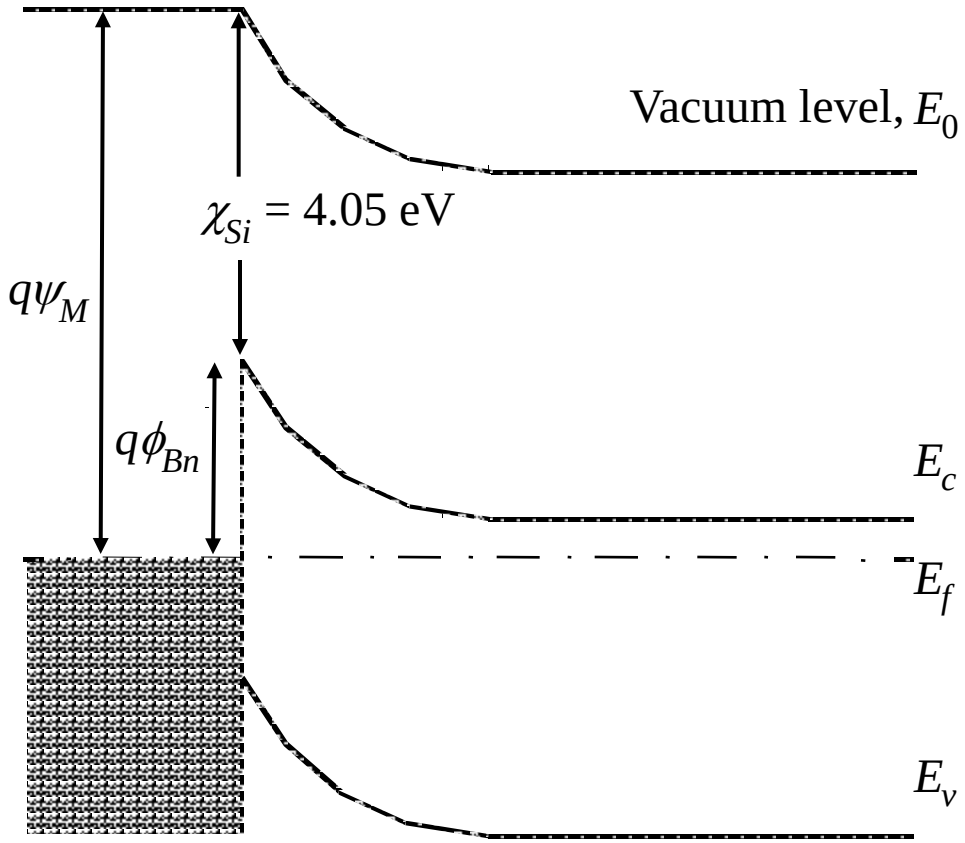
Part III: Metal-Semiconductor Junction

Two kinds of metal-semiconductor contacts:

- *Rectifying **Schottky diodes**: metal on lightly doped silicon*

- *Low-resistance **ohmic contacts**: metal on heavily doped silicon*

ϕ_{Bn} Increases with Increasing Metal Work Function



ψ_M : Work Function of metal

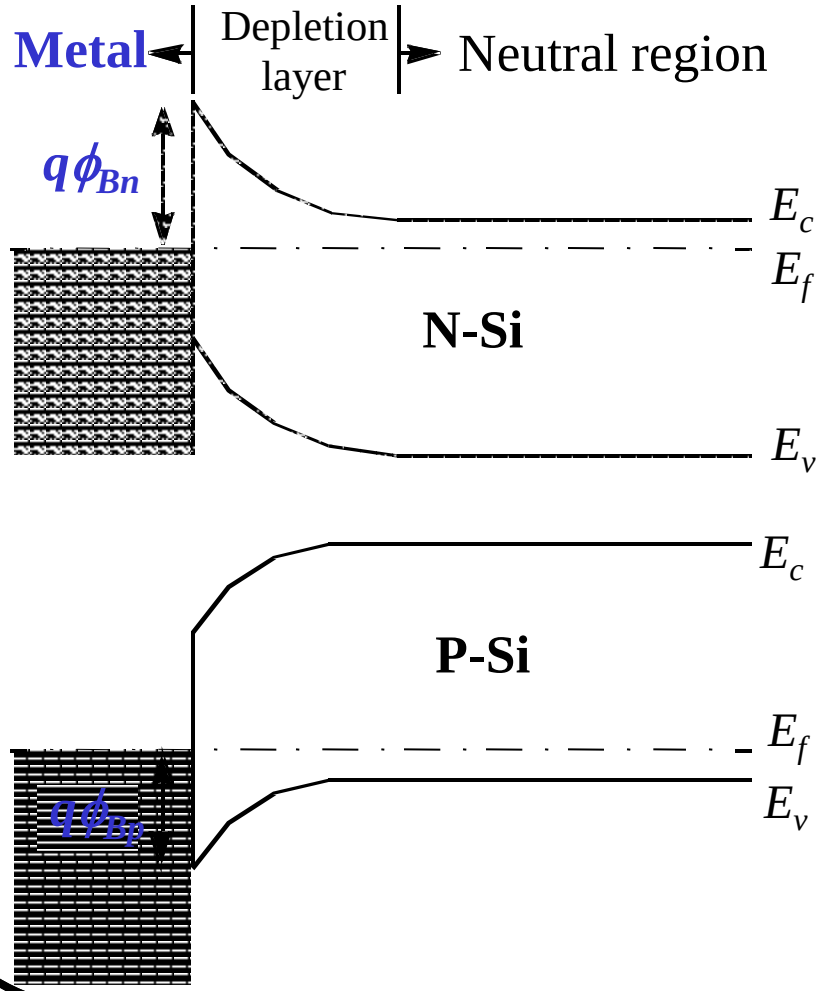
χ_{Si} : Electron Affinity of Si

Theoretically,

$$\phi_{Bn} = \psi_M - \chi_{Si}$$

4.16 Schottky Barriers

Energy Band Diagram of Schottky Contact



- Schottky barrier height, ϕ_B, is a function of the metal material.

- ϕ_B is the most important parameter. The sum of $q\phi_{Bn}$ and $q\phi_{Bp}$ is equal to E_g.

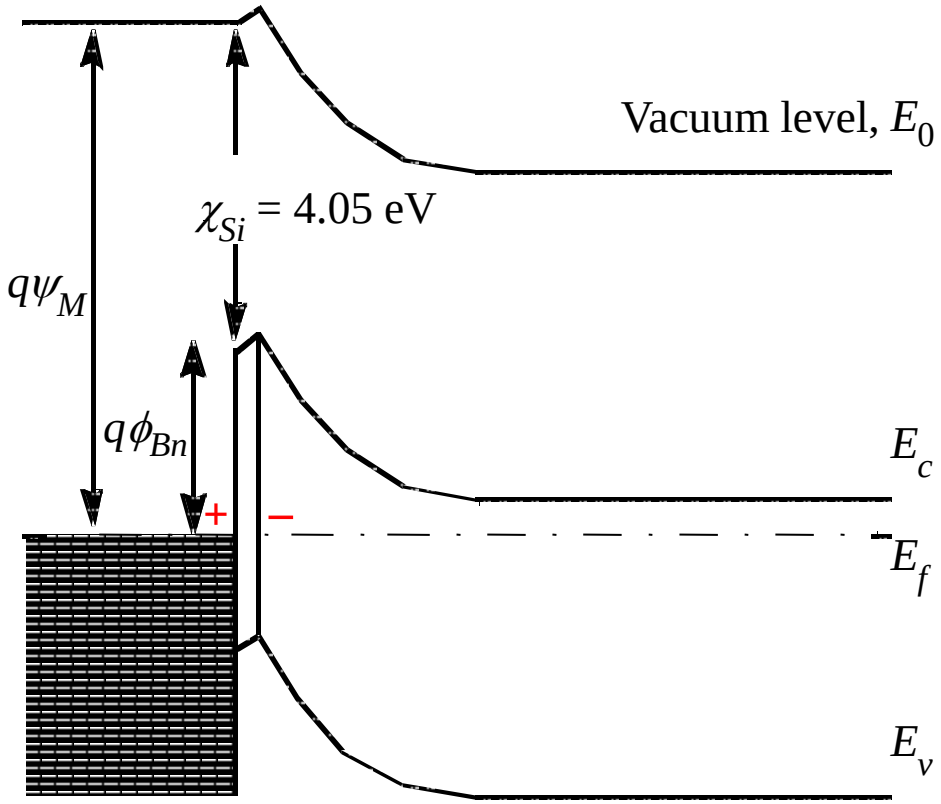
Schottky barrier heights for electrons and holes

Metal	Mg	Ti	Cr	W	Mo	Pd	Au	Pt
ϕ_{Bn} (V)	0.4	0.5	0.61	0.67	0.68	0.77	0.8	0.9
ϕ_{Bp} (V)		0.61	0.5		0.42		0.3	
Work Function ψ_m (V)	3.7	4.3	4.5	4.6	4.6	5.1	5.1	5.7

$$\phi_{Bn} + \phi_{Bp} \approx E_g$$

ϕ_{Bn} increases with increasing metal work function

Fermi Level Pinning



- A high density of energy states in the bandgap at the metal-semiconductor interface pins E_f to a narrow range and ***ϕ_{Bn} is typically 0.4 to 0.9 V***

- ***Question***: What is the typical range of ϕ_{Bp}?

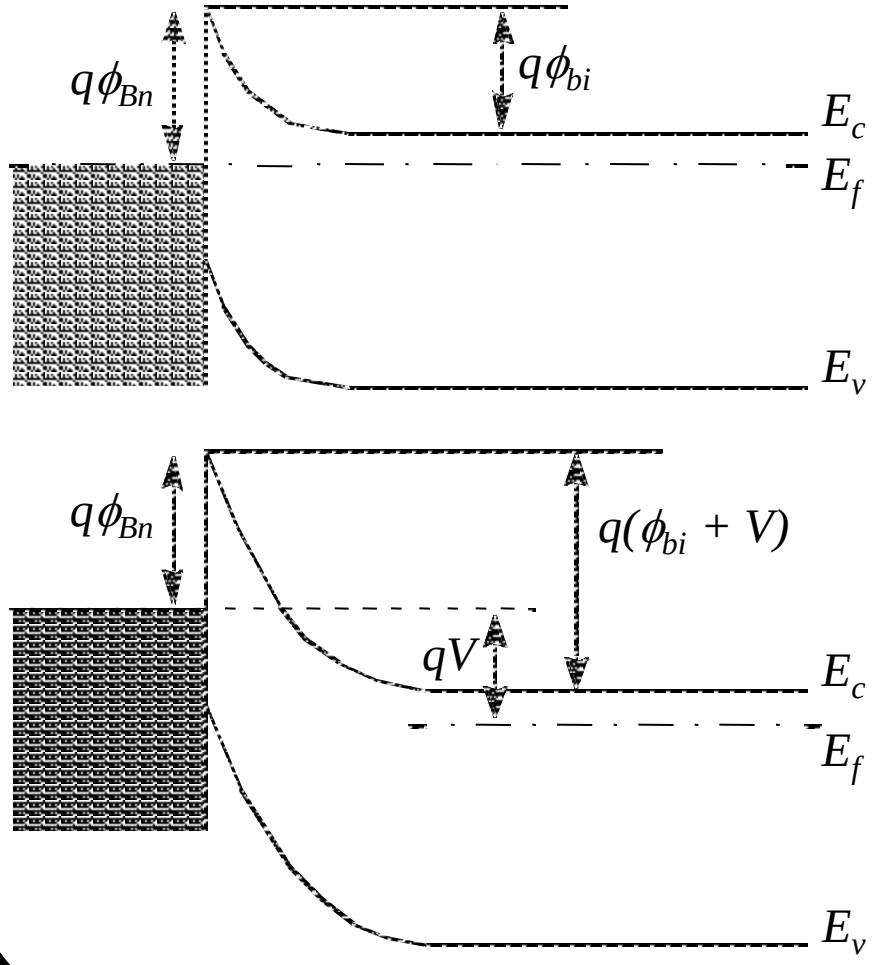
Schottky Contacts of Metal Silicide on Si

***Silicide*: A silicon and metal compound. It is conductive similar to a metal.**

Silicide-Si interfaces are more stable than metal-silicon interfaces. After metal is deposited on Si, an annealing step is applied to form a silicide-Si contact. The term ***metal-silicon contact*** includes and almost always means silicide-Si contacts.

Silicide	$ErSi_{1.7}$	HfSi	$MoSi_2$	$ZrSi_2$	$TiSi_2$	$CoSi_2$	WSi_2	$NiSi_2$	Pd_2Si	PtSi
ϕ_{Bn} (V)	0.28	0.45	0.55	0.55	0.61	0.65	0.67	0.67	0.75	0.87
ϕ_{Bp} (V)			0.55	0.49	0.45	0.45	0.43	0.43	0.35	0.23

Using C-V Data to Determine ϕ_B



$$q\phi_{bi} = q\phi_{Bn} - (E_c - E_f)$$

$$= q\phi_{Bn} - kT \ln \frac{N_c}{N_d}$$

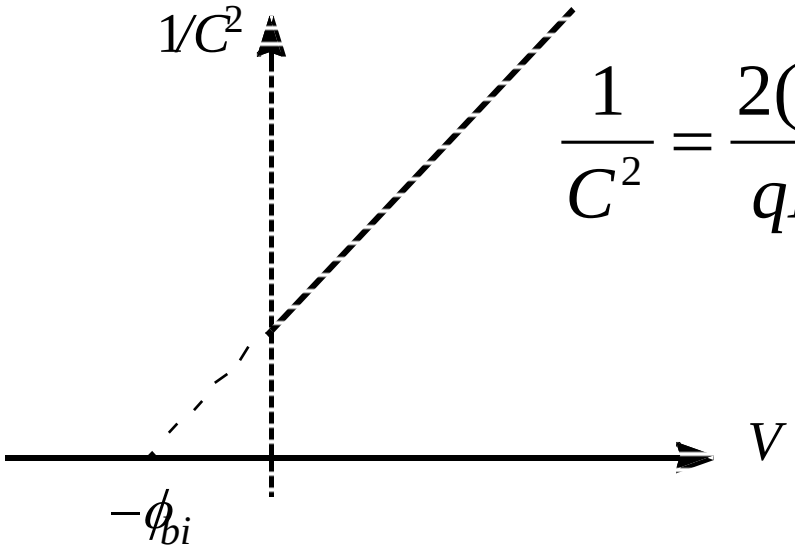
$$W_{dep} = \sqrt{\frac{2\varepsilon_s(\phi_{bi} + V)}{qN_d}}$$

$$C = \frac{\varepsilon_s}{W_{dep}} A$$

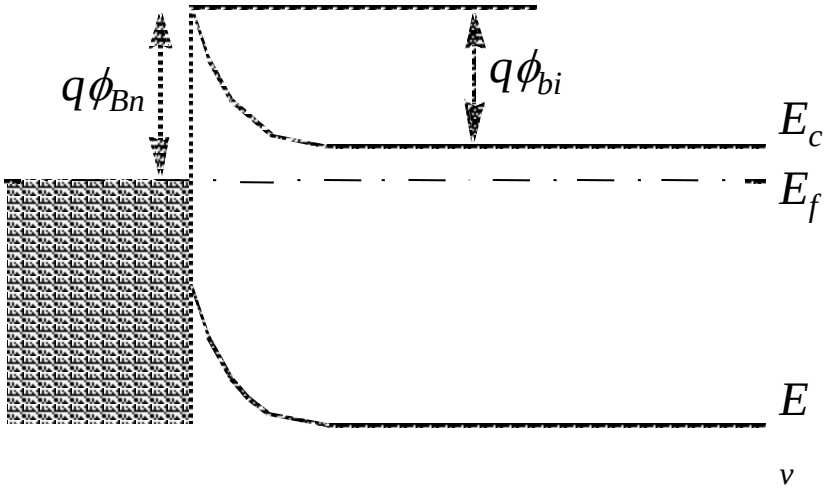
Question:

How should we plot the CV data to extract ϕ_{bi}?

Using CV Data to Determine ϕ_B



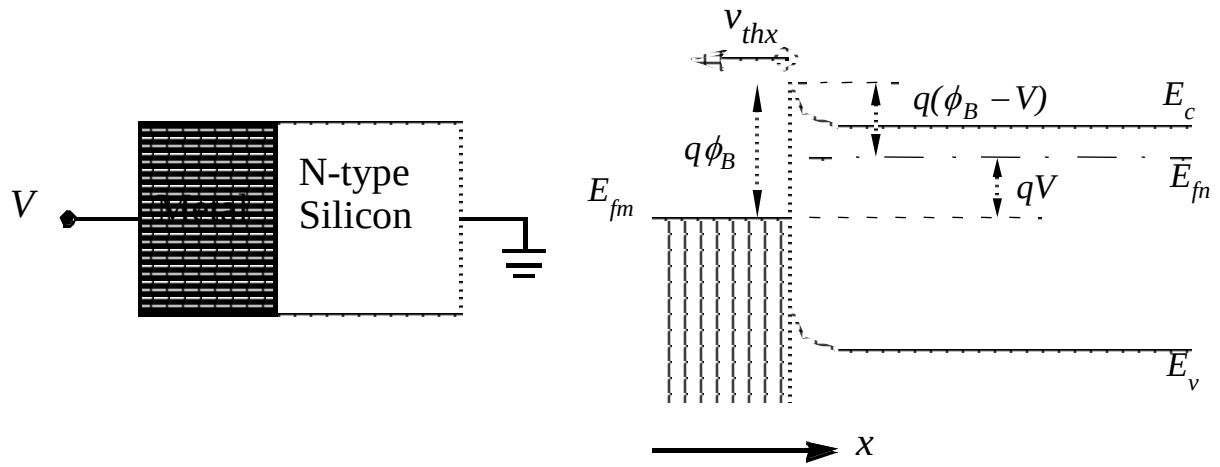
$$\frac{1}{C^2} = \frac{2(\phi_{bi} + V)}{qN_d\varepsilon_s A^2}$$



Once ϕ_{bi} is known, ϕ_B can be determined using

$$q\phi_{bi} = q\phi_{Bn} - (E_c - E_f) = q\phi_{Bn} - kT\ln\frac{N_c}{N_d}$$

4.17 Thermionic Emission Theory



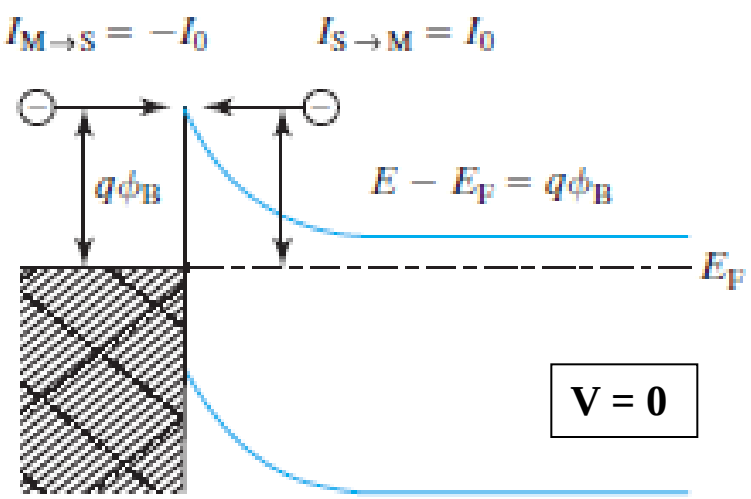
$$n = N_c e^{-q(\phi_B - V)/kT} = 2\left[\frac{2\pi m_n kT}{h^2}\right]^{3/2} e^{-q(\phi_B - V)/kT}$$

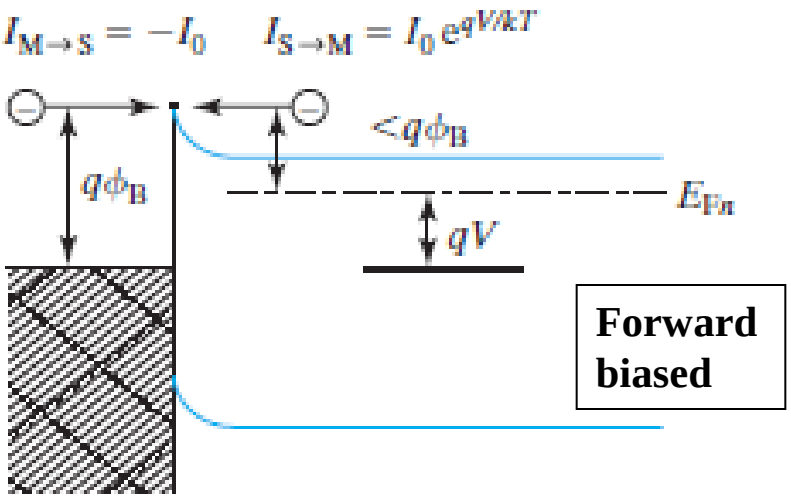
$$v_{th} = \sqrt{3kT/m_n} \qquad v_{thx} = -\sqrt{2kT/\pi m_n}$$

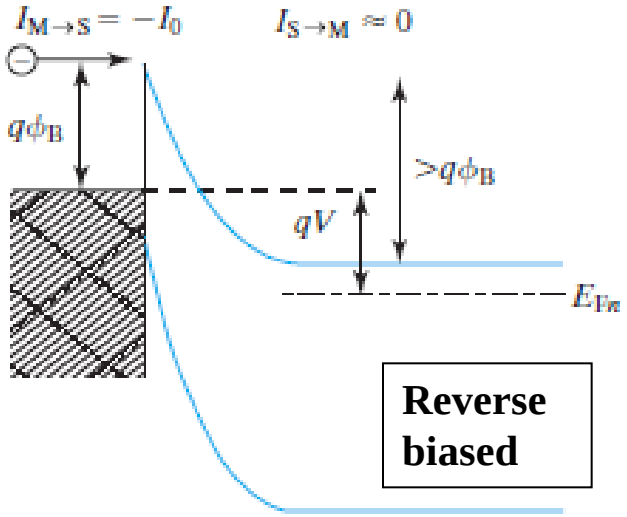
$$J_{S\to M} = -\frac{1}{2} qnv_{thx} = \frac{4\pi q m_n k^2}{h^3} T^2 e^{-q\phi_B/kT} e^{qV/kT}$$

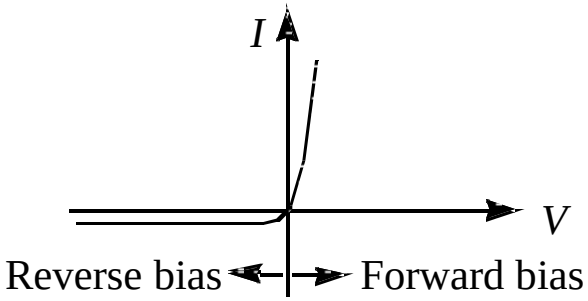
$$= J_0 e^{qV/kT}, \text{where } J_o \approx 100 e^{-q\phi_B/kT} \text{ A/cm}^2$$

4.18 Schottky Diodes

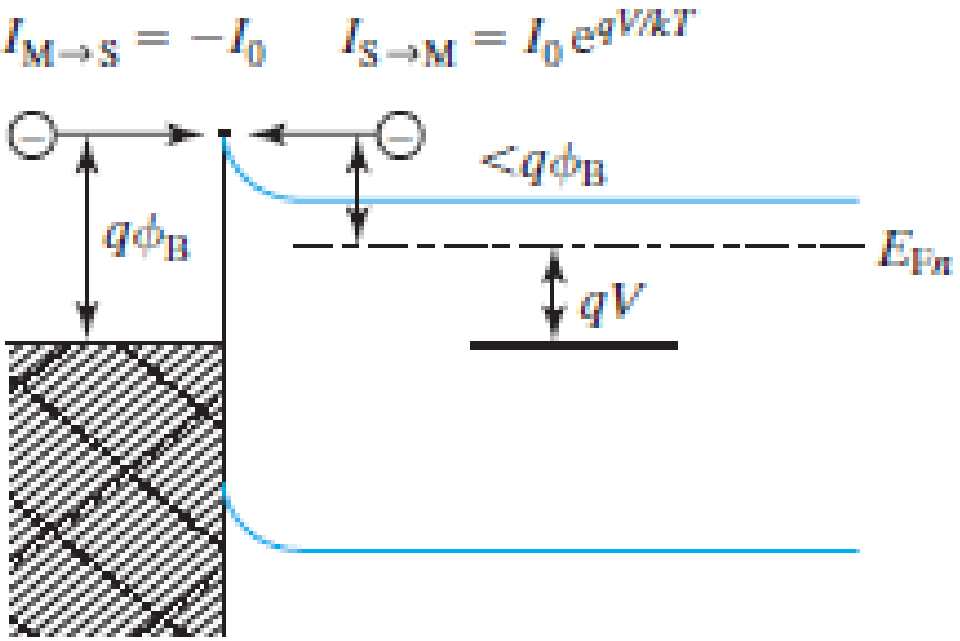








4.18 Schottky Diodes

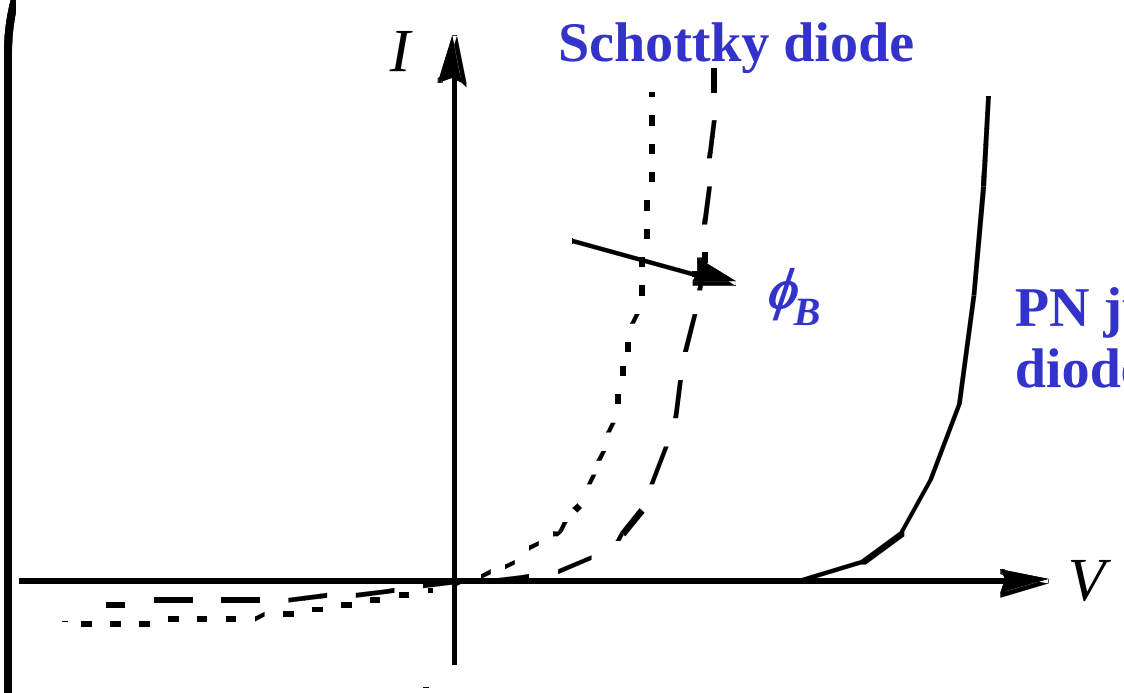


$$I_0 = AKT^2 e^{-q\phi_B/kT}$$

$$K = \frac{4\pi q m_n k^2}{h^3} \approx 100 \text{ A/(cm}^2 \cdot \text{K}^2\text{)}$$

$$I = I_{S\to M} + I_{M\to S} = I_0 e^{qV/kT} - I_0 = I_0 (e^{qV/kT} - 1)$$

4.19 Applications of Schottly Diodes

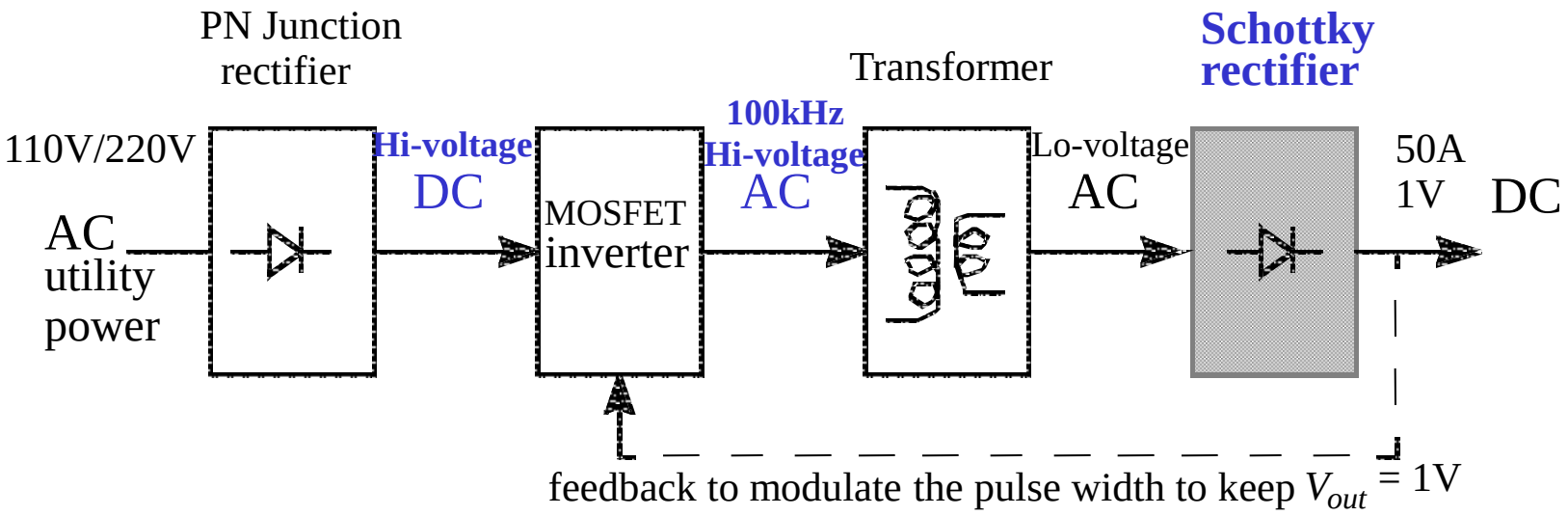


$$I = I_0(e^{qV/kT} - 1)$$

$$I_0 = AKT^2 e^{-q\phi_B/kT}$$

- I_0 of a Schottky diode is 10^3 to 10^8 times larger than a PN junction diode, depending on ϕ_B. A larger I_0 means a smaller forward drop V.
- A Schottky diode is the preferred rectifier in low voltage, high current applications.

Switching Power Supply



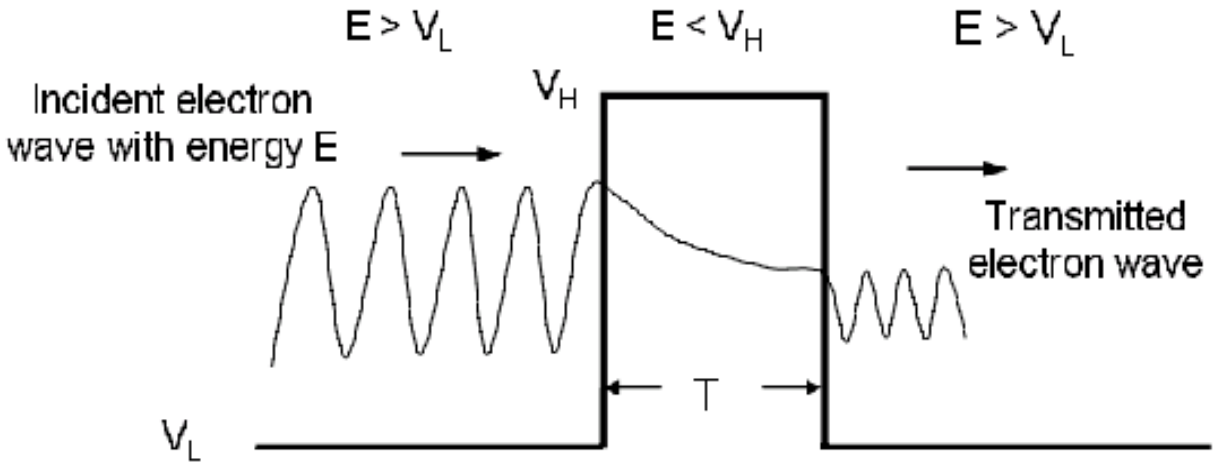
4.19 Applications of Schottky diodes

***Question**:* What sets the lower limit in a Schottky diode's forward drop?

• ***Synchronous Rectifier***: For an even lower forward drop, replace the diode with a wide-W MOSFET which is not bound by the tradeoff between diode *V* and leakage current.

• There is no minority carrier injection at the Schottky junction. Therefore, Schottky diodes can operate at higher frequencies than PN junction diodes.

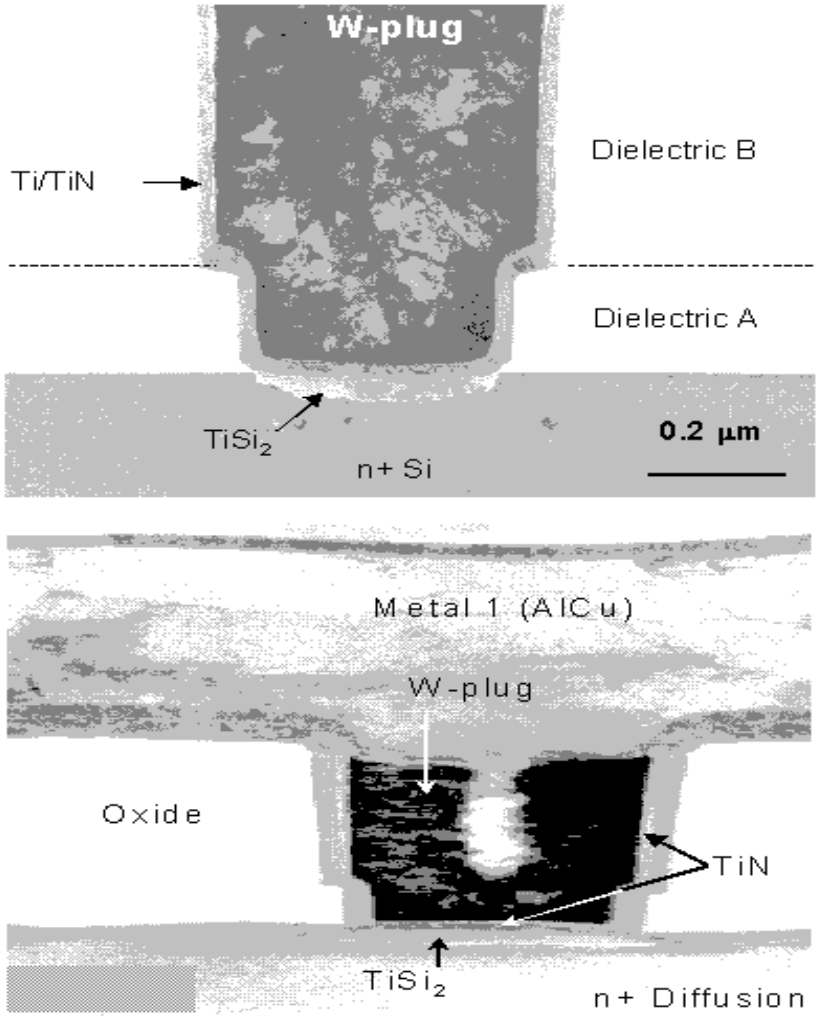
4.20 Quantum Mechanical Tunneling



***Tunneling probability*:**

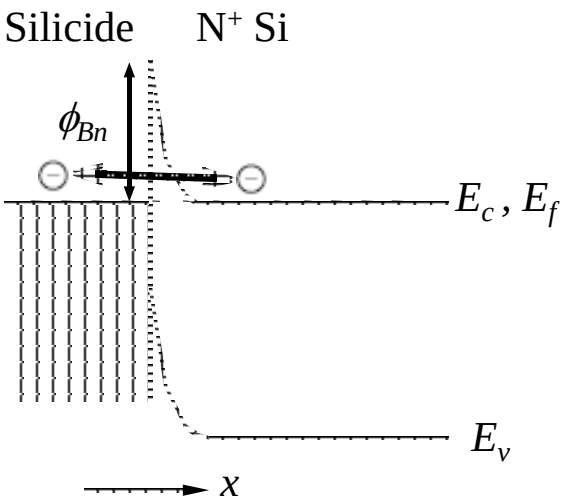
$$P \approx \exp\left(-2T\sqrt{\frac{8\pi^2 m}{h^2}(V_H - E)}\right)$$

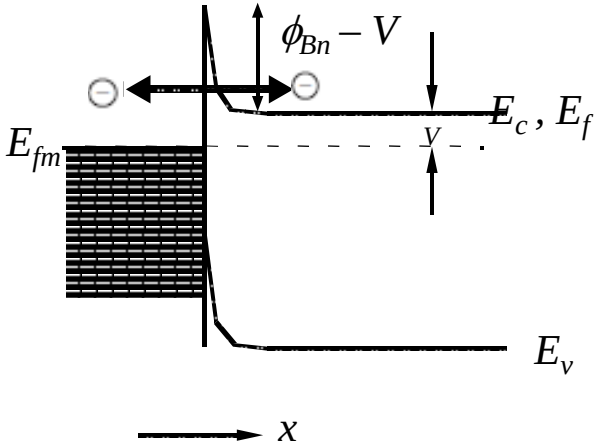
4.21 Ohmic Contacts



4.21 Ohmic Contacts

$$W_{dep} = \sqrt{\frac{2\varepsilon_s \phi_{Bn}}{qN_d}}$$





Tunneling probability:

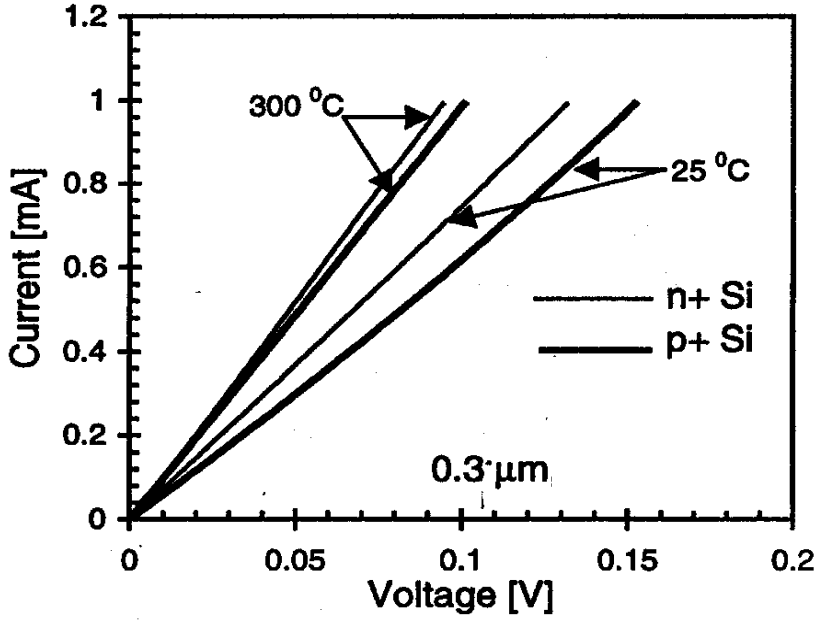
$$P \approx e^{-H\phi_{Bn}/\sqrt{N_d}}$$

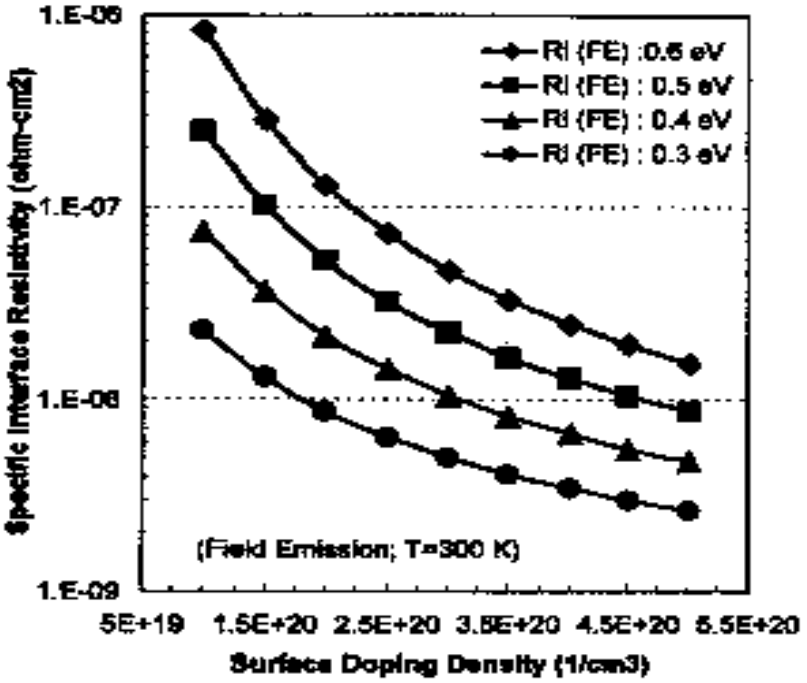
$$T \approx W_{dep}/2 = \sqrt{\varepsilon_s \phi_{Bn}/2qN_d}$$

$$H = \frac{4\pi}{h}\sqrt{\varepsilon_s m_n / q}$$

$$J_{S\to M} \approx \frac{1}{2} qN_d v_{thx} P = qN_d \sqrt{kT/2\pi m_n}\, e^{-H(\phi_{Bn}-V)/\sqrt{N_d}}$$

4.21 Ohmic Contacts





$$R_c \equiv \left(\frac{dJ_{S\to M}}{dV} \right)^{-1} = \frac{2e^{H\phi_{Bn}/\sqrt{N_d}}}{qv_{thx}H\sqrt{N_d}} \propto e^{H\phi_{Bn}/\sqrt{N_d}} \ \Omega\cdot\mathrm{cm}^2$$

4.22 Chapter Summary

Part I: PN Junction

$$\phi_{bi} = \frac{kT}{q} \ln \frac{N_d N_a}{n_i^2}$$

The potential barrier increases by 1 V if a 1 V reverse bias is applied

depletion width

$$W_{dep} = \sqrt{\frac{2\varepsilon_s \cdot potential\ barrier}{qN}}$$

junction capacitance

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}}$$

4.22 Chapter Summary

- Under forward bias, minority carriers are injected across the jucntion.

- The quasi-equilibrium boundary condition of minority carrier densities is:

$$n(x_p) = n_{P0} e^{qV/kT}$$

$$p(x_N) = p_{N0} e^{qV/kT}$$

- Most of the minority carriers are injected into the more lightly doped side.

4.22 Chapter Summary

- Steady-state continuity equation:

$$\frac{d^2 p'}{dx^2} = \frac{p'}{D_p \tau_p} = \frac{p'}{L_p^{\ 2}}$$

$$L_p \equiv \sqrt{D_p \tau_p}$$

- Minority carriers diffuse outward $\propto e^{-|x|/L_p}$ and $e^{-|x|/L_n}$
- L_p and L_n are the diffusion lengths

$$I = I_0 (e^{qV/kT} - 1)$$

$$I_0 = Aqn_i^{\ 2} \left(\frac{D_p}{L_p N_d} + \frac{D_n}{L_n N_a} \right)$$

4.22 Chapter Summary

Charge storage: $Q = I\tau_s$

Diffusion capacitance: $C = \tau_s G$

Diode conductance: $G = I_{DC} / \frac{kT}{q}$

4.22 Chapter Summary

Part II: Optoelectronic Applications

$$\text{Solar cell power} = I_{sc} \times V_{oc} \times FF$$

•~100um Si or <1um direct–gap semiconductor can absorb most of solar photons with energy larger than E_g.

•Carriers generated within diffusion length from the junction can be collected and contribute to the Short Circuit Current I_{sc}.

•Theoretically, the highest efficiency (~24%) can be obtained with 1.9eV >E_g>1.2eV. Larger E_g lead to too low I_{sc} (low light absorption); smaller E_g leads to too low Open Circuit VoltageVoc.

•Si cells with ~15% efficiency dominate the market. >2x cost reduction (including package and installation) is required to achieve cost parity with base-load non-renewable electricity.

4.22 Chapter Summary

LED and Solid-State Lighting

•Electron-hole recombination in direct-gap semiconductors such as GaAs produce light.

•Tenary semiconductors such as GaAsP provide tunable E_g and LED color.

•Quatenary semiconductors such as AlInGaP provide tunable E_g and lattice constants for high quality epitaxial growth on inexpensive substrates.

•Beyond displays, communication, and traffic lights, a new application is space lighting with luminous efficacy >5x higher than incandescent lamps. White light can be obtained with UV LED and phosphors. Cost still an issue.

•Organic semiconductor is an important low-cost LED material class.

4.22 Chapter Summary

Laser Diodes

•Light is amplified under the condition of population inversion – states at higher E have higher probability of occupation than states at lower E.

•Population inversion occurs when diode forward bias $qV > E_g$.

•Optical feedback is provided with cleaved surfaces or distributed Bragg reflectors.

•When the round-trip gain (including loss at reflector) exceeds unity, laser threshold is reached.

•Quantum-well structures significantly reduce the threshold currents.

•Purity of laser light frequency enables long-distance fiber-optic communication. Purity of light direction allows focusing to tiny spots and enables DVD writer/reader and other application.

4.22 Chapter Summary

Part III: Metal-Semiconductor Junction

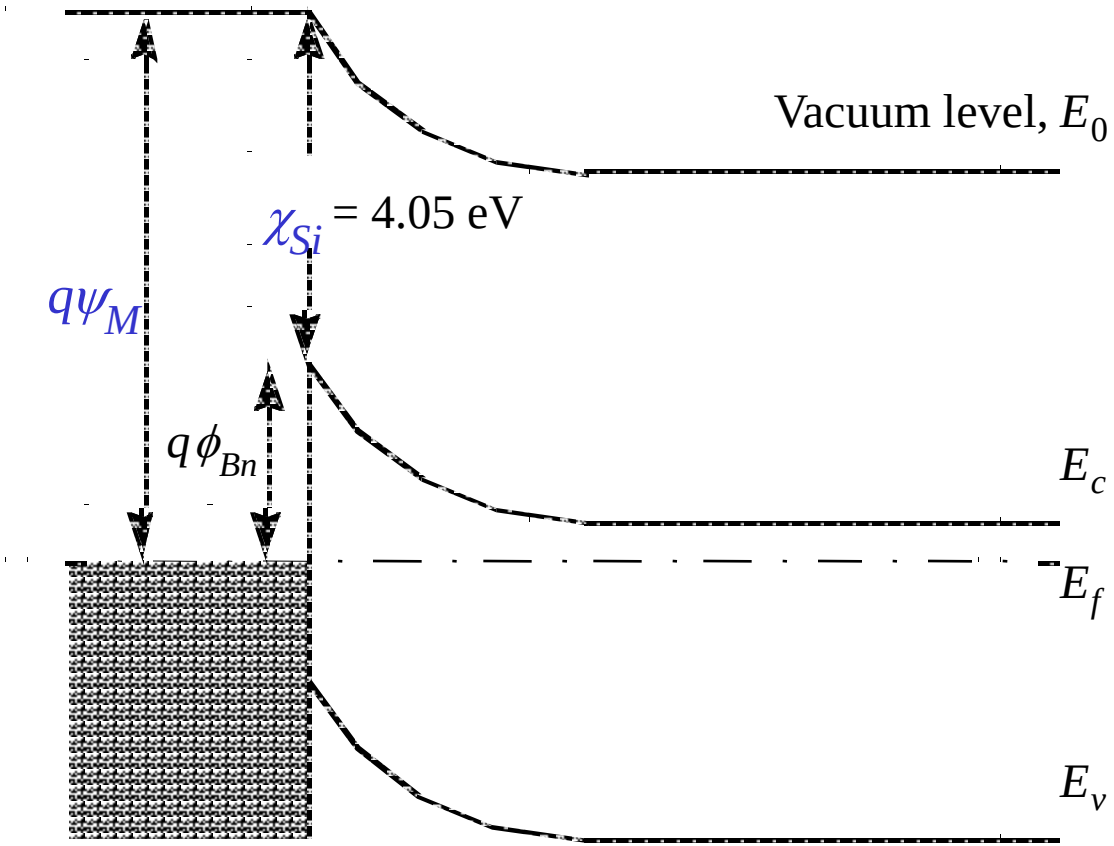
$$I_0 = AKT^2 e^{-q\phi_B / kT}$$

•Schottky diodes have large reverse saturation current, determined by the Schottky barrier height ϕ_B, and therefore lower forward voltage at a given current density.

•Ohmic contacts relies on tunneling. Low resistance contact requires low ϕ_B and higher doping concentration.

$$R_c \propto e^{-(\frac{4\pi}{h}\phi_B \sqrt{\varepsilon_s m_n / qN_d})} \Omega \cdot \text{cm}^2$$

ϕ_{Bn} Increases with Increasing Metal Work Function

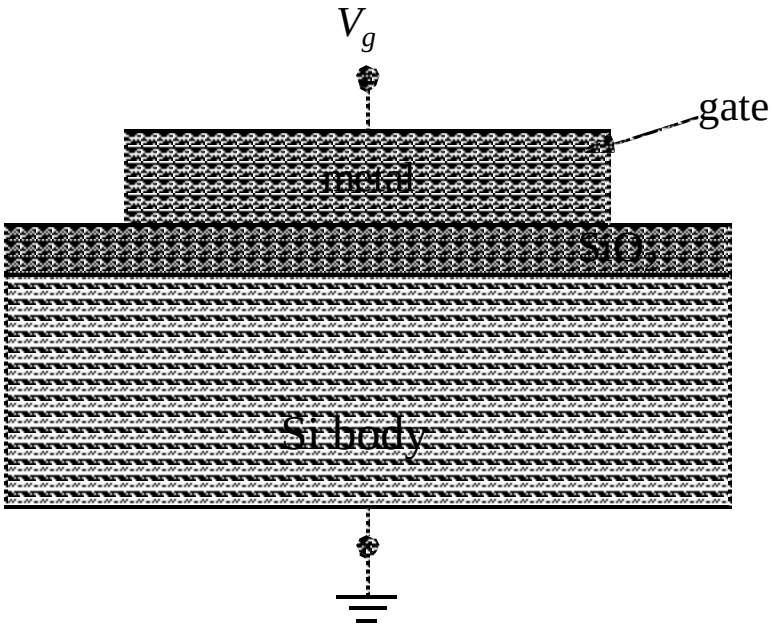


Ideally,

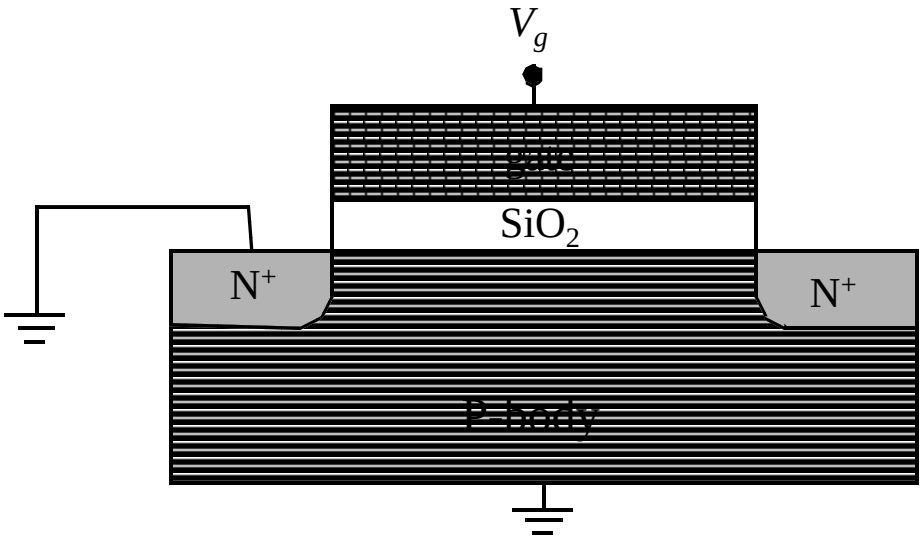
$\phi_{Bn} = \psi_M - \chi_{Si}$

Chapter 5 MOS Capacitor

MOS: Metal-Oxide-Semiconductor

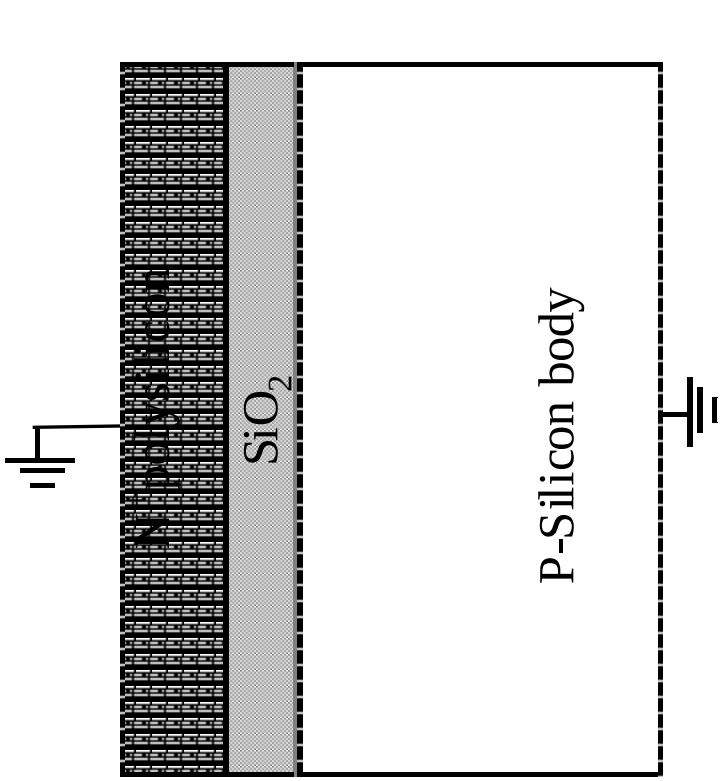


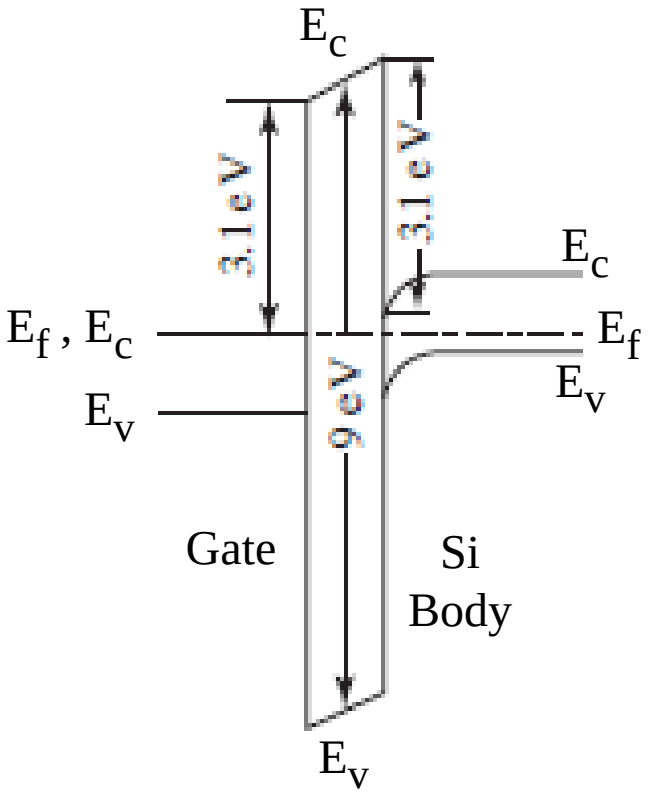
MOS capacitor



MOS transistor

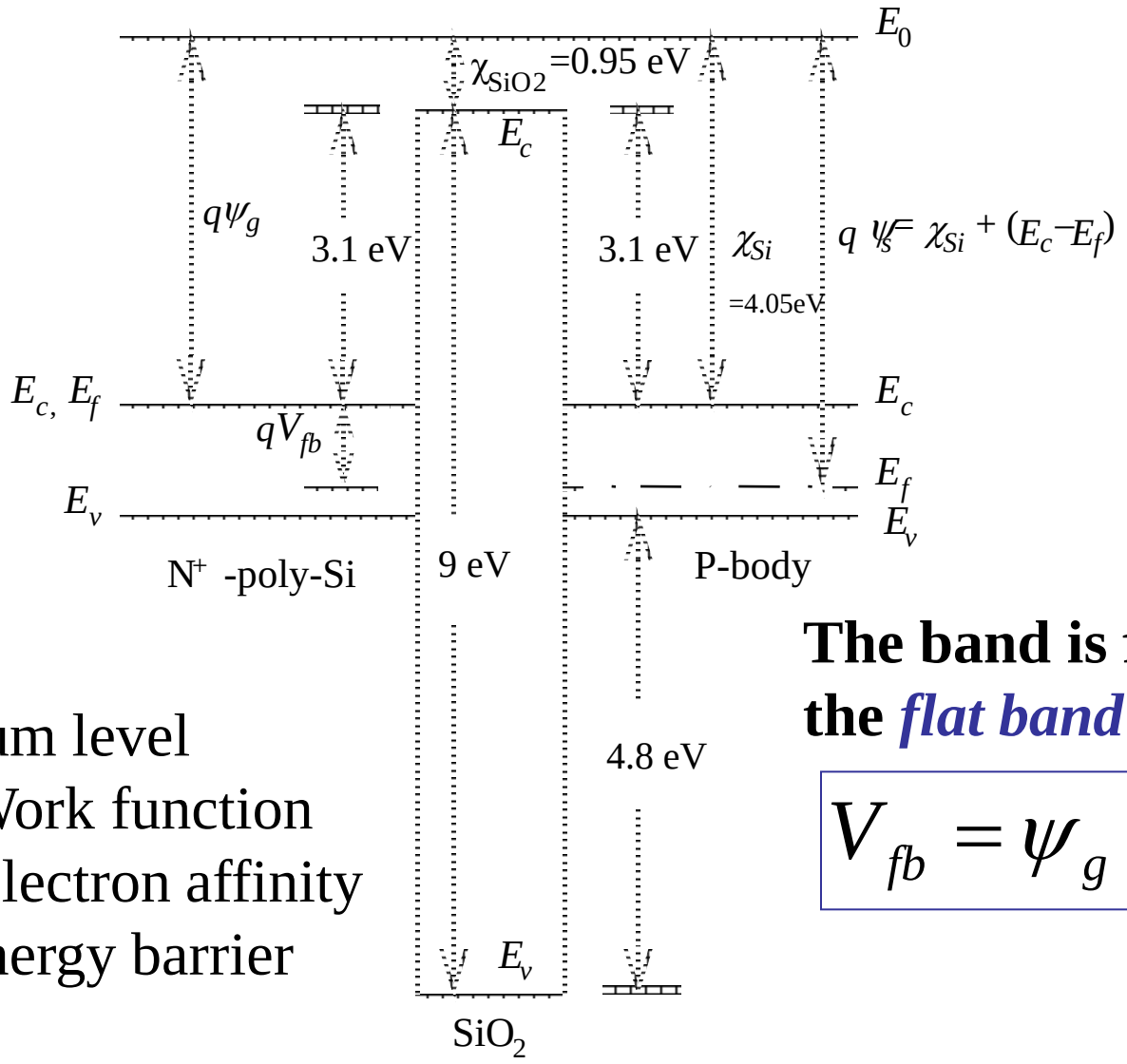
Chapter 5 MOS Capacitor





This energy-band diagram for $V_g = 0$ is not the simplest one.

5.1 Flat-band Condition and Flat-band Voltage



E_0 : Vacuum level

$E_0 - E_f$: Work function

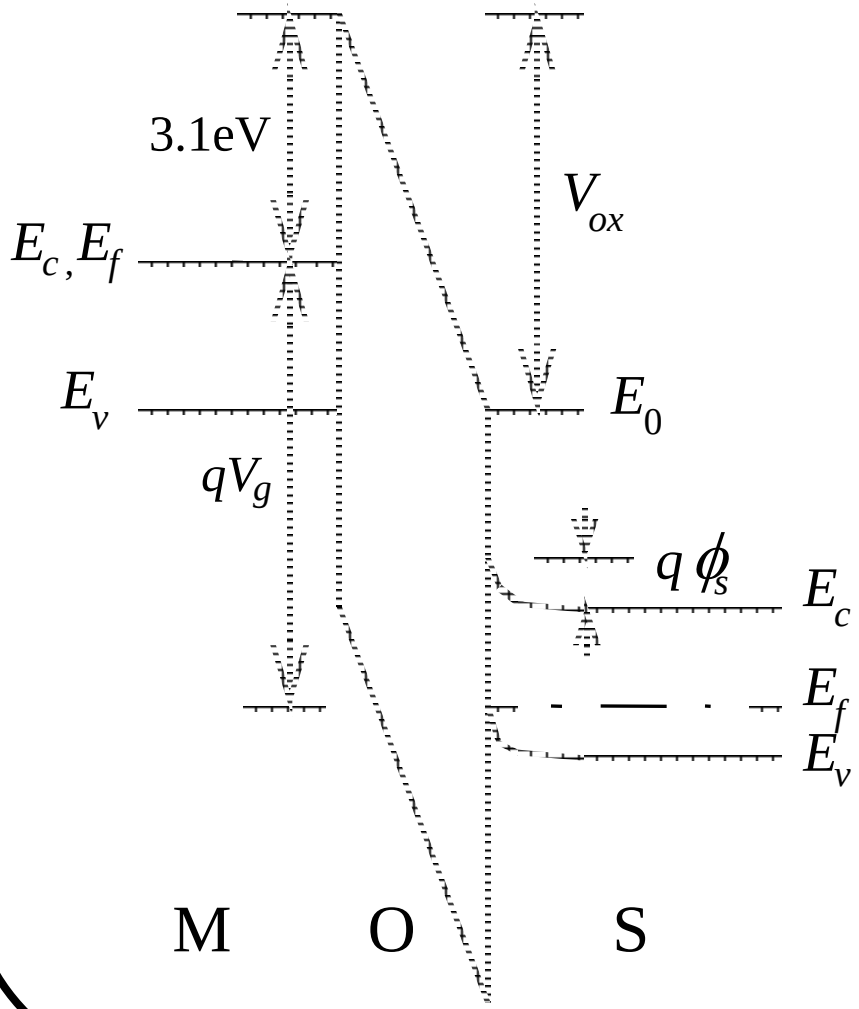
$E_0 - E_c$: Electron affinity

Si/SiO_2 energy barrier

The band is flat at the *flat band voltage*.

$$V_{fb} = \psi_g - \psi_s$$

5.2 Surface Accumulation



Make $V_g < V_{fb}$

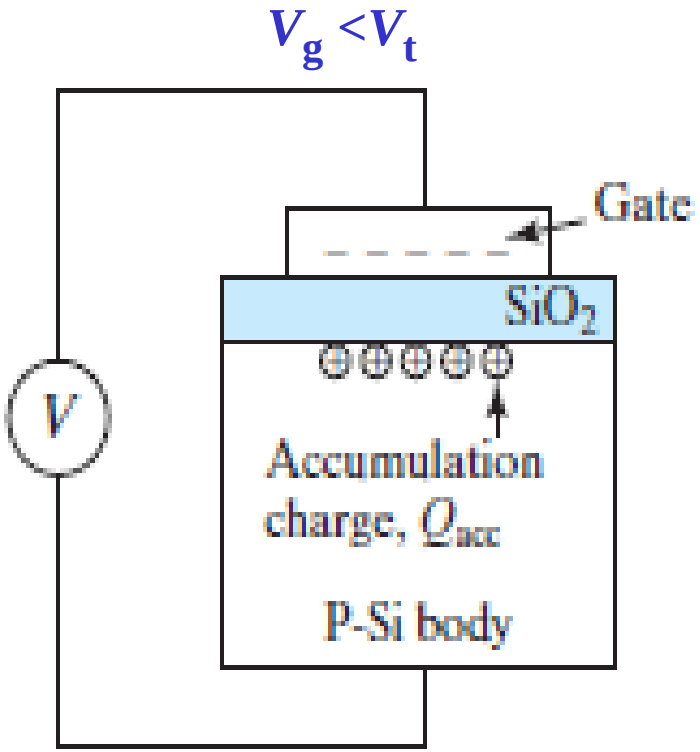
$$V_g = V_{fb} + \phi_s + V_{ox}$$

ϕ_s : surface potential, band bending

V_{ox}: voltage across the oxide

ϕ_s is negligible when the surface is in accumulation.

5.2 Surface Accumulation



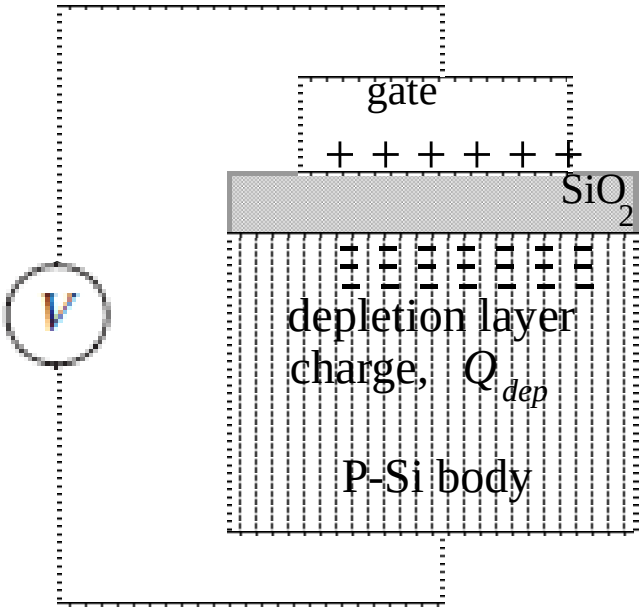
$$V_{ox} = V_g - V_{fb}$$

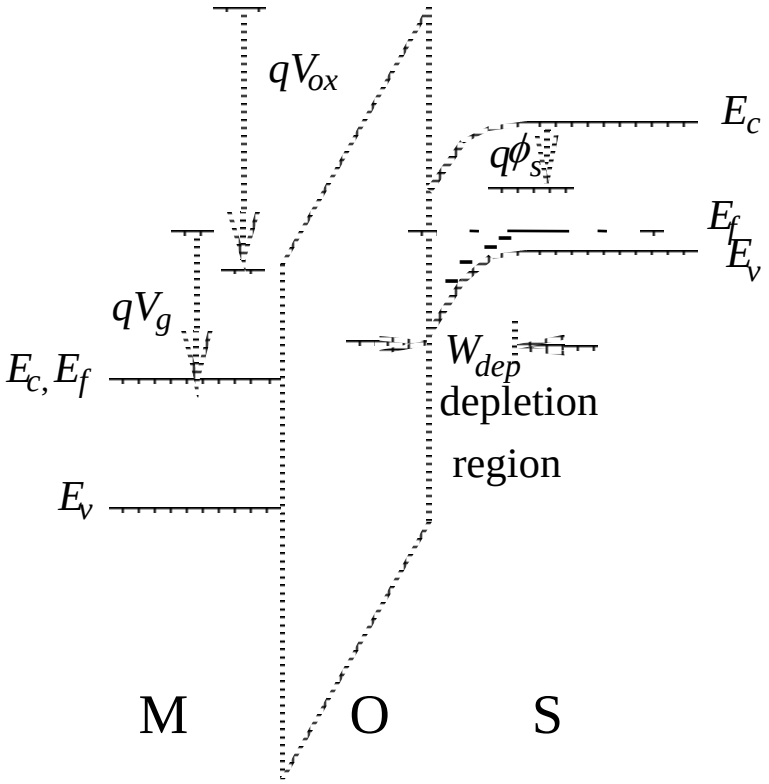
Gauss's Law $\rightarrow V_{ox} = -Q_{acc} / C_{ox}$

$$Q_{acc} = -C_{ox}(V_g - V_{fb})$$

$$V_{ox} = -Q_s / C_{ox}$$

5.3 Surface Depletion ($V_g > V_{fb}$)





$$V_{ox} = -\frac{Q_s}{C_{ox}} = -\frac{Q_{dep}}{C_{ox}} = \frac{qN_a W_{dep}}{C_{ox}} = \frac{\sqrt{qN_a 2\varepsilon_s \phi_s}}{C_{ox}}$$

5.3 Surface Depletion

$$V_g = V_{fb} + \phi_s + V_{ox} = V_{fb} + \phi_s + \frac{\sqrt{qN_a 2\varepsilon_s \phi_s}}{C_{ox}}$$

This equation can be solved to yield ϕ_s .

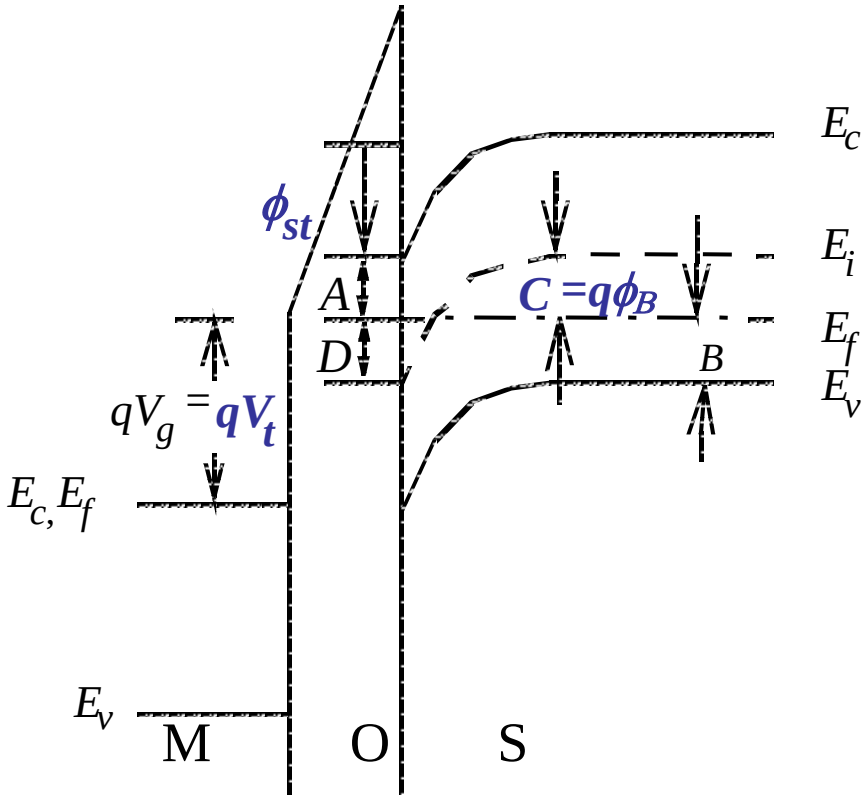
5.4 Threshold Condition and Threshold Voltage

Threshold (of inversion):

$n_s = N_a$, or

$(E_c - E_f)_{surface} = (E_f - E_v)_{bulk}$, or

✧ $A = B$, and $C = D$



$$\phi_{st} = 2\phi_B = 2\frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

$$q\phi_B = \frac{E_g}{2} - (E_f - E_v)\big|_{bulk} = \frac{kT}{q}\ln\left(\frac{N_v}{n_i}\right) - \frac{kT}{q}\ln\left(\frac{N_v}{N_a}\right) = \frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

Threshold Voltage

$$V_g = V_{fb} + \varphi_s + V_{ox}$$

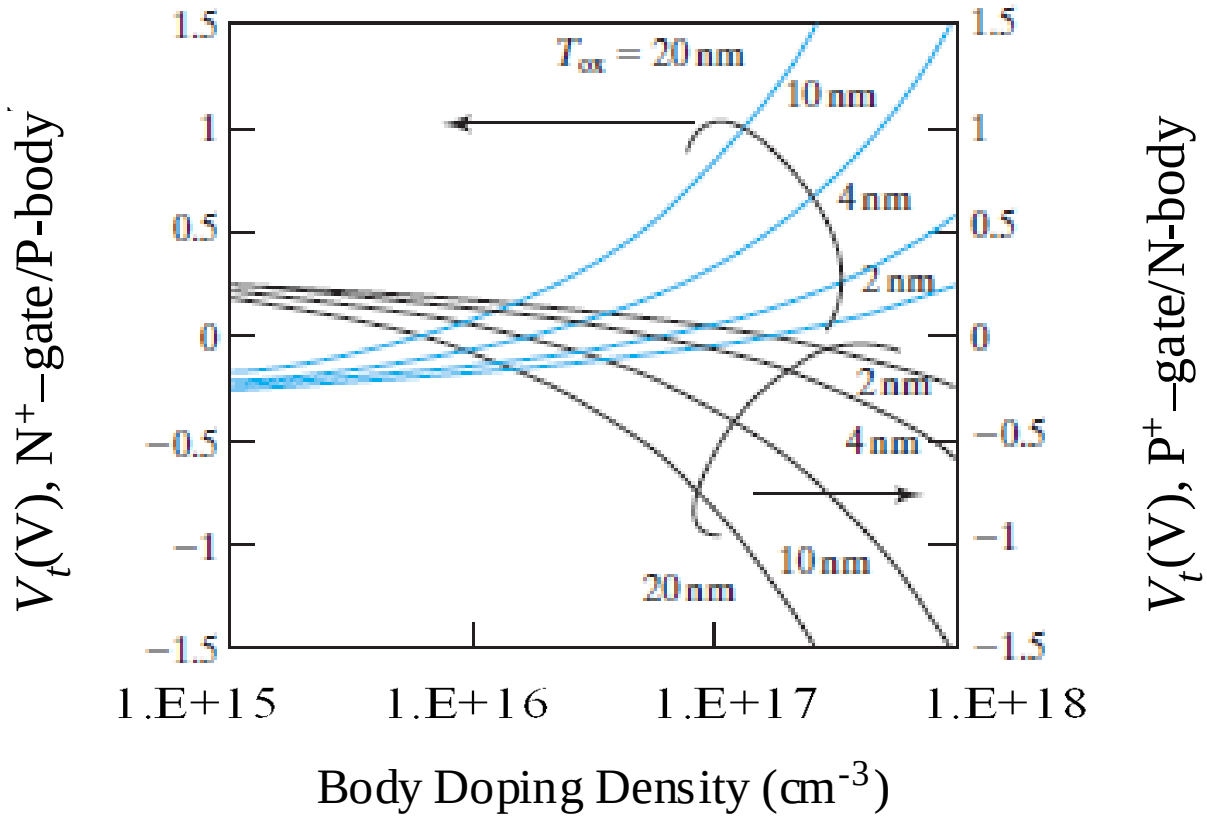
At threshold,

$$\varphi_{st} = 2\phi_B = 2\frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

$$V_{ox} = \frac{\sqrt{qN_a 2\varepsilon_s 2\phi_B}}{C_{ox}}$$

$$V_t = V_g \text{ at threshold} = V_{fb} + 2\phi_B + \frac{\sqrt{qN_a 2\varepsilon_s 2\phi_B}}{C_{ox}}$$

Threshold Voltage



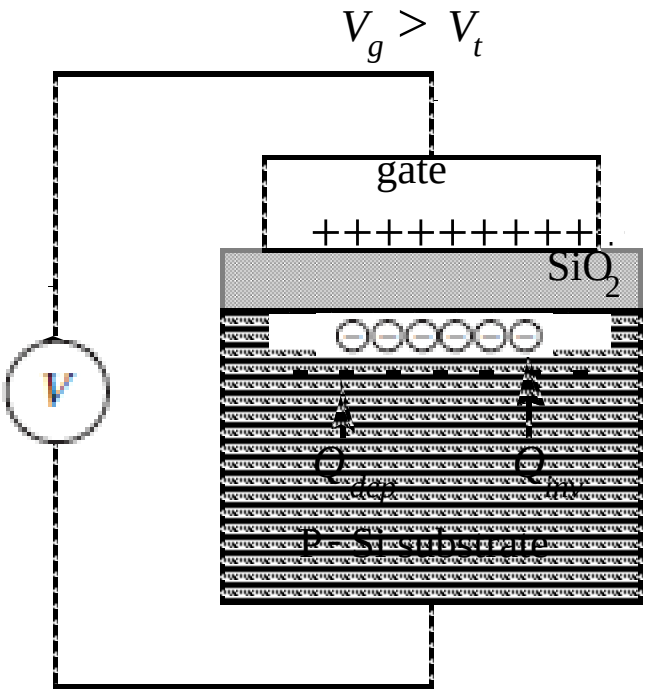
$$V_t = V_{fb} \pm 2\phi_B \pm \frac{\sqrt{qN_{sub}\,2\varepsilon_s\,2\phi_B}}{C_{ox}}$$

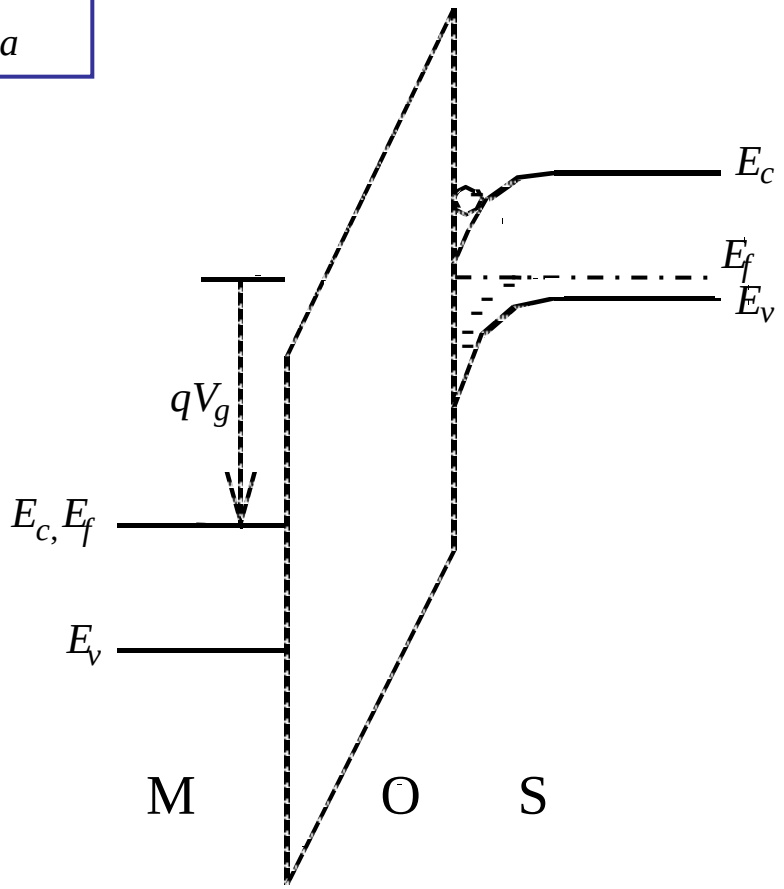
+ for P-body,
– for N-body

5.5 Strong Inversion–Beyond Threshold

$V_g > V_t$

$$W_{dep} = W_{dmax} = \sqrt{\frac{2\varepsilon_s 2\phi_B}{qN_a}}$$

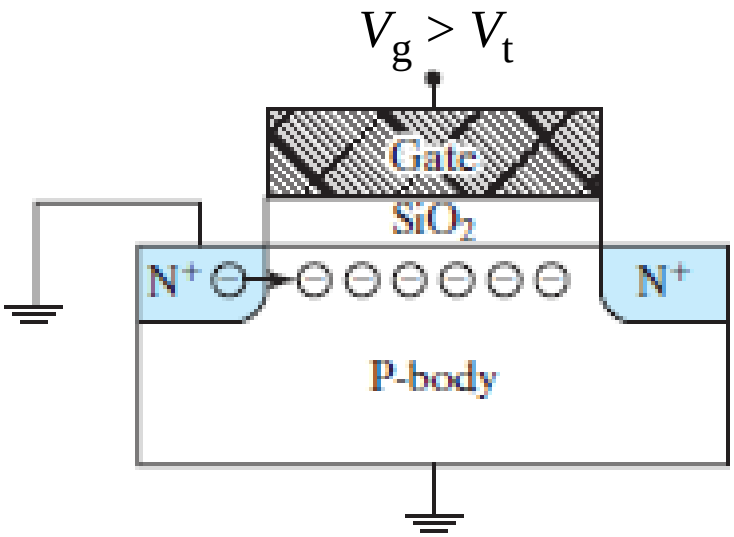


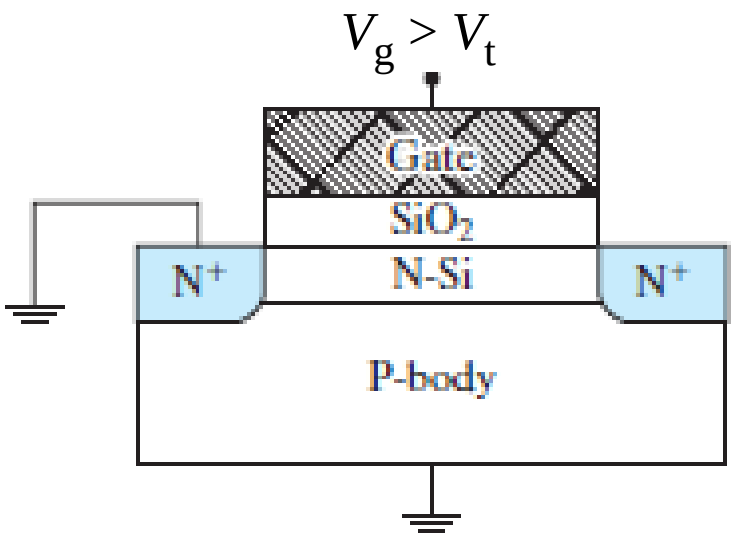


Inversion Layer Charge, Q_{inv} (C/cm²)

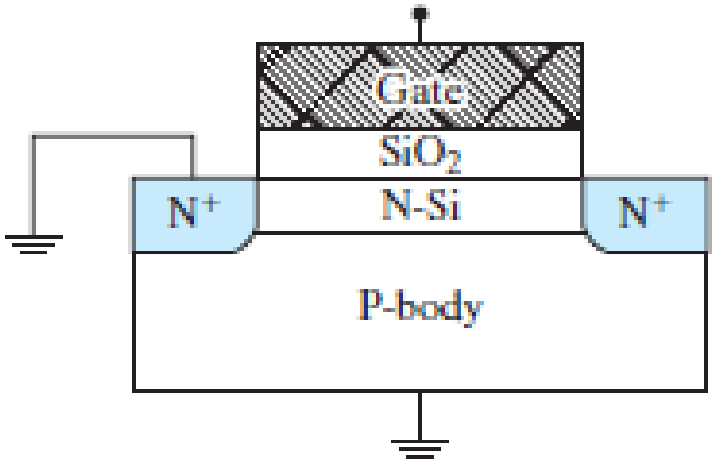
$$V_g = V_{fb} + 2\phi_B - \frac{Q_{dep}}{C_{ox}} - \frac{Q_{inv}}{C_{ox}} = V_{fb} + 2\phi_B + \frac{\sqrt{qN_a 2\varepsilon_s 2\phi_B}}{C_{ox}} - \frac{Q_{inv}}{C_{ox}}$$

$$= V_t - \frac{Q_{inv}}{C_{ox}} \qquad \therefore \qquad \boxed{Q_{inv} = -C_{ox}(V_g - V_t)}$$





5.5.1 Choice of V_t and Gate Doping Type

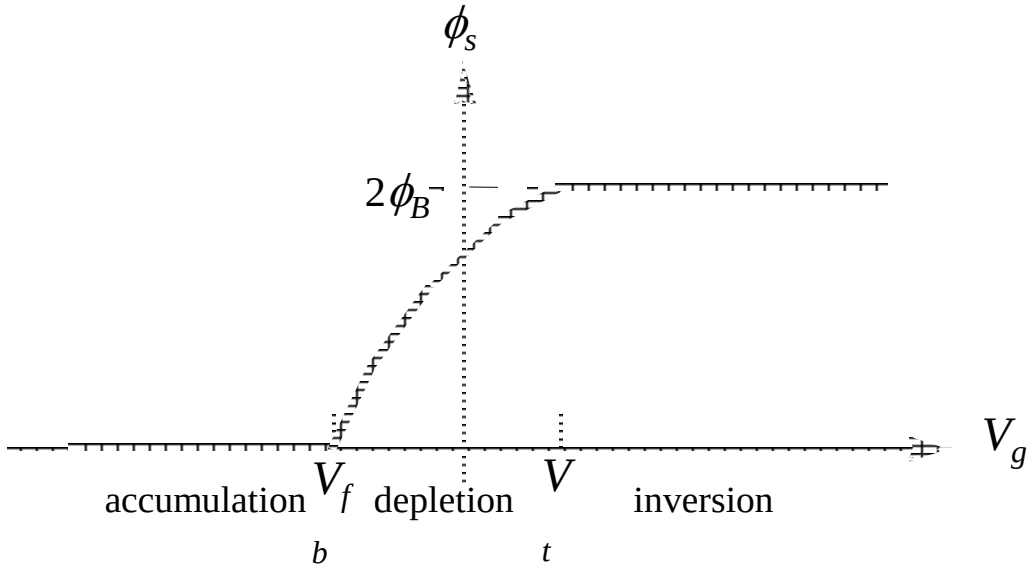


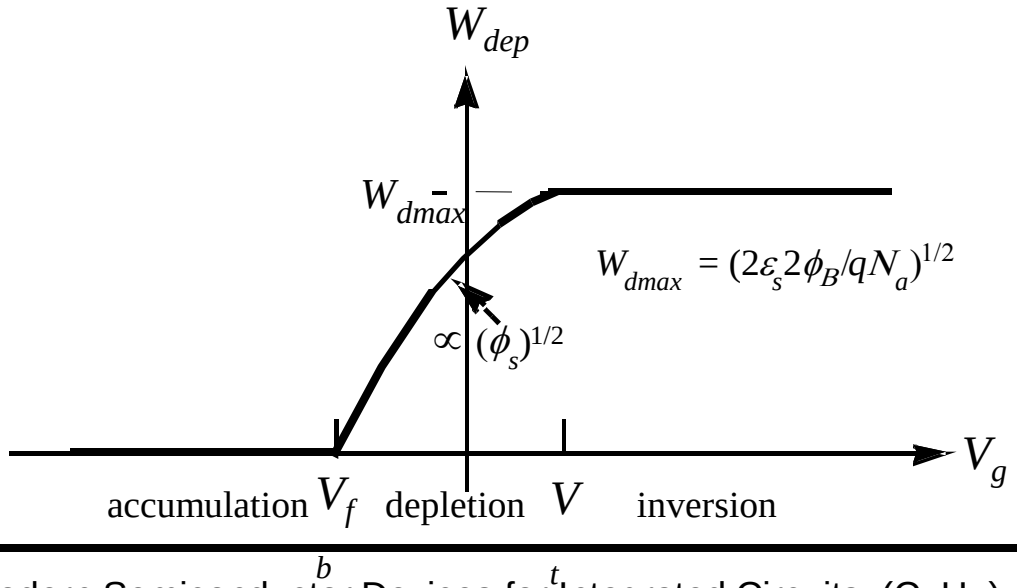
V_t is generally set at a small positive value so that, at V_g = 0, the transistor does not have an inversion layer and current does not flow between the two N^+ regions

• P-body is normally paired with N^+-gate to achieve a small positive threshold voltage.

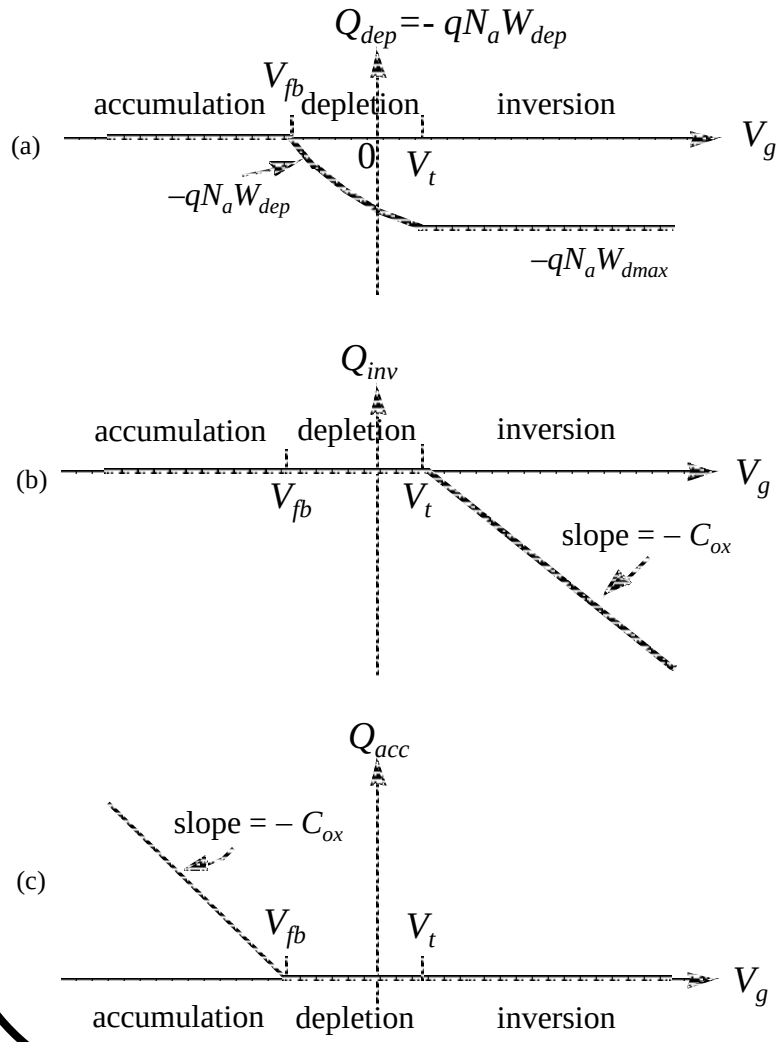
• N-body is normally paired with P^+-gate to achieve a small negative threshold voltage.

Review : Basic MOS Capacitor Theory



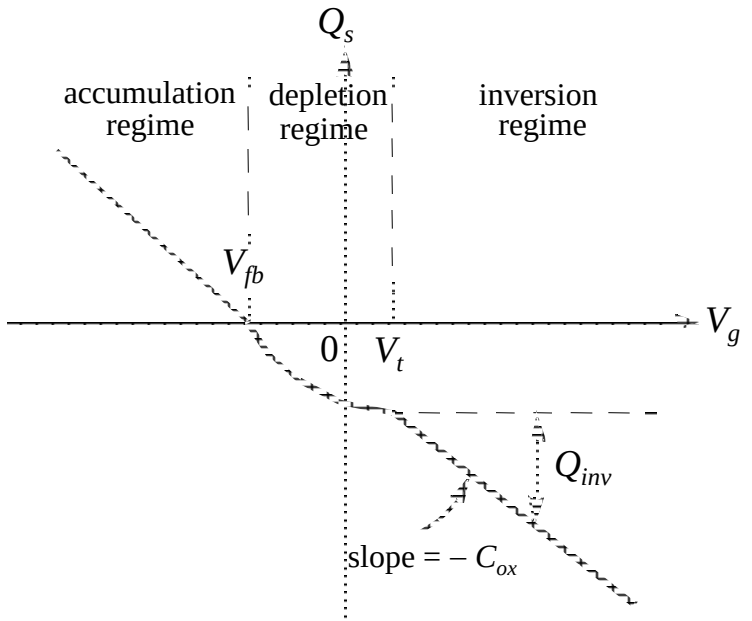


Review : Basic MOS Capacitor Theory

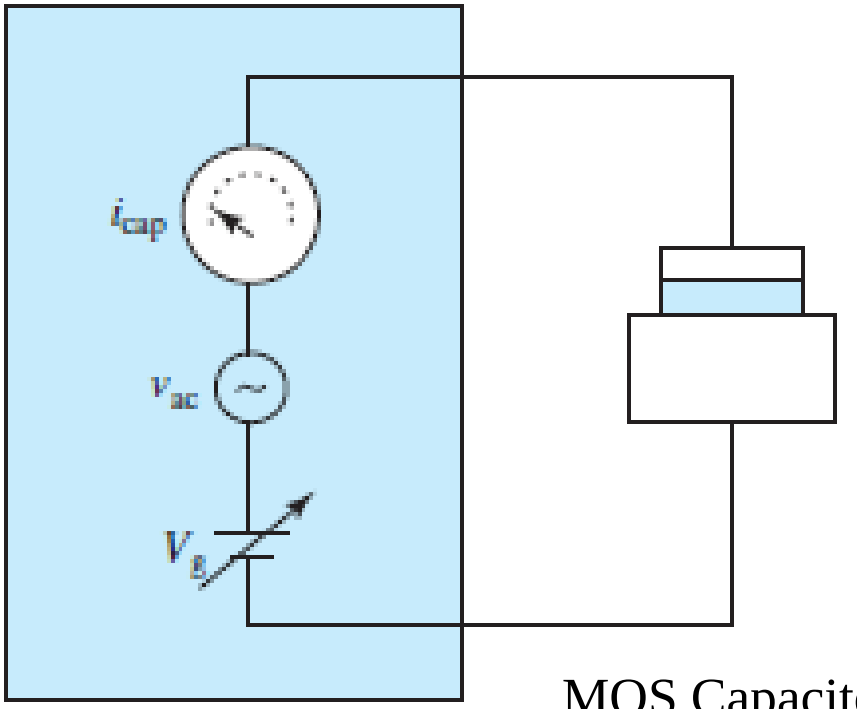


total substrate charge, Q_s

$$Q_s = Q_{acc} + Q_{dep} + Q_{inv}$$



5.6 MOS CV Characteristics



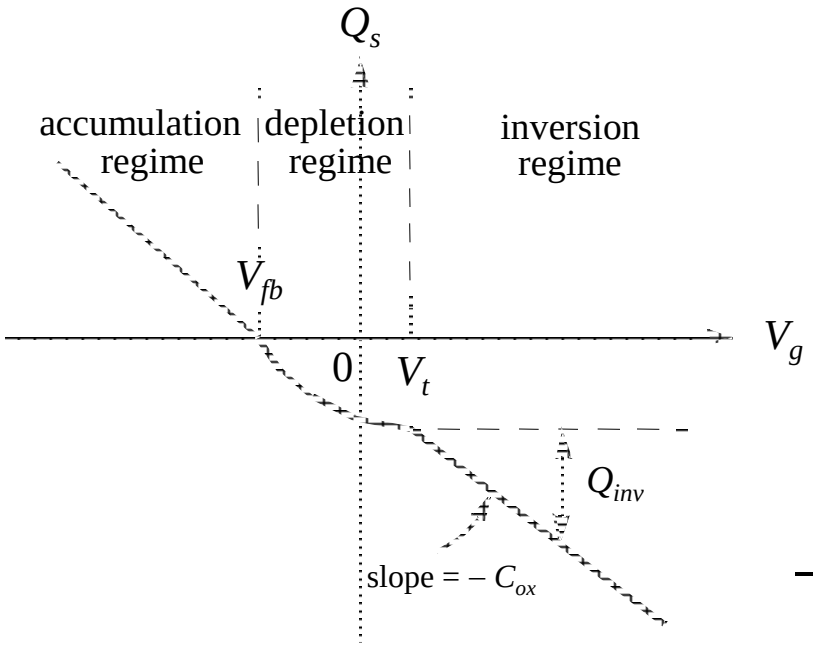
C-V Meter

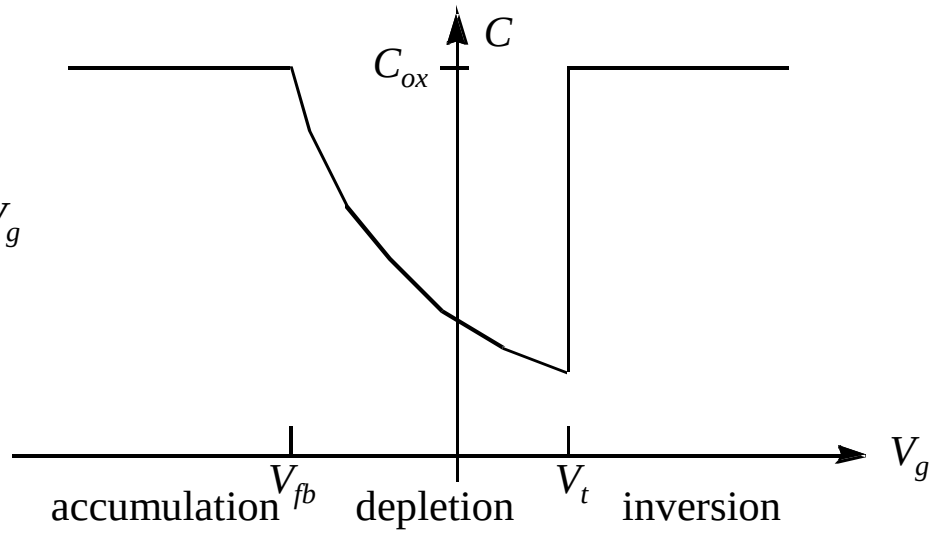
MOS Capacitor

$$C = \frac{dQ_g}{dV_g} = -\frac{dQ_s}{dV_g}$$

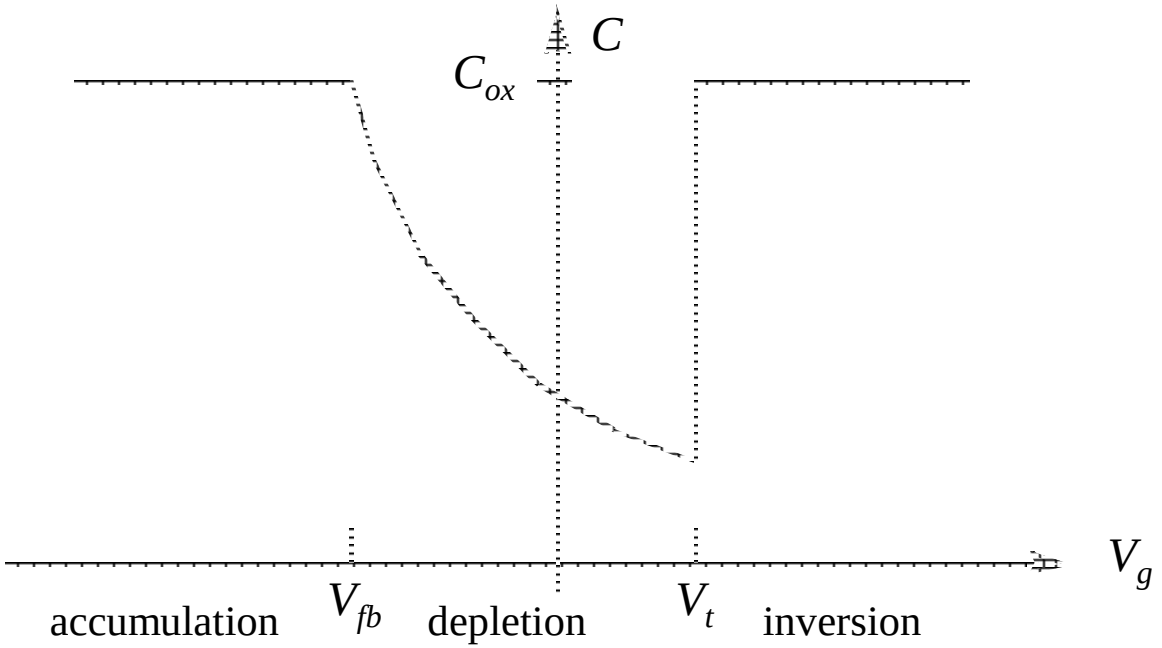
5.6 MOS CV Characteristics

$$C = \frac{dQ_g}{dV_g} = -\frac{dQ_s}{dV_g}$$





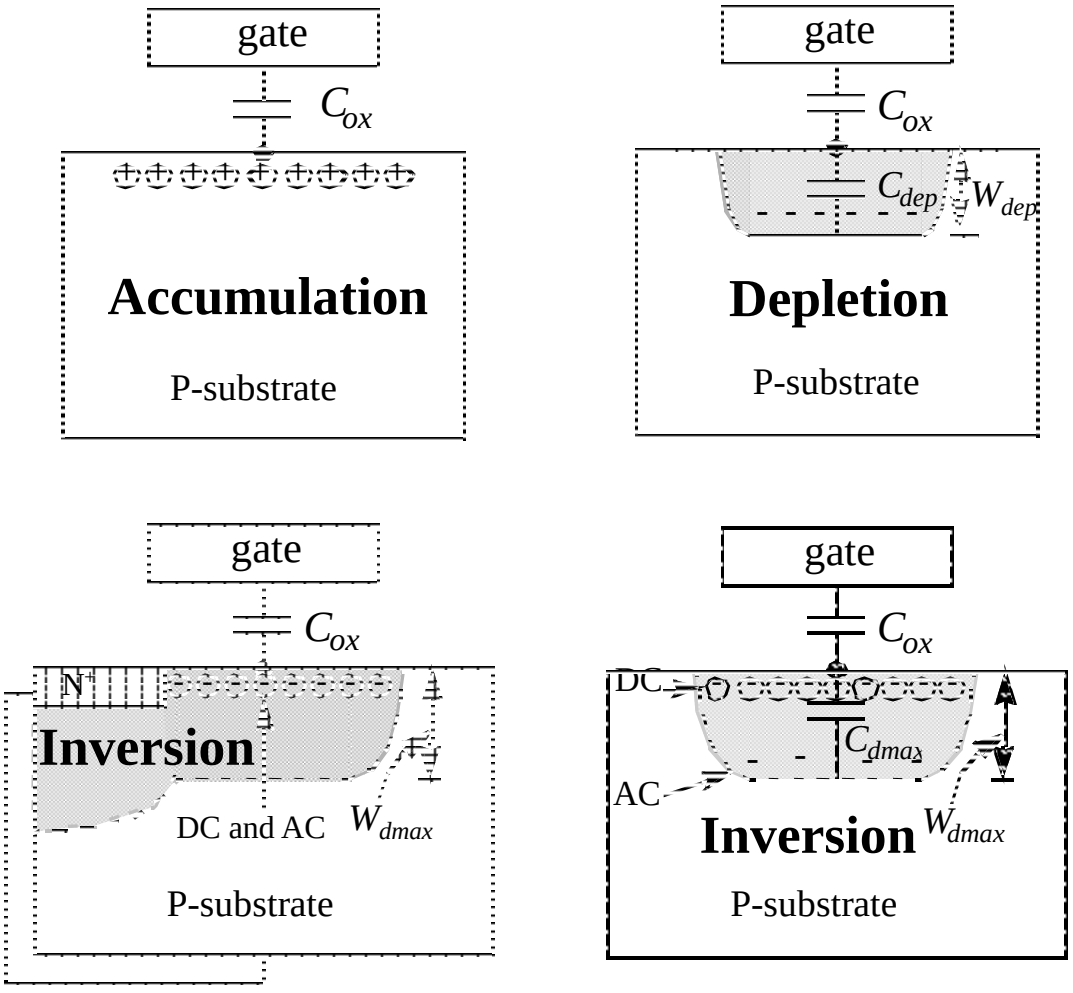
CV Characteristics



In the depletion regime: $\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}}$

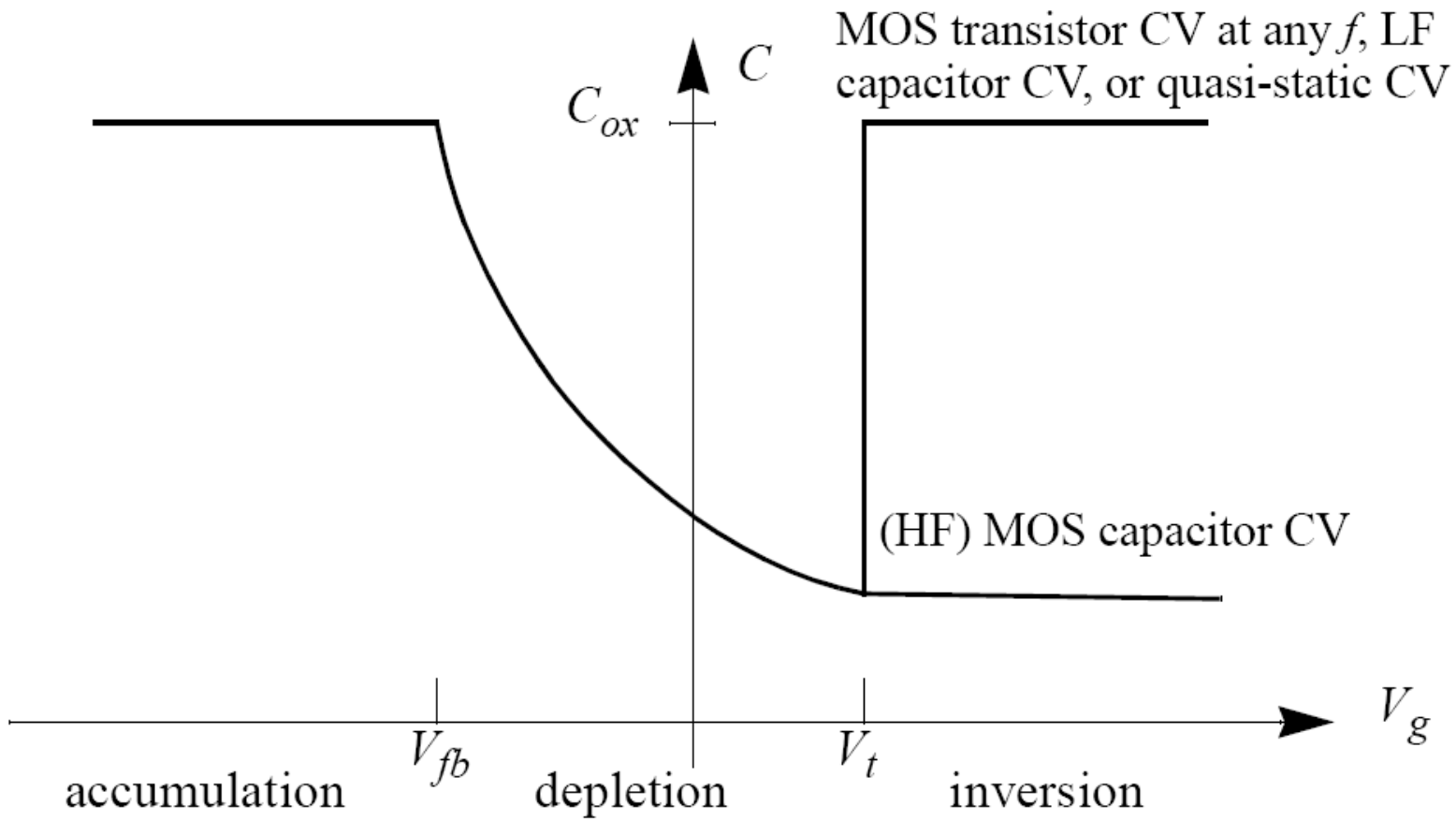
$$\frac{1}{C} = \sqrt{\frac{1}{C_{ox}^2} + \frac{2(V_g - V_{fb})}{qN_a\varepsilon_s}}$$

Supply of Inversion Charge May be Limited

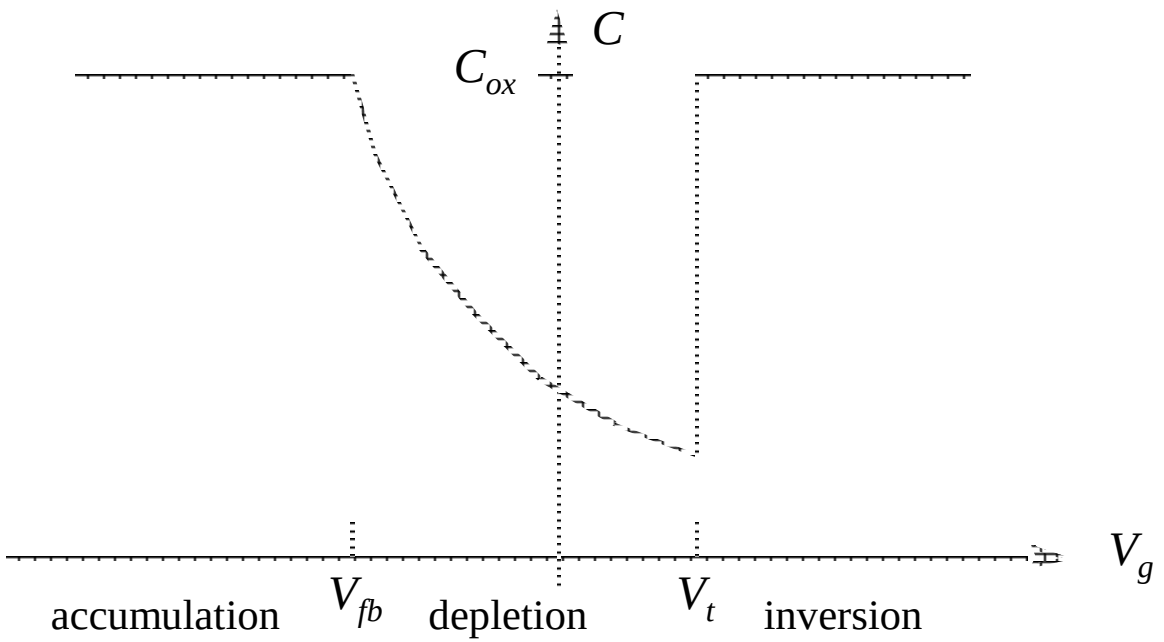


In each case, C = ?

Capacitor and Transistor CV (or HF and LF CV)

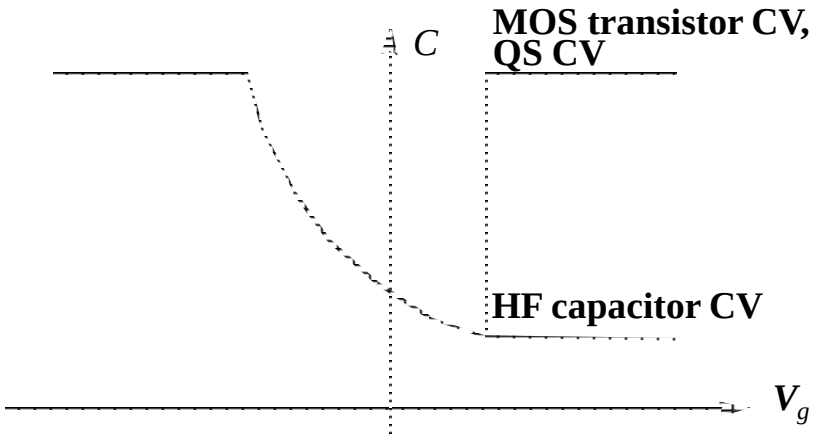


Quasi-Static CV of MOS Capacitor



The quasi-static CV is obtained by the application of a slow linear-ramp voltage (< 0.1V/s) to the gate, while measuring I_g with a very sensitive DC ammeter. C is calculated from $I_g = C \cdot dV_g/dt$. This allows sufficient time for Q_{inv} to respond to the slow-changing V_g .

EXAMPLE : CV of MOS Capacitor and Transistor



Does the QS CV or the HF capacitor CV apply?

(1) MOS transistor, 10kHz. *(Answer: QS CV).*

(2) MOS transistor, 100MHz. *(Answer: QS CV).*

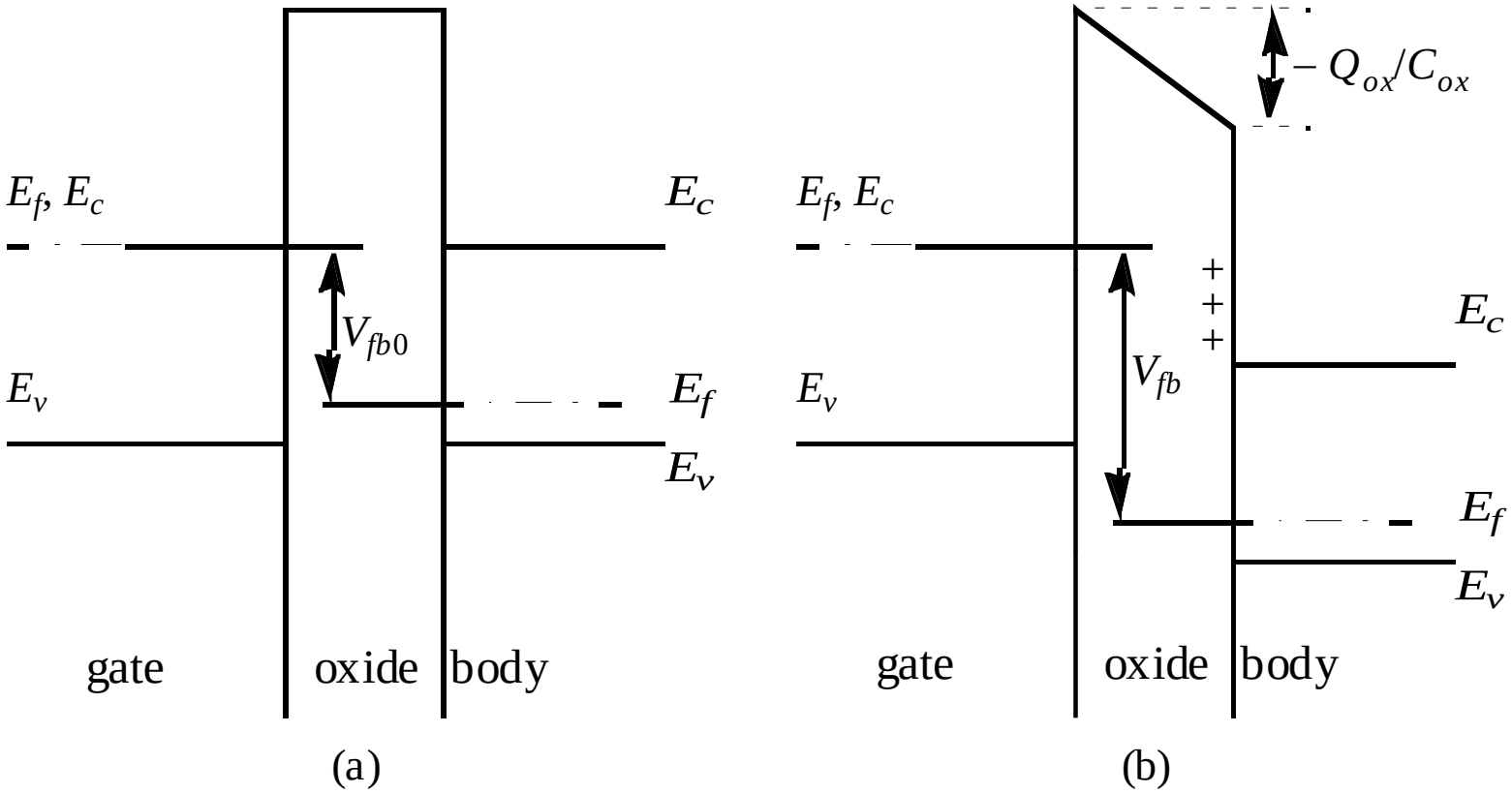
(3) MOS capacitor, 100MHz. *(Answer: HF capacitor CV).*

(4) MOS capacitor, 10kHz. *(Answer: HF capacitor CV).*

(5) MOS capacitor, slow V_g ramp. *(Answer: QS CV).*

(6) MOS transistor, slow V_g ramp. *(Answer: QS CV).*

5.7 Oxide Charge–A Modification to V_{fb} and V_t



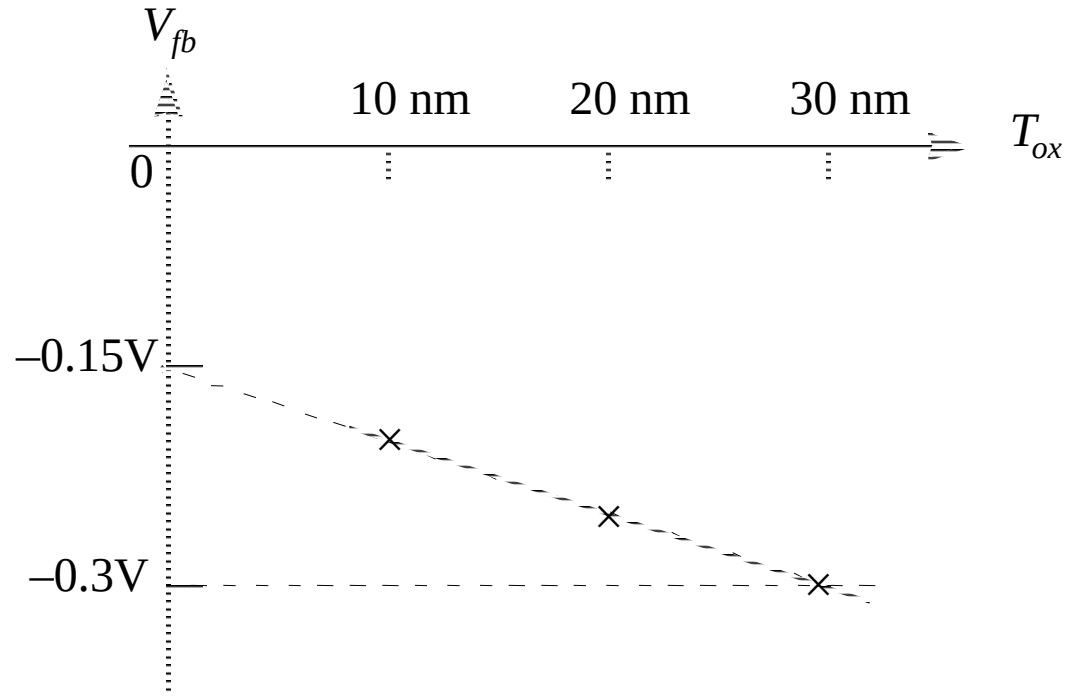
$$V_{fb} = V_{fb0} - Q_{ox}/C_{ox} = \psi_g - \psi_s - Q_{ox}/C_{ox}$$

5.7 Oxide Charge–A Modification to V_{fb} and V_t

Types of oxide charge:

- Fixed oxide charge, Si^+
- Mobile oxide charge, due to Na^+contamination
- Interface traps, neutral or charged depending on Vg.
- Voltage/temperature stress induced charge and traps--a reliability issue

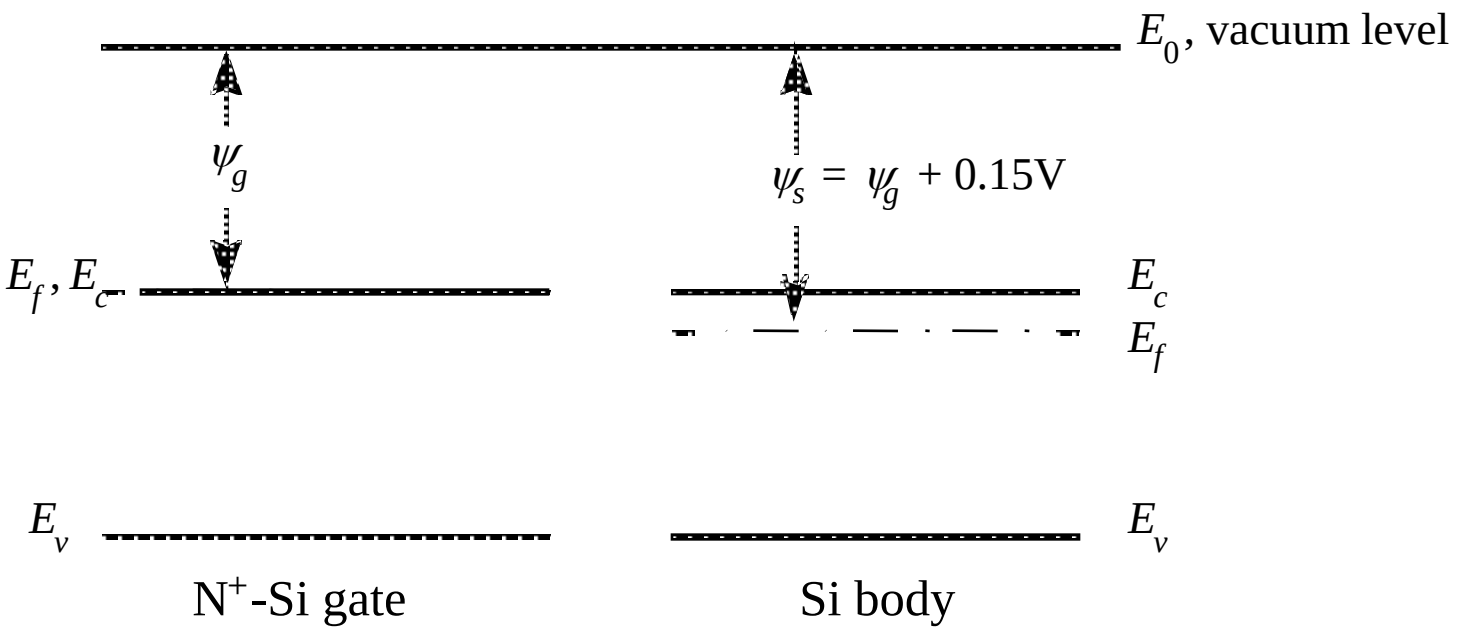
EXAMPLE: *Interpret this measured V_{fb} dependence on oxide thickness. The gate electrode is N^+ poly-silicon.*



What does it tell us? Body work function? Doping type? Other?

Solution: $V_{fb} = \psi_g - \psi_s - Q_{ox}T_{ox} / \varepsilon_{ox}$

$$\textit{from intercept} \longrightarrow \psi_g - \psi_s = -0.15\ \text{V}$$

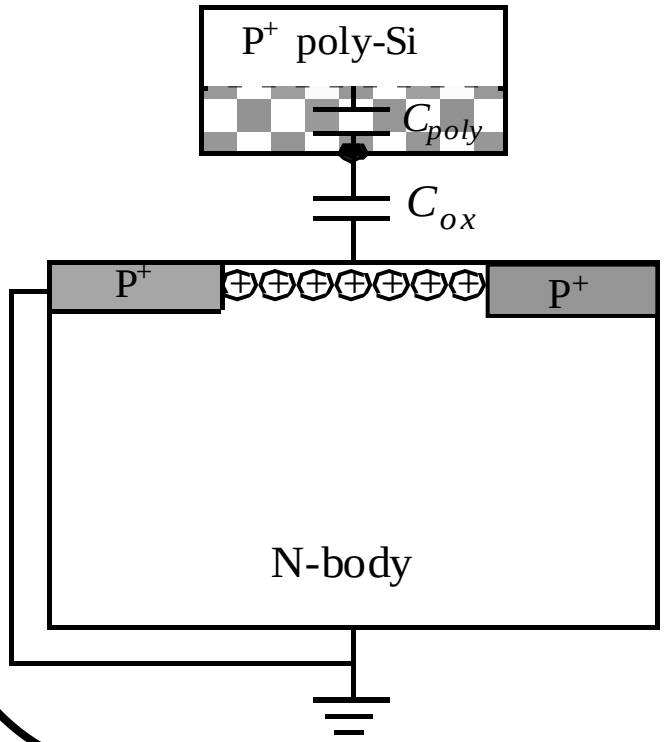


$$\textit{N-type substrate,}\ \ N_d = n = N_c e^{-0.15\,\text{eV}/kT} \approx 10^{17}\ \text{cm}^{-3}$$

$$\textit{from slope} \longrightarrow Q_{ox} = 1.7\times10^{-8}\ \text{C/cm}^2$$

5.8 Poly-Silicon Gate Depletion–Effective Increase in T_{ox}

Gauss's Law $W_{dpoly} = \varepsilon_{ox}\mathbf{E}_{ox} / qN_{poly}$



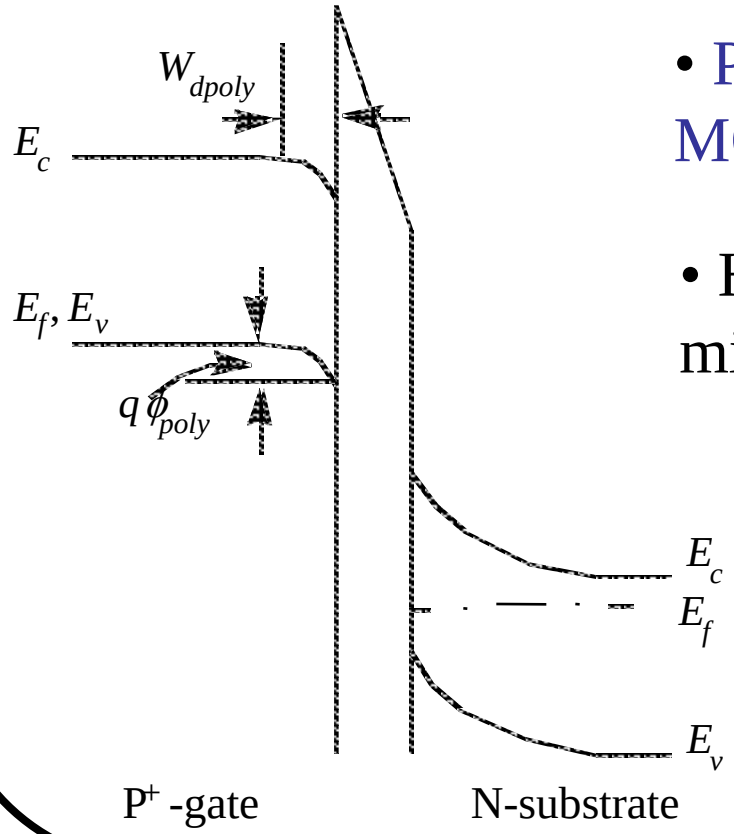
$$C = \left(\frac{1}{C_{ox}} + \frac{1}{C_{poly}} \right)^{-1} = \left(\frac{T_{ox}}{\varepsilon_{ox}} + \frac{W_{dpoly}}{\varepsilon_s} \right)^{-1}$$

$$= \frac{\varepsilon_{ox}}{T_{ox} + W_{dpoly} / 3}$$

If W_{dpoly}= 15 Å, what is the effective increase in T_{ox}?

Effect of Poly-Gate Depletion on Q_{inv}

$$Q_{inv} = C_{ox}(V_g - \phi_{poly} - V_t)$$



- Poly-gate depletion degrades MOSFET current and circuit speed.
- How can poly-depletion be minimized?

EXAMPLE : Poly-Silicon Gate Depletion

V_{ox} , the voltage across a 2 nm thin oxide, is –1 V. The P^+ poly-gate doping is $N_{poly} = 8 \times 10^{19}\ cm^{-3}$ and substrate N_d is $10^{17} cm^{-3}$. Find (a) W_{dpoly}, (b) ϕ_{poly}, and (c) V_g .

Solution:

(a) $$W_{dpoly} = \varepsilon_{ox}\mathcal{E}_{ox} / qN_{poly} = \varepsilon_{ox}V_{ox} / T_{ox}qN_{poly}$$

$$= \frac{3.9 \times 8.85 \times 10^{-14}\,(\mathrm{F/cm}) \times 1\,\mathrm{V}}{2 \times 10^{-7}\,\mathrm{cm} \times 1.6 \times 10^{-19}\,\mathrm{C} \times 8 \times 10^{19}\,\mathrm{cm}^{-3}}$$

$$= 1.3\,\mathrm{nm}$$

EXAMPLE : Poly-Silicon Gate Depletion

(b) $$W_{dpoly} = \sqrt{\frac{2\varepsilon_s \phi_{poly}}{qN_{poly}}}$$

$$\phi_{dpoly} = qN_{poly}W_{dpoly}^2 / 2\varepsilon_s = 0.11\,\text{V}$$

(c) $$V_g = V_{fb} + \phi_{st} + V_{ox} + \phi_{poly}$$

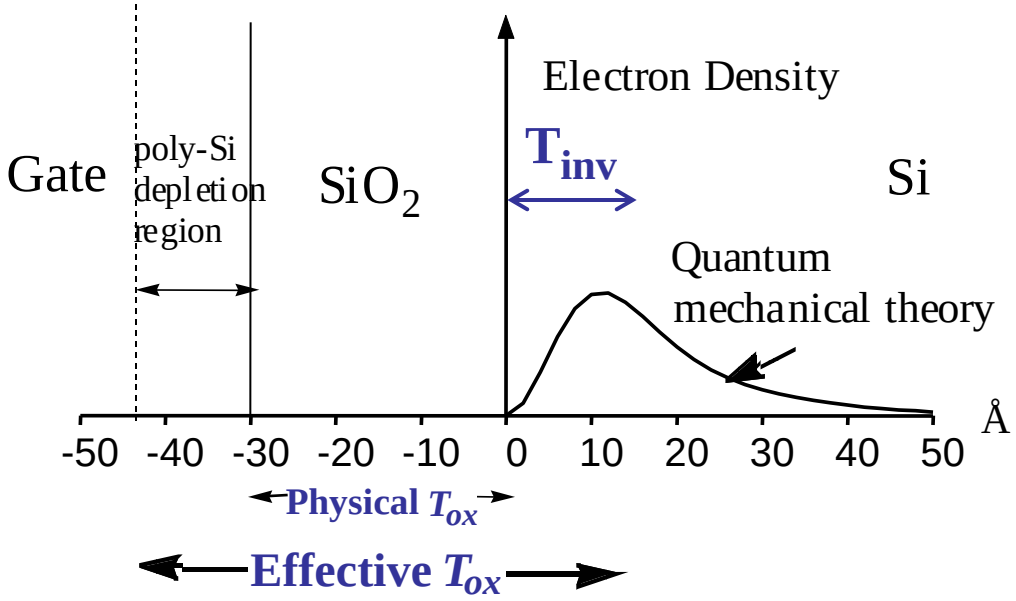
$$V_{fb} = \frac{E_g}{q} - \frac{kT}{q}\ln\left(\frac{N_c}{N_d}\right) = 1.1\,\text{V} - 0.15\,\text{V} = 0.95\,\text{V}$$

$$V_g = 0.95\,\text{V} - 0.85\,\text{V} - 1\,\text{V} - 0.11\,\text{V} = -1.01\,\text{V}$$

Is the loss of 0.11 V from the 1.01 V significant?

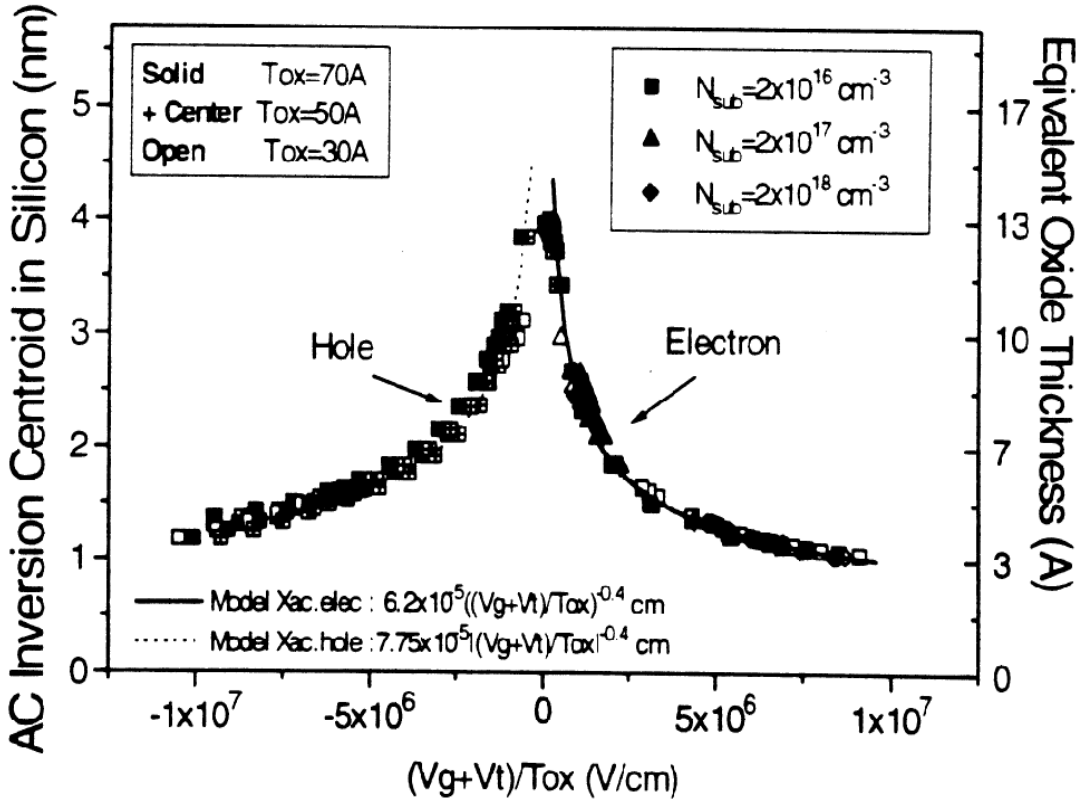
5.9 Inversion and Accumulation Charge-Layer Thickness–Quantum Mechanical Effect

Average inversion-layer location below the Si/SiO_2 interface is called the ***inversion-layer thickness,*** T_{inv} .



$n(x)$ is determined by Schrodinger's eq., Poisson eq., and Fermi function.

Electrical Oxide Thickness, T_{oxe}



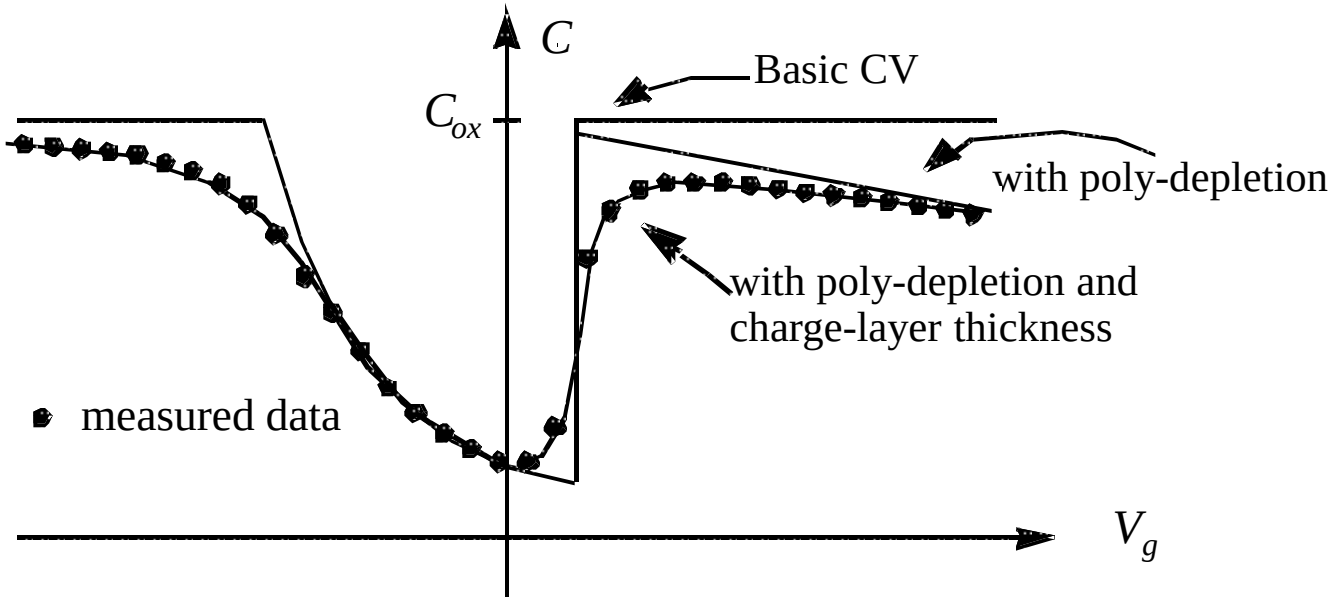
- T_{inv} is a function of the average electric field in the inversion layer, which is $(V_g + V_t)/6T_{ox}$ (Sec. 6.3.1).
- T_{inv} of holes is larger than that of electrons because of difference in effective mass.
- **T_{oxe} is the electrical oxide thickness.**

$$T_{oxe} = T_{ox} + W_{dpoly}/3 + T_{inv}/3 \quad \text{at } V_g = V_{dd}$$

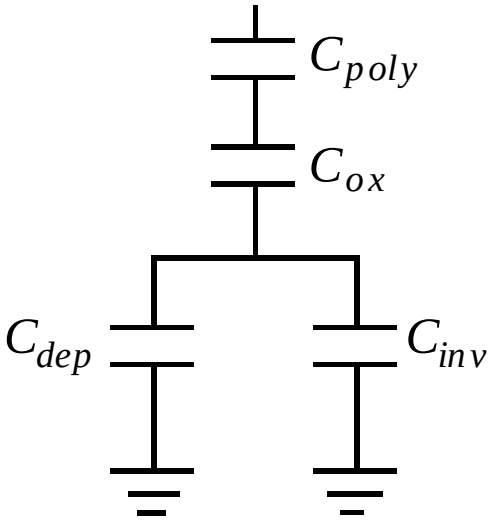
Effective Oxide Thickness and Effective Oxide Capacitance

$$Q_{inv} = C_{oxe}(V_g - V_t)$$

$$T_{oxe} = T_{ox} + W_{dpoly}/3 + T_{inv}/3$$

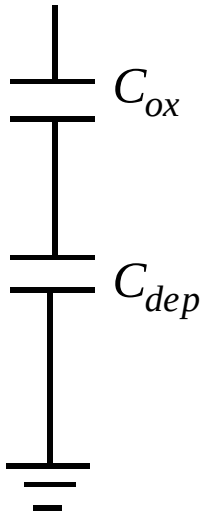


Equivalent circuit in the depletion and the inversion regimes



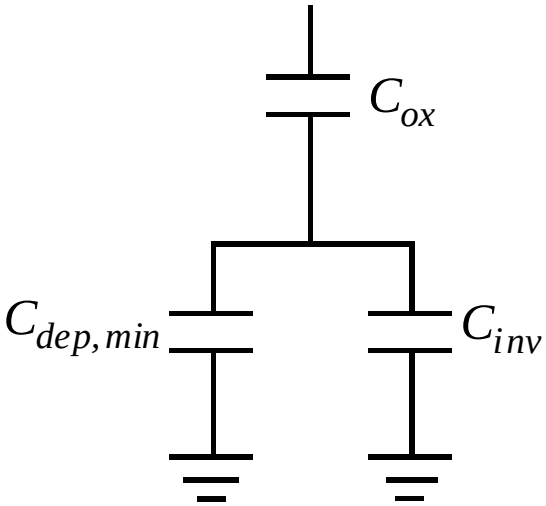
(a)

General case for both depletion and inversion regions.



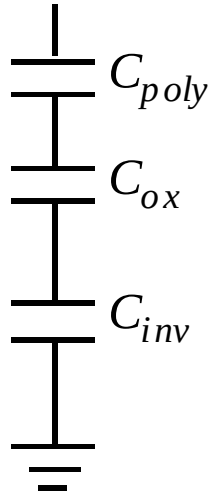
(b)

In the depletion regions



(c)

$V_g \approx V_t$

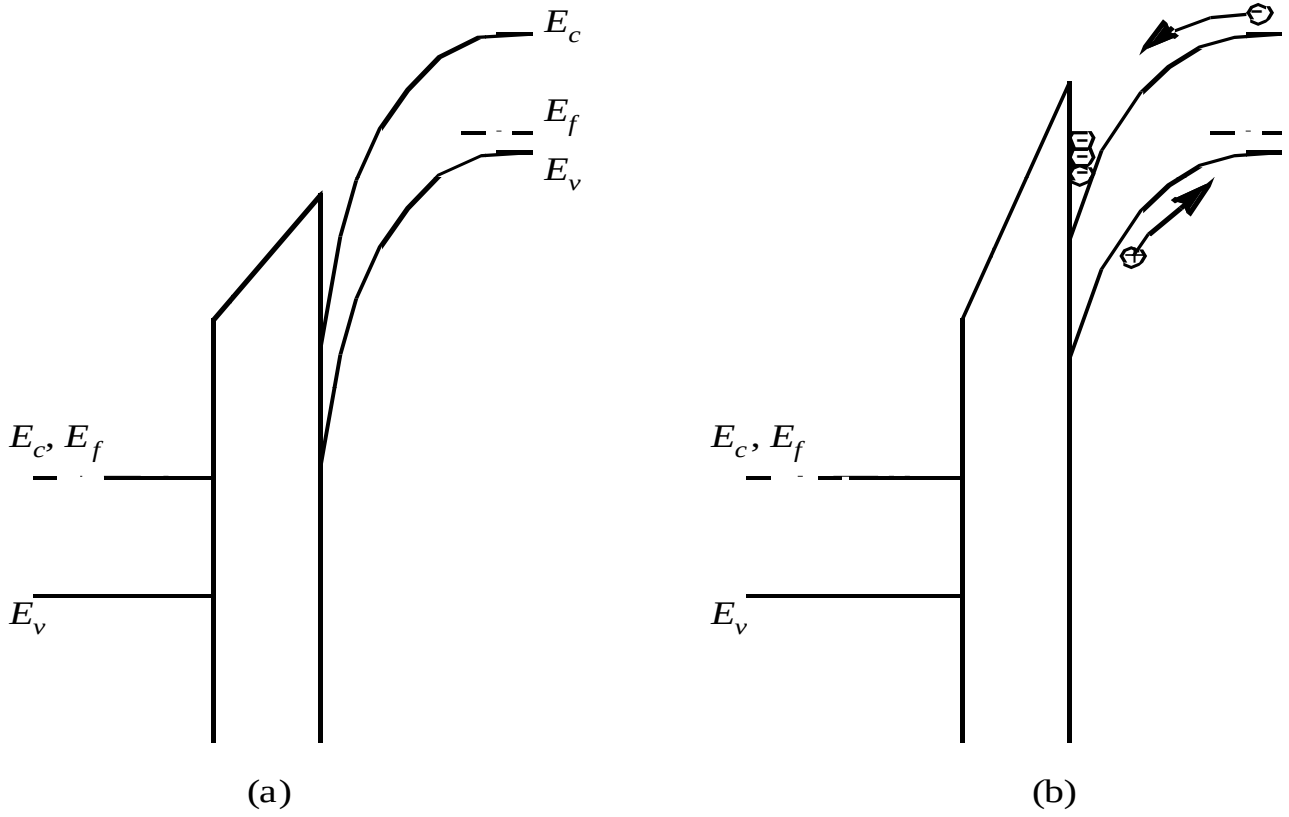


(d)

Strong inversion

5.10 CCD Imager and CMOS Imager

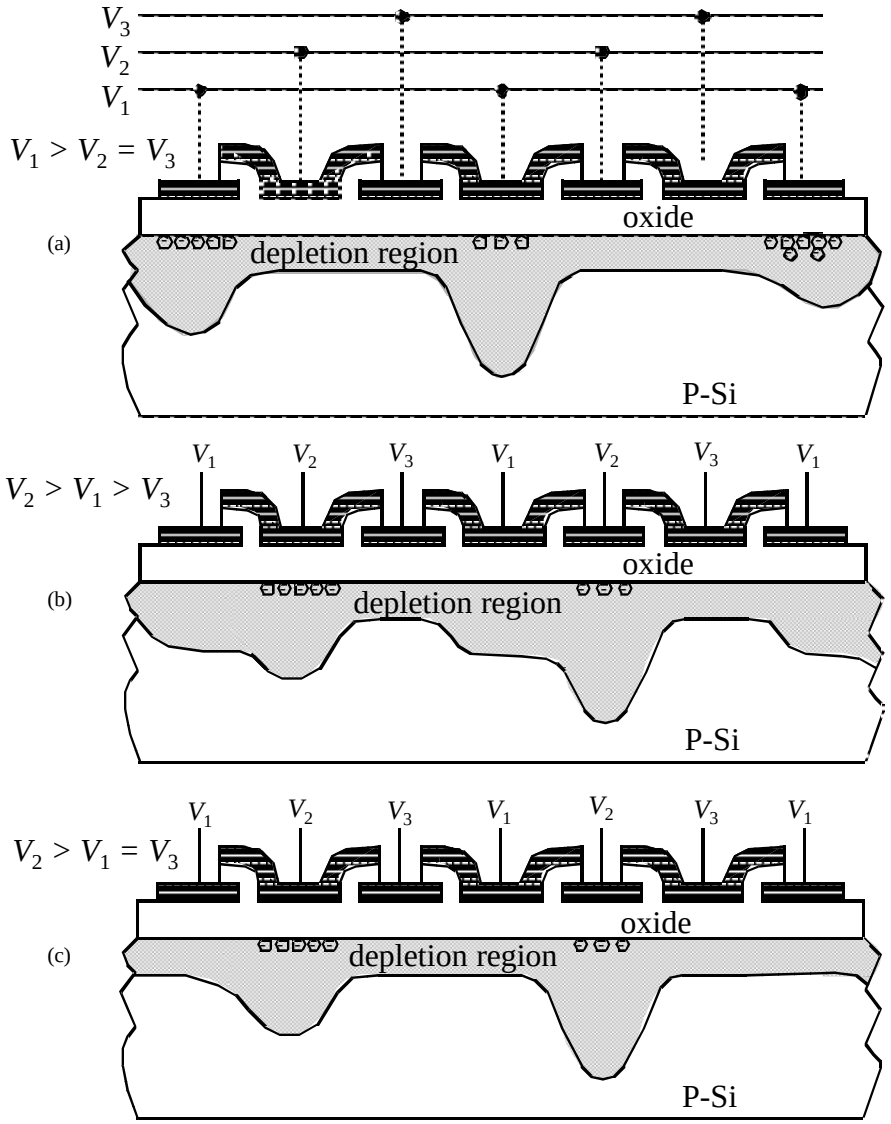
5.10.1 CCD Imager



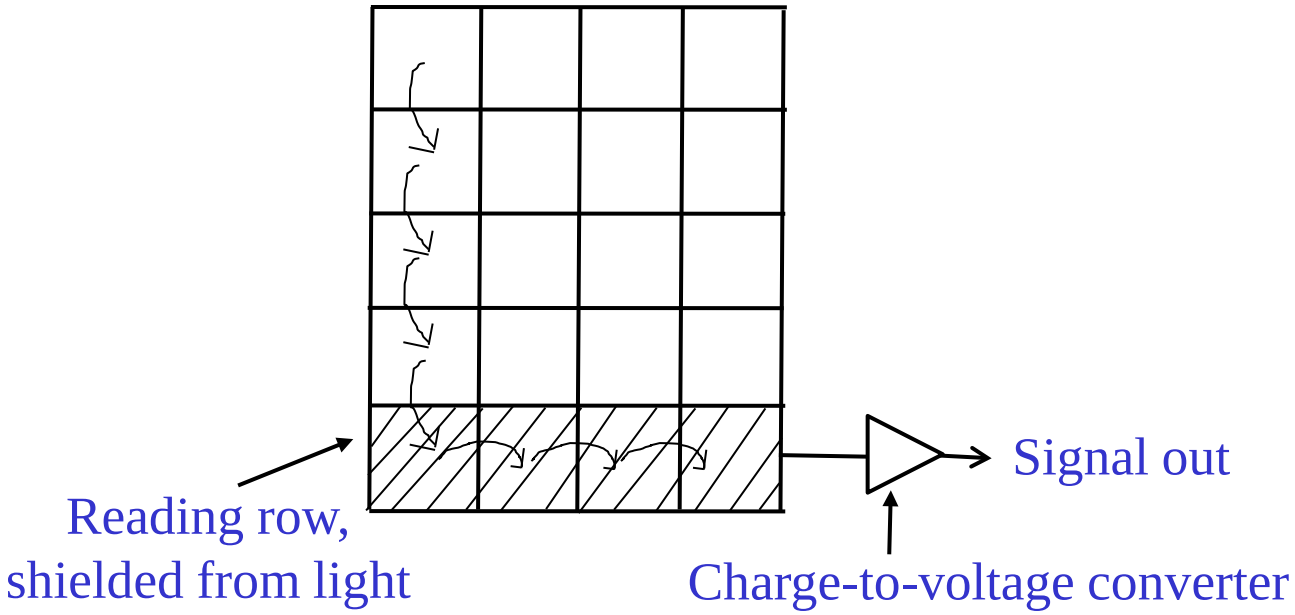
Deep depletion, Q_{inv}= 0

Exposed to light

CCD Charge Transfer

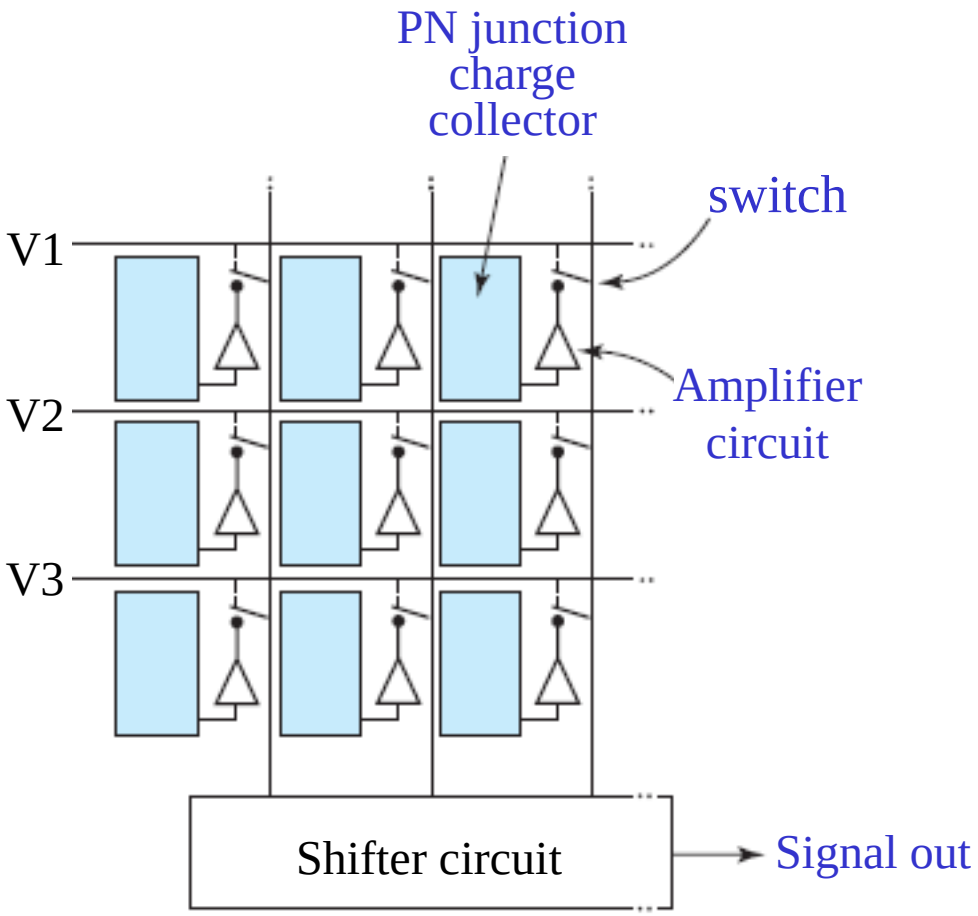


two-dimensional CCD imager



The reading row is shielded from the light by a metal film.
The 2-D charge packets are read row by row.

5.10.2 CMOS Imager



CMOS imagers can be integrated with signal processing and control circuitries to further reduce system costs. However, The size constrain of the sensing circuits forces the CMOS imager to use very simple circuits

5.11 Chapter Summary

N-type device: N^+-polysilicon gate over P-body

P-type device: P^+-polysilicon gate over N-body

$$V_{fb} = \psi_g - \psi_s + (-Q_{ox} / C_{ox})$$

$$V_g = V_{fb} + \phi_s + V_{ox} + \phi_{poly}$$
$$= V_{fb} + \phi_s - Q_s / C_{ox} + \phi_{poly}$$

5.11 Chapter Summary

$$\phi_{st} = \pm 2\phi_B \quad \text{or} \quad \pm(\phi_B + 0.45\,\text{V})$$

$$\phi_B = \frac{kT}{q} \ln \frac{N_{sub}}{n_i}$$

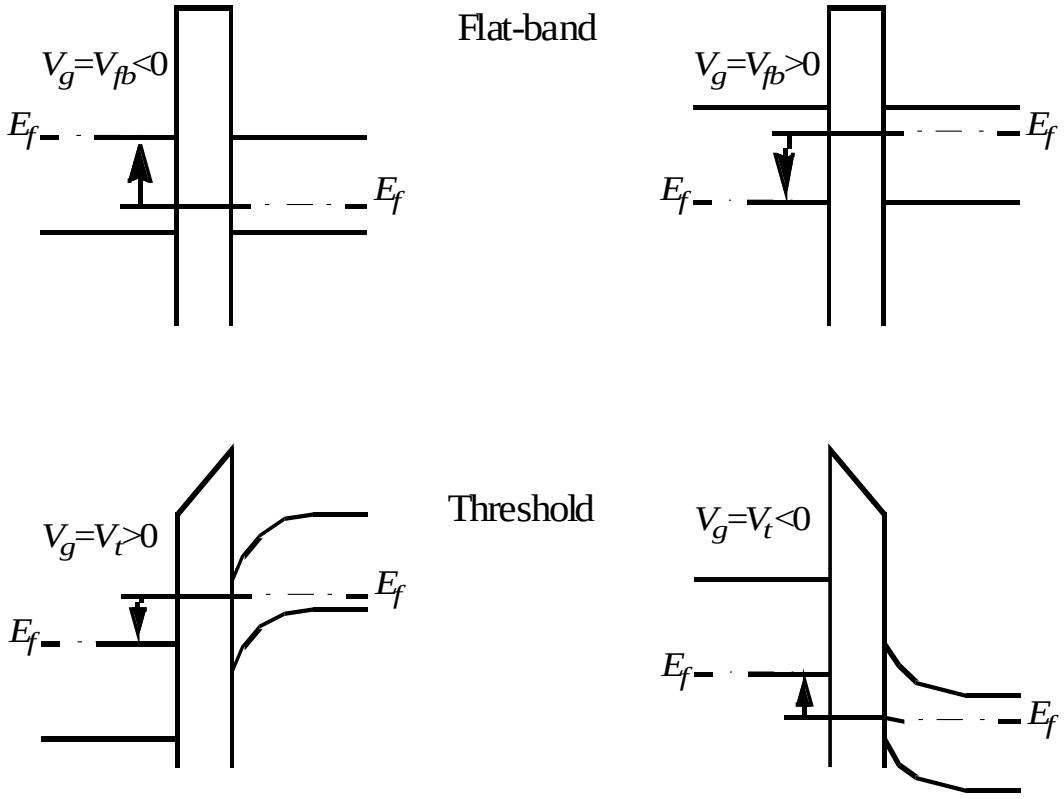
$$V_t = V_{fb} + \phi_{st} \pm \frac{\sqrt{qN_{sub}\,2\varepsilon_s\,|\phi_{st}|}}{C_{ox}}$$

+ : N-type device, – : P-type device

5.11 Chapter Summary

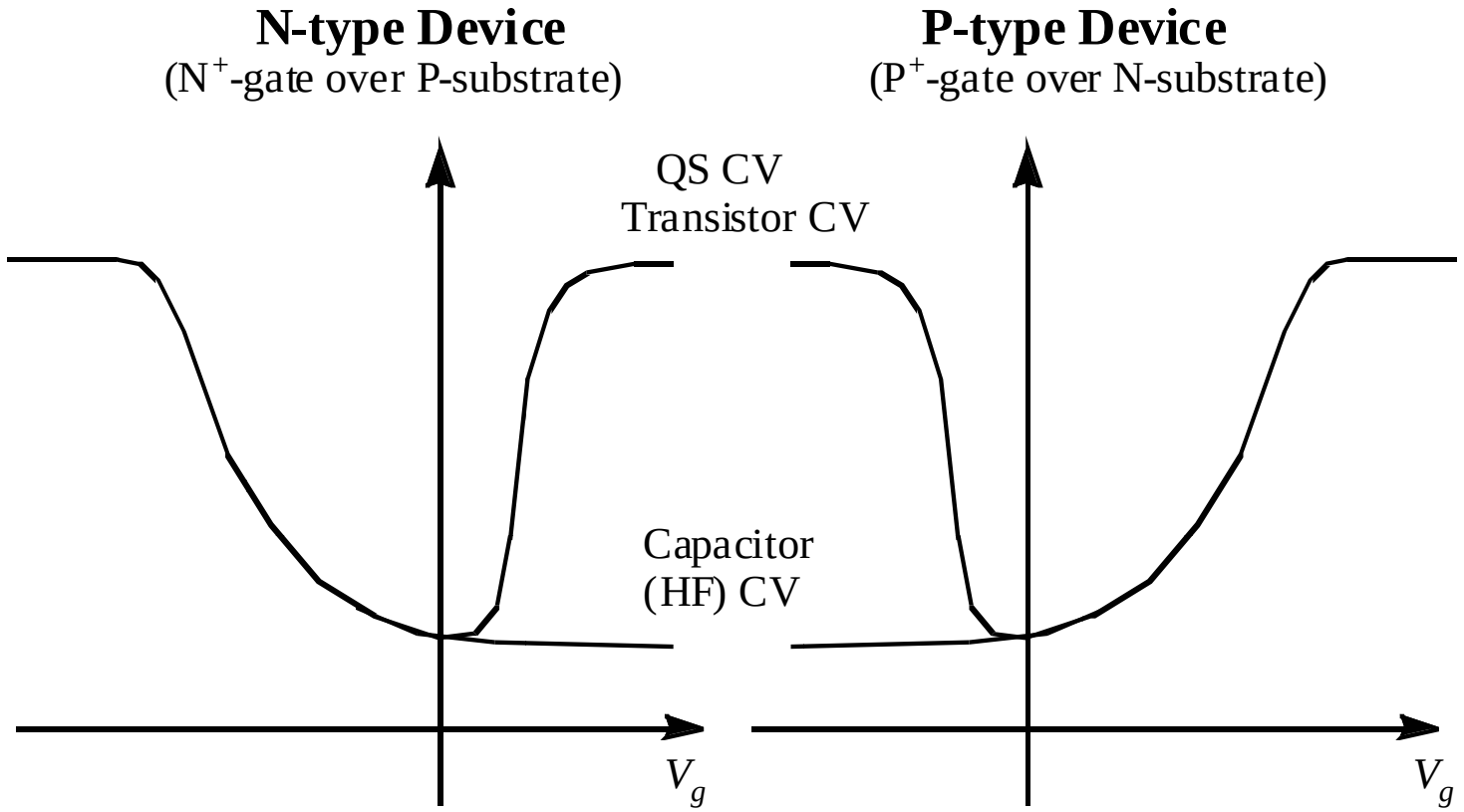
N-type Device (N^+-gate over P-substrate)

P-type Device (P^+-gate over N-substrate)



What's the diagram like at $V_g > V_t$? at $V_g = 0$?

5.11 Chapter Summary



What is the root cause of the low *C* in the HF CV branch?

Chapter 6 MOSFET

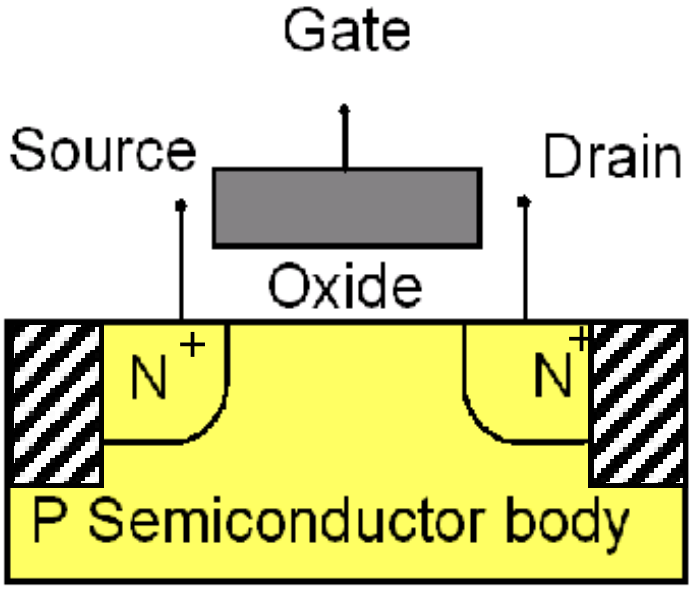
The MOSFET (MOS Field-Effect Transistor) is the building block of Gb memory chips, GHz microprocessors, analog, and RF circuits.

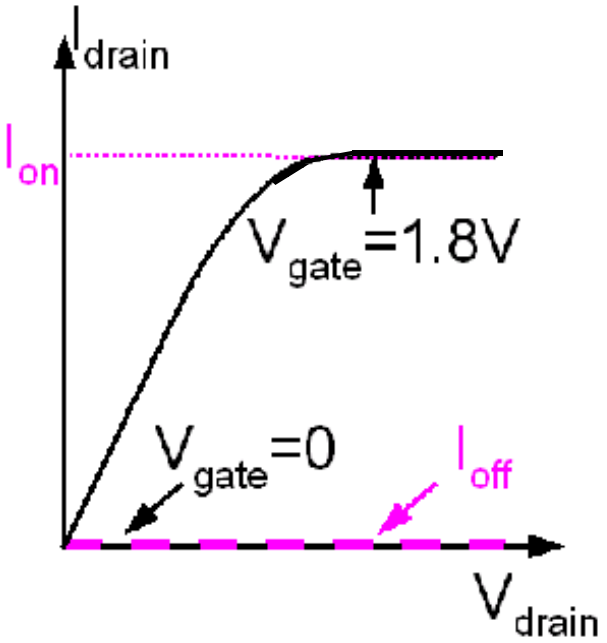
Match the following MOSFET characteristics with their applications:

- small size
- high speed
- low power
- high gain

6.1 Introduction to the MOSFET

Basic MOSFET structure and IV characteristics



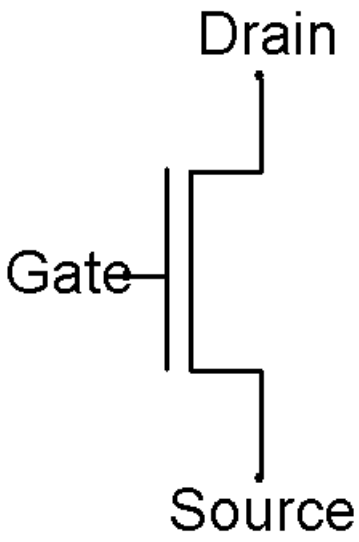


What is desirable: large I_{on}, small I_{off}

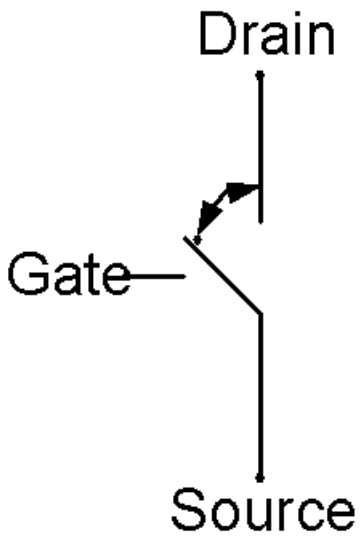
6.1 Introduction to the MOSFET

Two ways of representing a MOSFET:

Circuit Symbol



Simple Switch



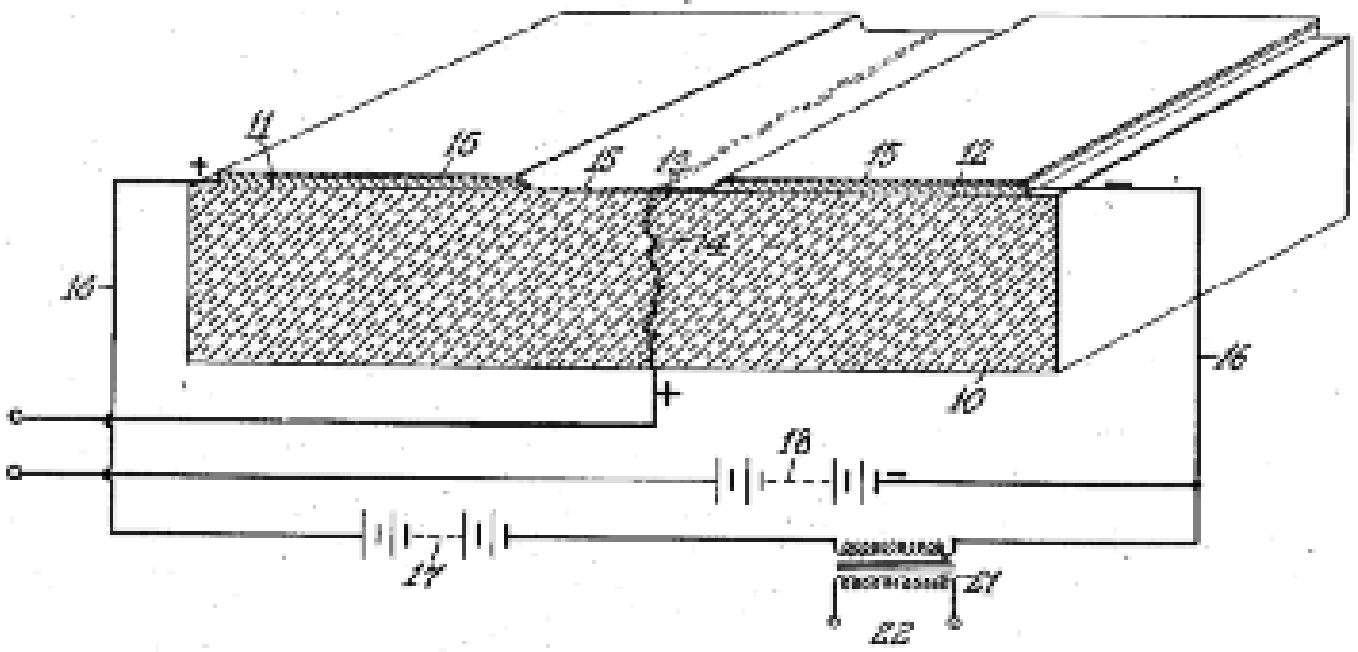
Early Patents on the Field-Effect Transistor

Jan. 28, 1930. J. E. LILIENFELD 1,745,175

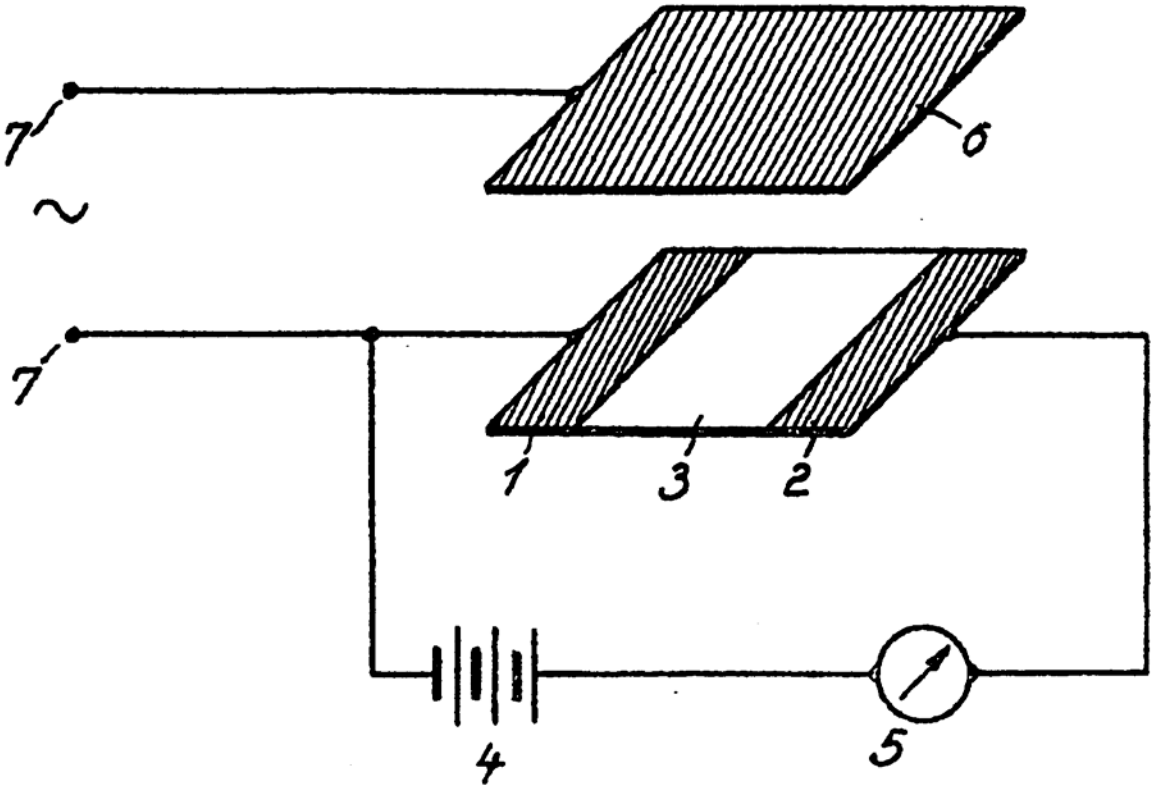
METHOD AND APPARATUS FOR CONTROLLING ELECTRIC CURRENTS

Filed Oct. 8, 1926

Fig. 1.



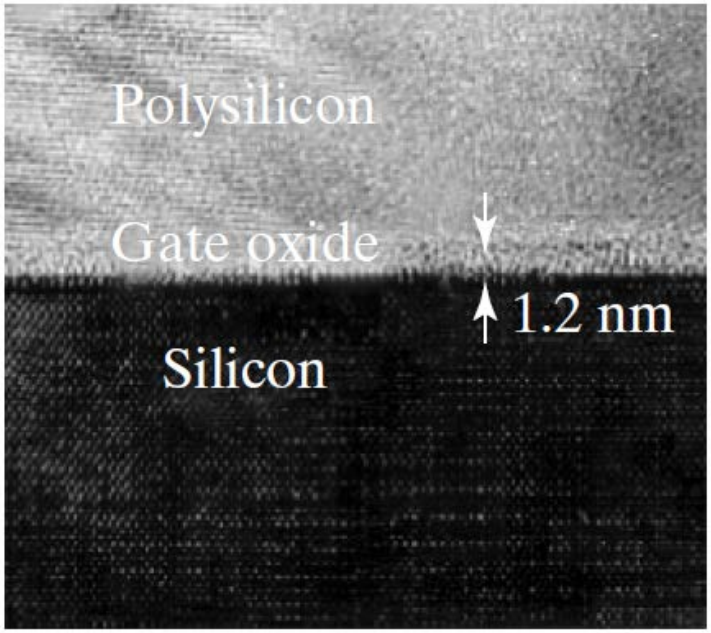
Early Patents on the Field-Effect Transistor



In 1935, a British patent was issued to Oskar Heil.
A working MOSFET was not demonstrated until 1955.
Using today's terminology, what are 1, 2, and 6?

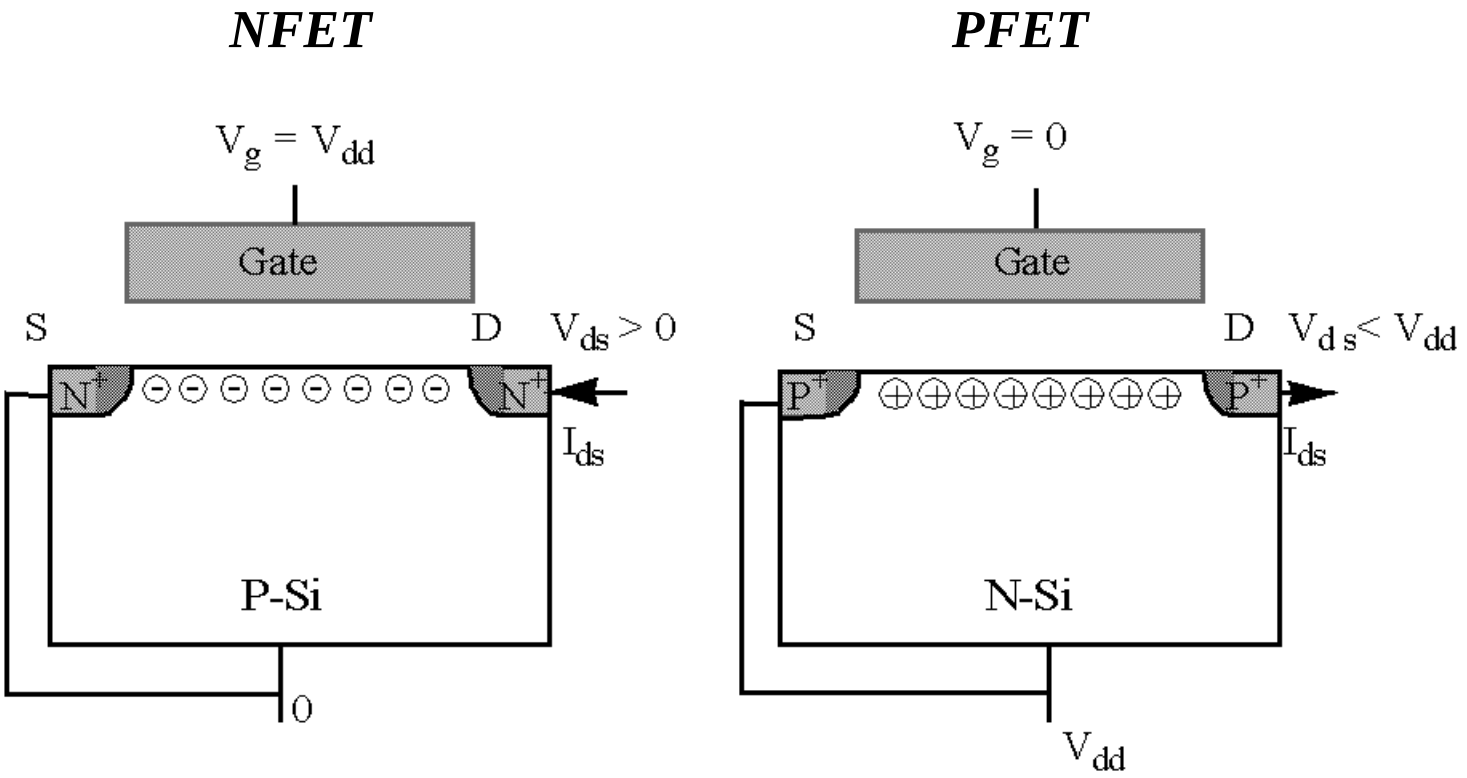
6.2 MOSFETs Technology

Polysilicon gate and 1.2nm SiO_2



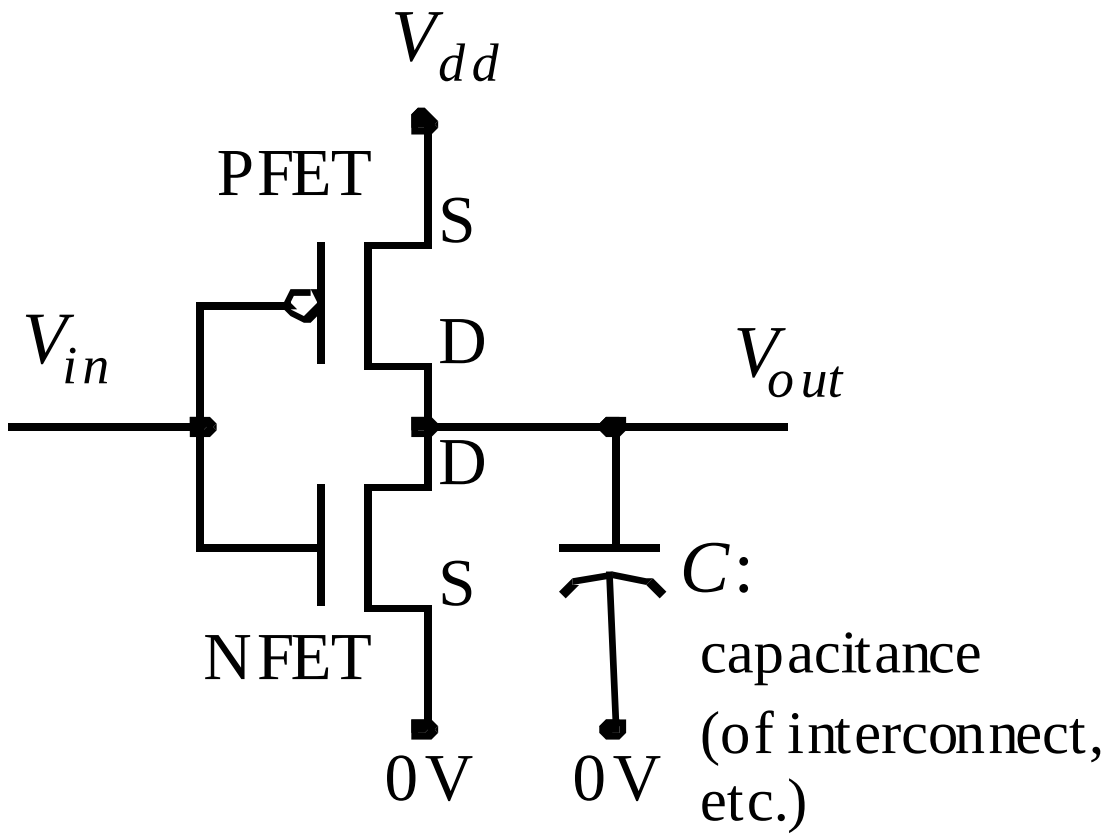
- 1.2 nm SiO_2 used in production. Leakage current through the oxide limits further thickness reduction.

6.2 Complementary MOSFETs Technology



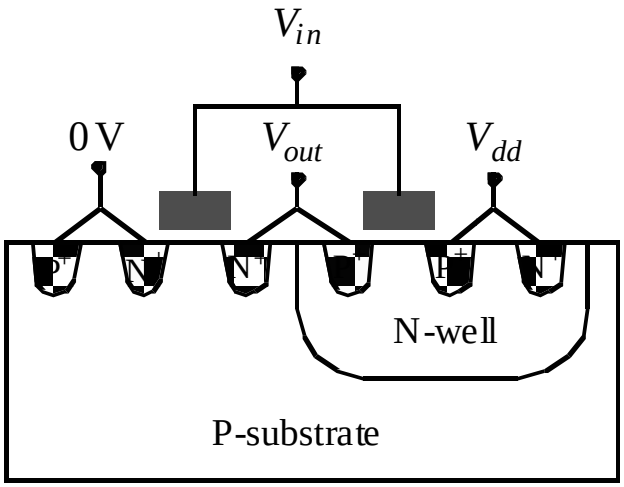
When $V_g = V_{dd}$, the NFET is on and the PFET is off.
When $V_g = 0$, the PFET is on and the NFET is off.

CMOS (Complementary MOS) Inverter

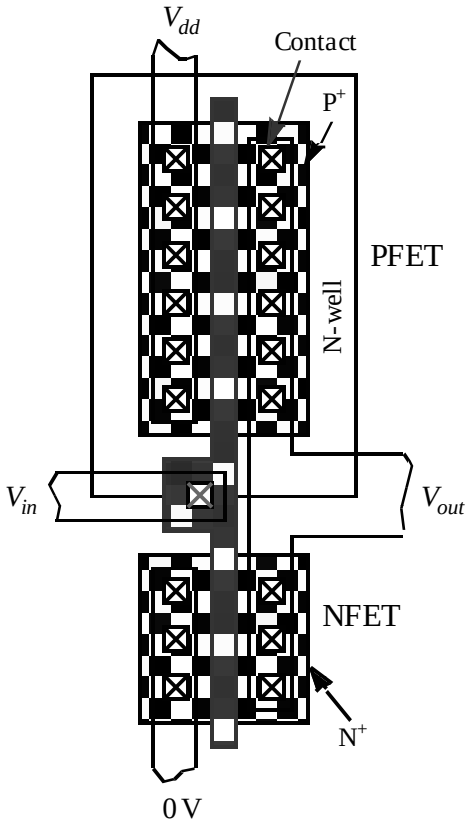


A CMOS inverter is made of a PFET ***pull-up device*** and a NFET ***pull-down device***. V_{out} = ? if V_{in} = 0 V.

CMOS (Complementary MOS) Inverter



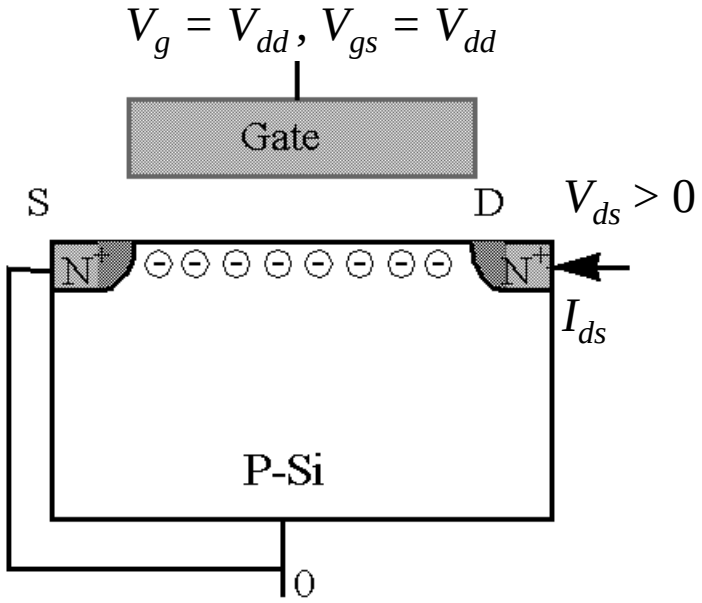
• NFET and PFET can be fabricated on the same chip.



• basic layout of a CMOS inverter

6.3 Surface Mobilities and High-Mobility FETs

6.3.1 Surface Mobilities



How to measure the surface mobility:

$$I_{ds} = W \times Q_{inv} \times v = W Q_{inv} \mu_{ns} \mathcal{E} = W Q_{inv} \mu_{ns} V_{ds} / L$$

$$= W C_{oxe} (V_{gs} - V_t) \mu_{ns} V_{ds} / L$$

Mobility is a function of the average of the fields at the bottom and the top of the inversion charge layer, $\mathcal{E}_b$ and $\mathcal{E}_t$.

From Gauss's Law,

$$\mathcal{E}_b = -Q_{dep}/\varepsilon_s$$

$$V_t = V_{fb} + \phi_{st} - Q_{dep}/C_{oxe}$$

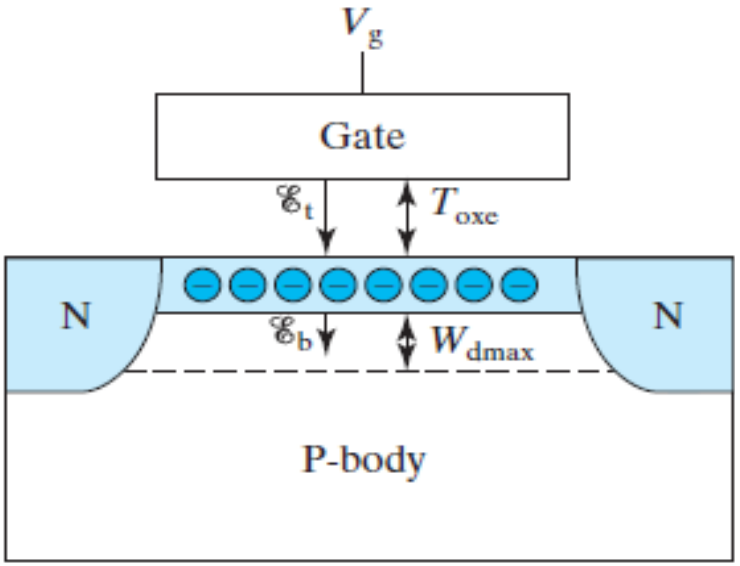
Therefore,

$$\mathcal{E}_b = \frac{C_{oxe}}{\varepsilon_s}(V_t - V_{fb} - \phi_{st})$$

$$\mathcal{E}_t = -(Q_{dep} + Q_{inv})/\varepsilon_s$$

$$= \mathcal{E}_b - Q_{inv}/\varepsilon_s = \mathcal{E}_b + \frac{C_{oxe}}{\varepsilon_s}(V_{gs} - V_t)$$

$$= \frac{C_{oxe}}{\varepsilon_s}(V_{gs} - V_{fb} - \phi_{st})$$

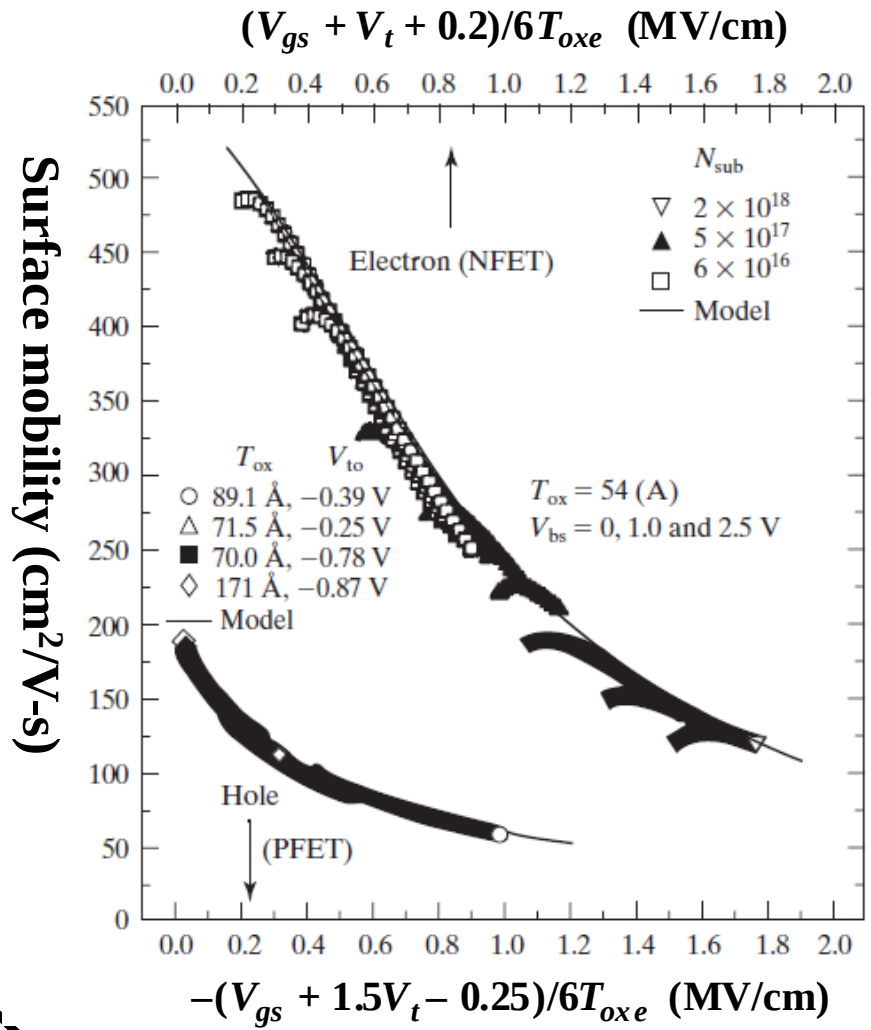


$$\therefore \frac{1}{2}(\mathcal{E}_b + \mathcal{E}_t) = \frac{C_{oxe}}{2\varepsilon_s}(V_{gs} + V_t - 2V_{fb} - 2\phi_{st})$$

$$\approx \frac{C_{oxe}}{2\varepsilon_s}(V_{gs} + V_t + 0.2\,\mathrm{V})$$

$$= \frac{V_{gs} + V_t + 0.2\,\mathrm{V}}{6T_{oxe}}$$

Universal Surface Mobilities



- Surface roughness scattering is stronger (mobility is lower) at higher V_g, higher V_t, and thinner T_{oxe}.

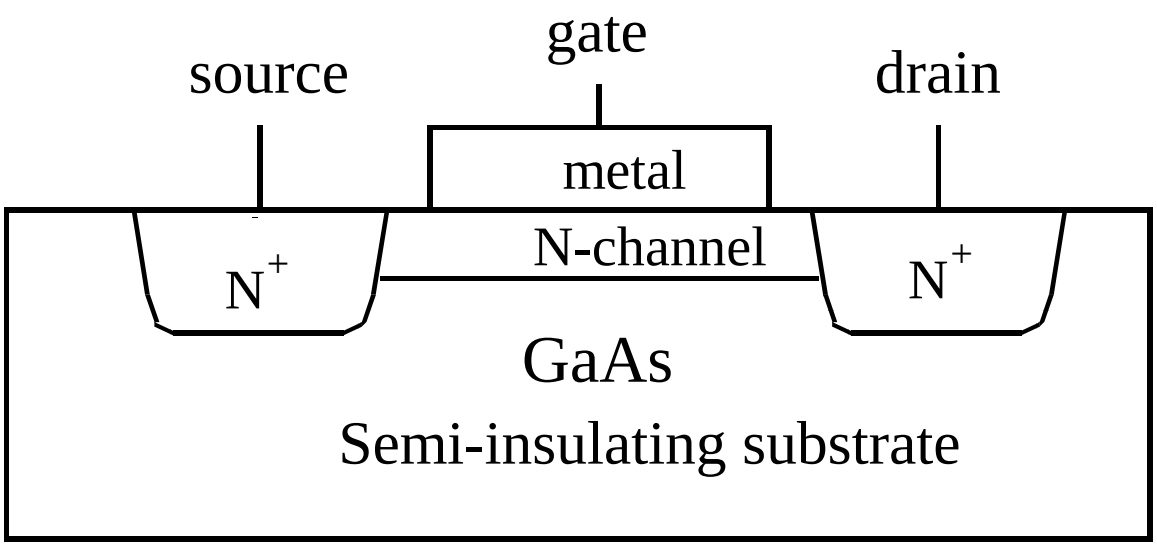
EXAMPLE: *What is the surface mobility at V_{gs}=1 V in an N-channel MOSFET with V_t=0.3 V and T_{oxe}=2 nm?*

Solution:

$$(V_{gs} + V_t + 0.2)/6T_{oxe}$$
$$= 1.5\,\text{V}/12\times10^{-7}\,\text{cm}$$
$$= 1.25\,\text{MV/cm}$$

1 MV is a megavolt (10^6 V). From the mobility figure, μ_{ns}=190 cm2/Vs, which is several times smaller than the bulk mobility.

6.3.2 GaAs MESFET

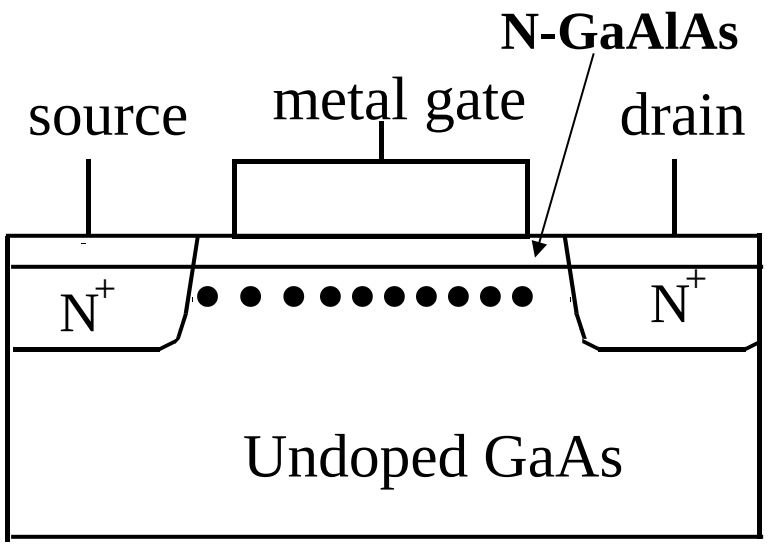


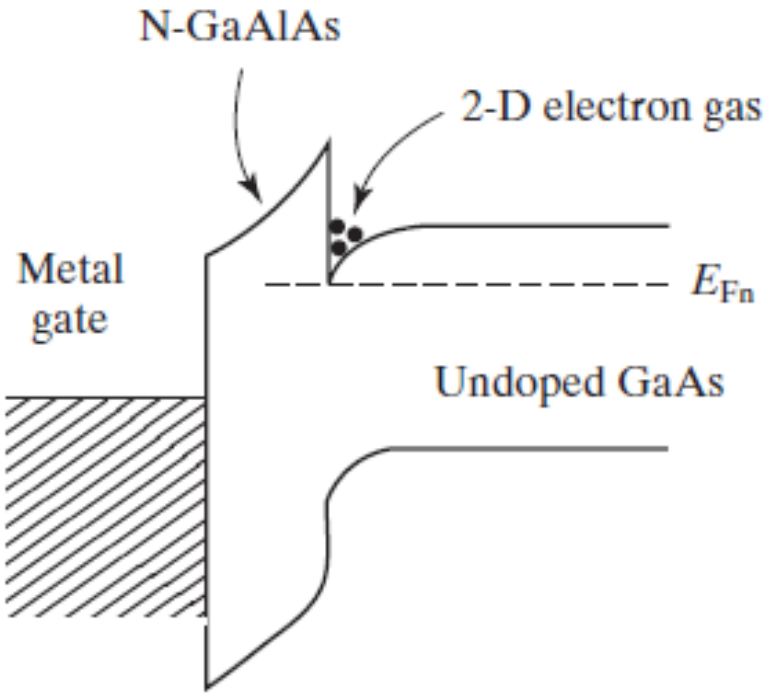
MESFET IV characteristics are similar to MOSFET's but does not require a gate oxide.

Question: What is the advantage of GaAs FET over Si FET?

Terms: ***depletion-mode transistor, enhancement-mode transistor***

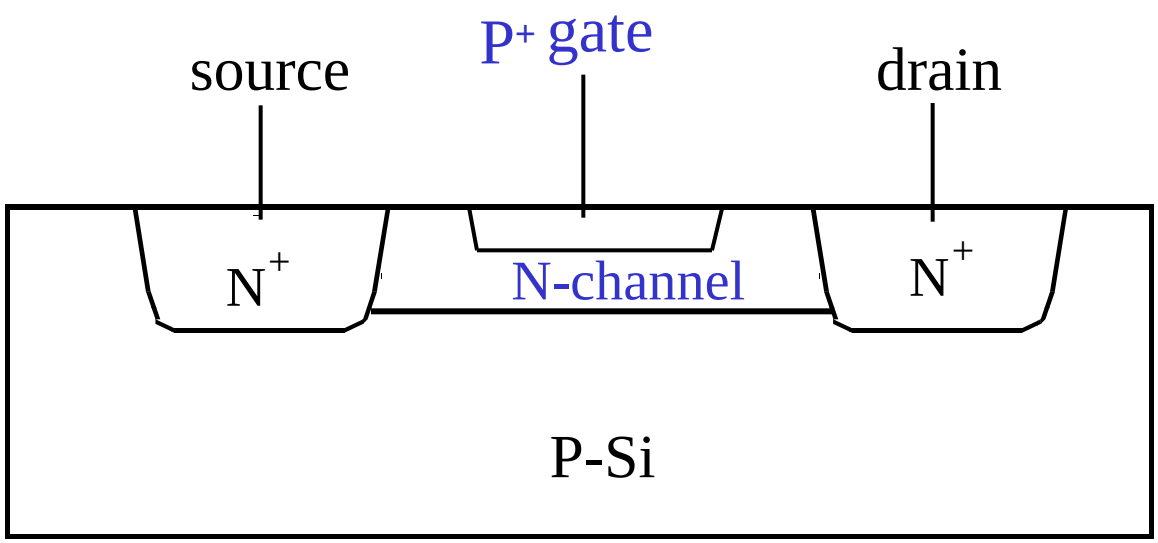
6.3.3 HEMT, High Electron Mobility Transistor





- A large-Eg semiconductor serves as the "gate dielectric".
- The layer of electrons is called **2D-electron-gas**, the equivalent of the inversion or accumulation layer of a MOSFET.

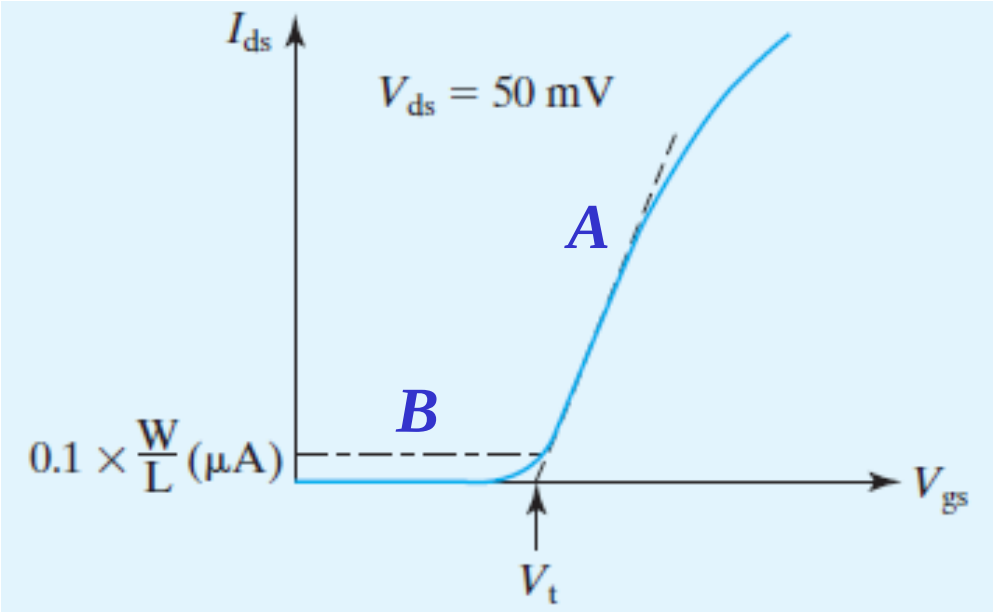
6.3.4 JFET



- The gate is a P^+N junction.
- The FET is a **junction field-effect transistor (JFET)**.

6.4 V_t and Body Effect

How to Measure the V_t of a MOSFET



- *Method A*. V_t is measured by extrapolating the I_{ds} versus V_{gs} curve to $I_{ds} = 0$.

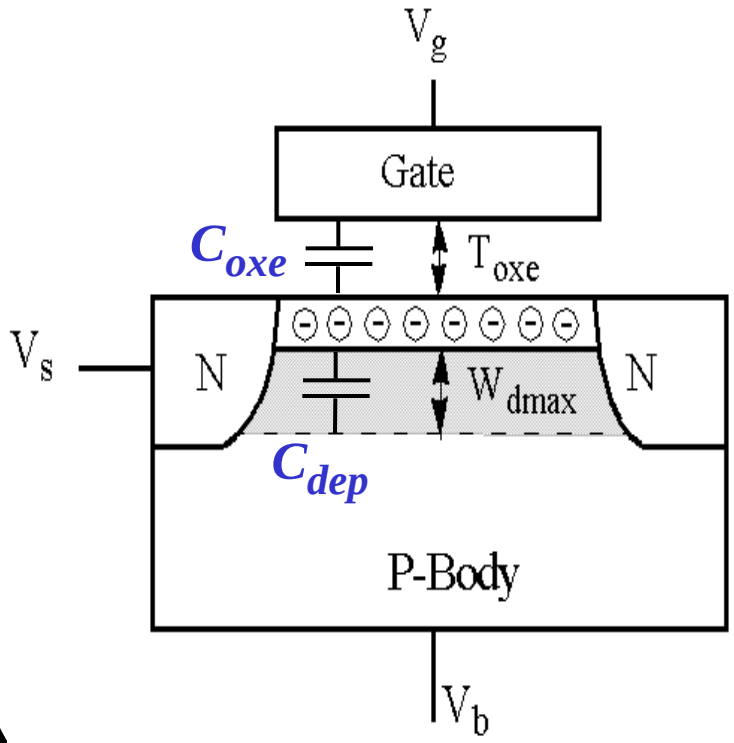
$$I_{dsat} = \frac{W}{L} C_{oxe} (V_{gs} - V_t) \mu_{ns} V_{ds} \propto V_{gs} - V_t$$

- *Method B*. The Vg at which $Ids = 0.1\mu A \times W/L$

MOSFET V_t and the Body Effect

- **Two capacitors => two charge components**

$$C_{dep} = \frac{\varepsilon_s}{W_{d\max}}$$



$$Q_{inv} = -C_{oxe}(V_{gs} - V_t) + C_{dep}V_{sb}$$

$$= -C_{oxe}(V_{gs} - (V_t + \frac{C_{dep}}{C_{oxe}}V_{sb}))$$

- Redefine V_t as

$$V_t(V_{sb}) = V_{t0} + \frac{C_{dep}}{C_{oxe}}V_{sb} = V_{t0} + \alpha V_{sb}$$

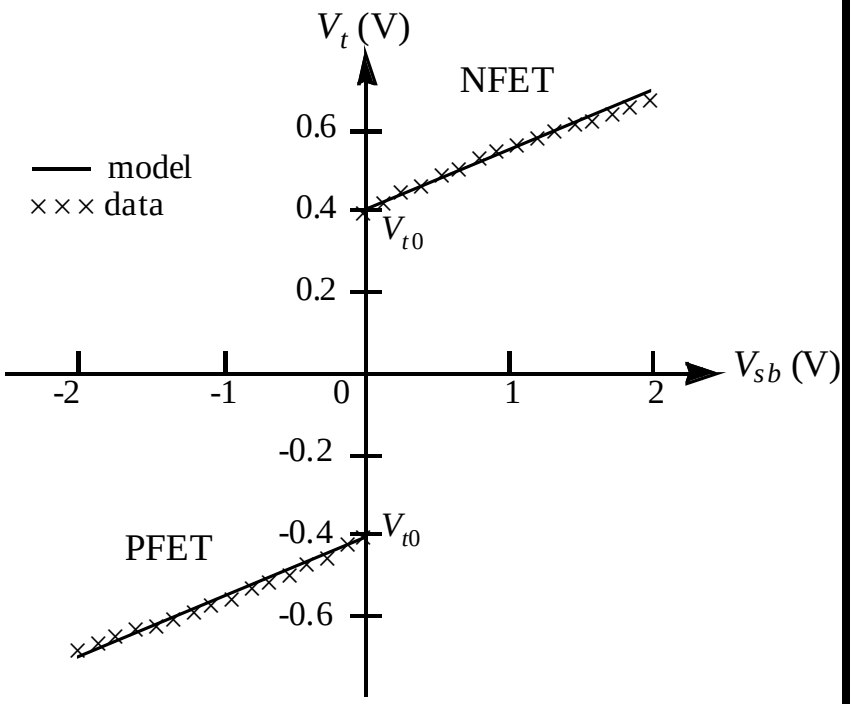
MOSFET V_t and the Body Effect

• ***Body effect***: V_t is a function of V_{sb}. When the source-body junction is reverse-biased, $|V_t|$ increases.

• ***Body effect coefficient***:

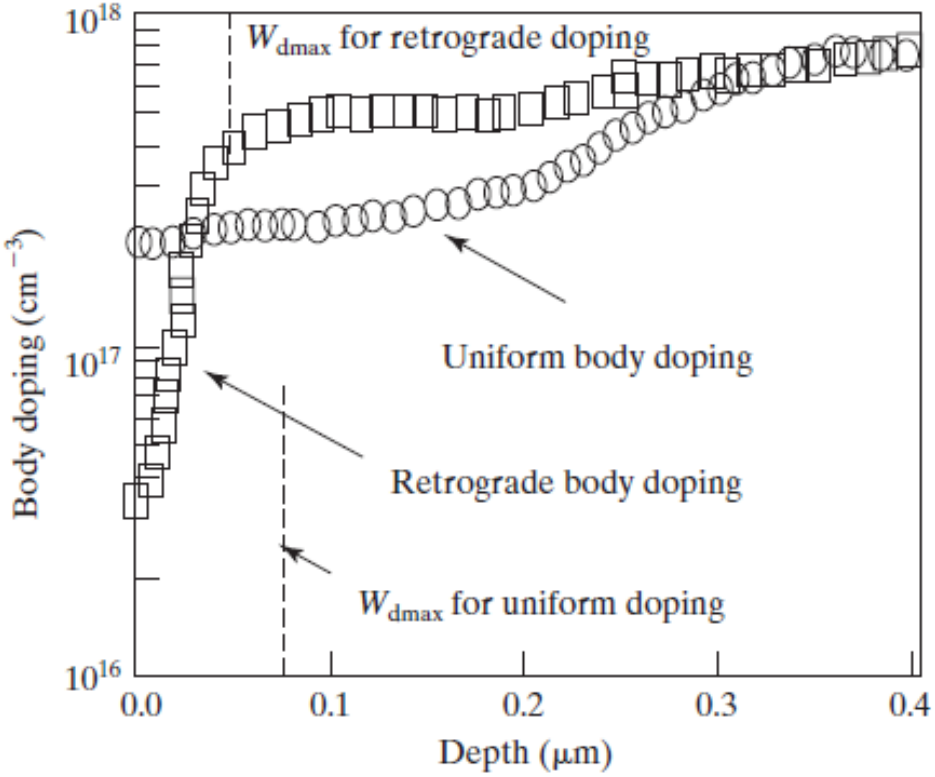
$$V_t = V_{t0} + \alpha V_{sb}$$

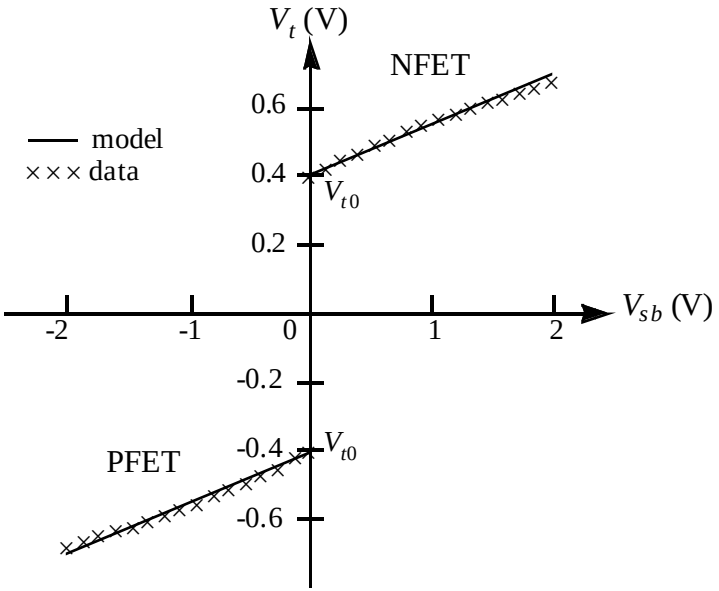
$$\alpha = C_{dep}/C_{oxe}$$
$$= 3T_{oxe} / W_{dep}$$



Body effect slows down circuits? How can it be reduced?

Retrograde Body Doping Profiles





- W_{dep} does not vary with V_{sb}.
- Retrograde doping is popular because it reduces off-state leakage and allows higher surface mobility.

Uniform Body Doping

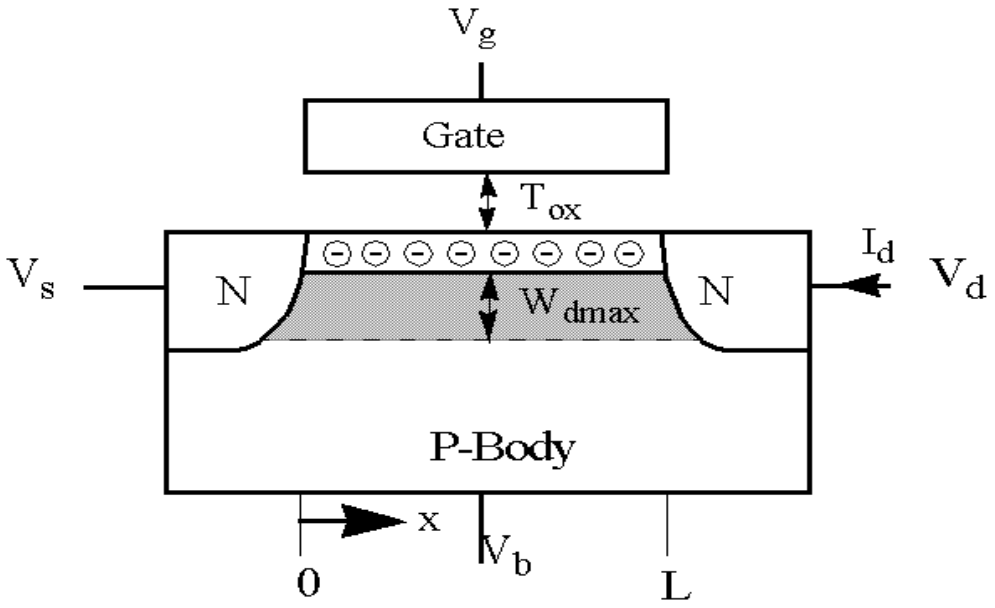
When the source/body junction is reverse-biased, there are two quasi-Fermi levels (E_{fn} and E_{fp}) which are separated by qV_{sb}. An NMOSFET reaches threshold of inversion when E_c is close to E_{fn}, not E_{fp}. This requires the band-bending to be $2\phi_B + V_{sb}$, not $2\phi_B$.

$$V_t = V_{t0} + \frac{\sqrt{qN_a 2\varepsilon_s}}{C_{oxe}}\left(\sqrt{2\phi_B + V_{sb}} - \sqrt{2\phi_B}\right)$$

$$\equiv V_{t0} + \gamma\left(\sqrt{2\phi_B + V_{sb}} - \sqrt{2\phi_B}\right)$$

γ is the ***body-effect parameter***.

6.5 Q_{inv} in MOSFET



- Channel voltage
  $V_c = V_s$ at $x = 0$ and
  $V_c = V_d$ at $x = L$.

- $Q_{inv} = -C_{oxe}(V_{gs} - V_{cs} - V_{t0} - \alpha\,(V_{sb} + V_{cs})$
  $= -C_{oxe}(V_{gs} - V_{cs} - (V_{t0} + \alpha\, V_{sb}) - \alpha\, V_{cs})$
  $= -C_{oxe}(V_{gs} - mV_{cs} - V_t)$

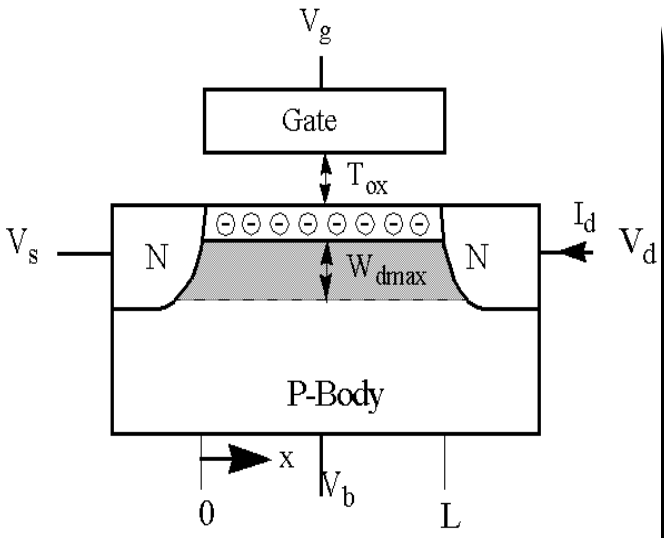
- $m \equiv 1 + \alpha = 1 + 3T_{oxe}/W_{dmax}$
  m is called the ***body-effect factor*** or ***bulk-charge factor***

6.6 Basic MOSFET IV Model

$$I_{ds} = WQ_{inv}v = WQ_{inv}\mu_{ns}\mathbf{E}$$
$$= WC_{oxe}(V_{gs} - mV_{cs} - V_t)\mu_{ns}dV_{cs}/dx$$

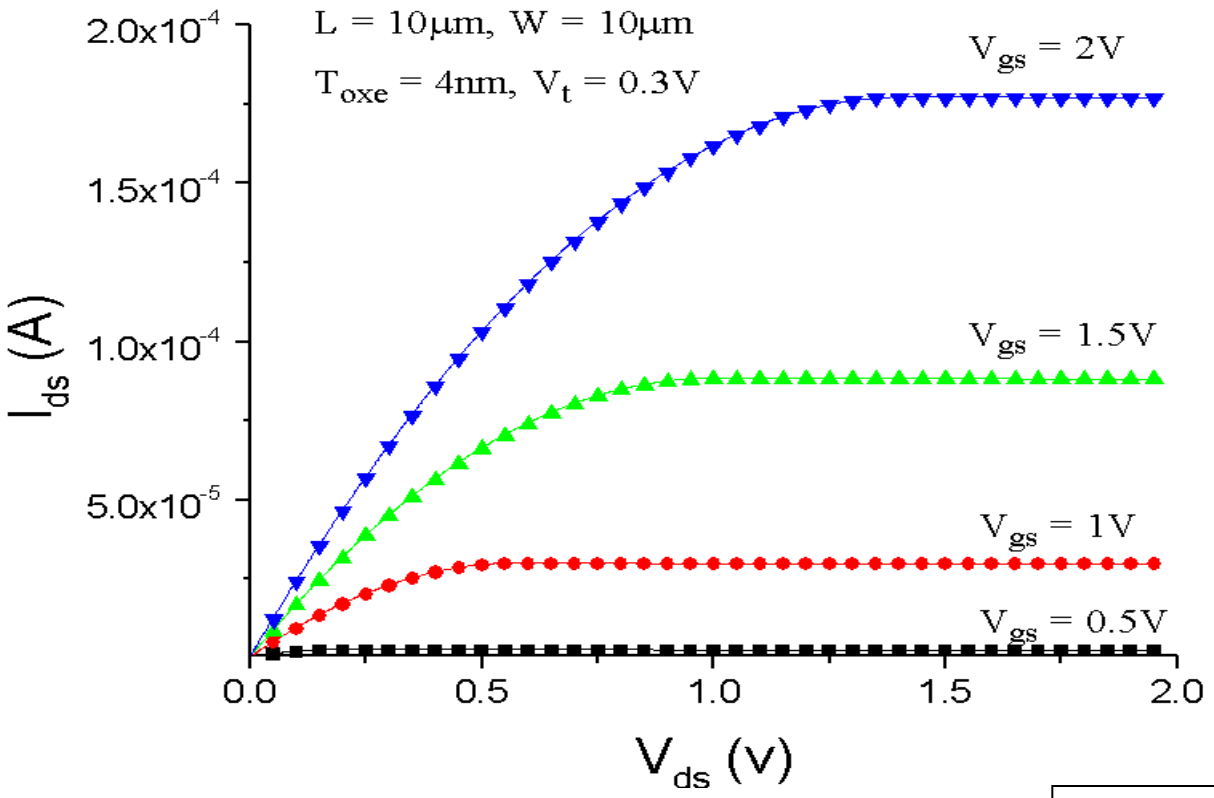
$$\int_0^L I_{ds}dx = WC_{oxe}\mu_{ns}\int_0^{V_{ds}}(V_{gs} - mV_{cs} - V_t)dV_{cs}$$

$$I_{ds}L = WC_{oxe}\mu_{ns}(V_{gs} - V_t - mV_{ds}/2)V_{ds}$$

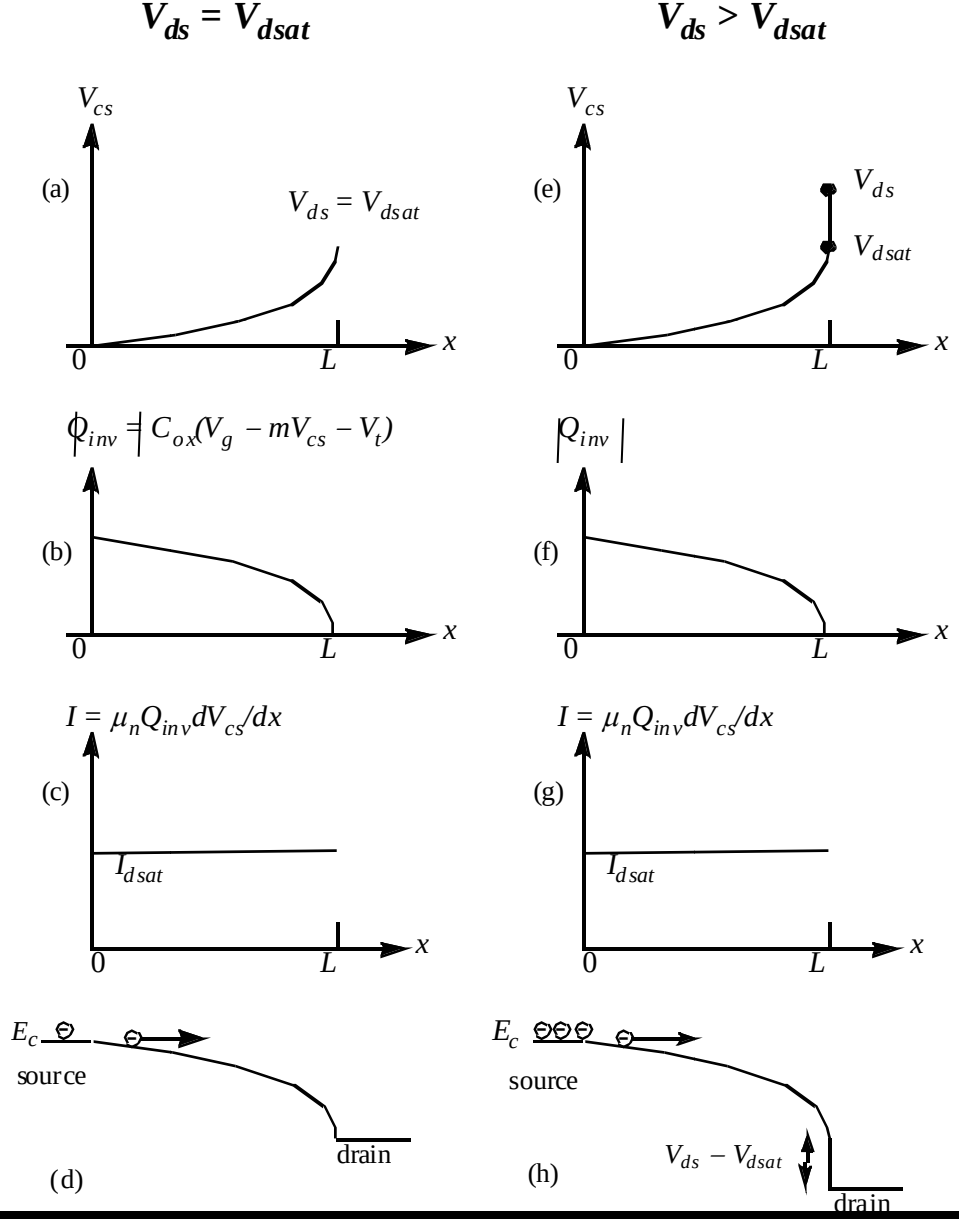


$$I_{ds} = \frac{W}{L}C_{oxe}\mu_s(V_{gs} - V_t - \frac{m}{2}V_{ds})V_{ds}$$

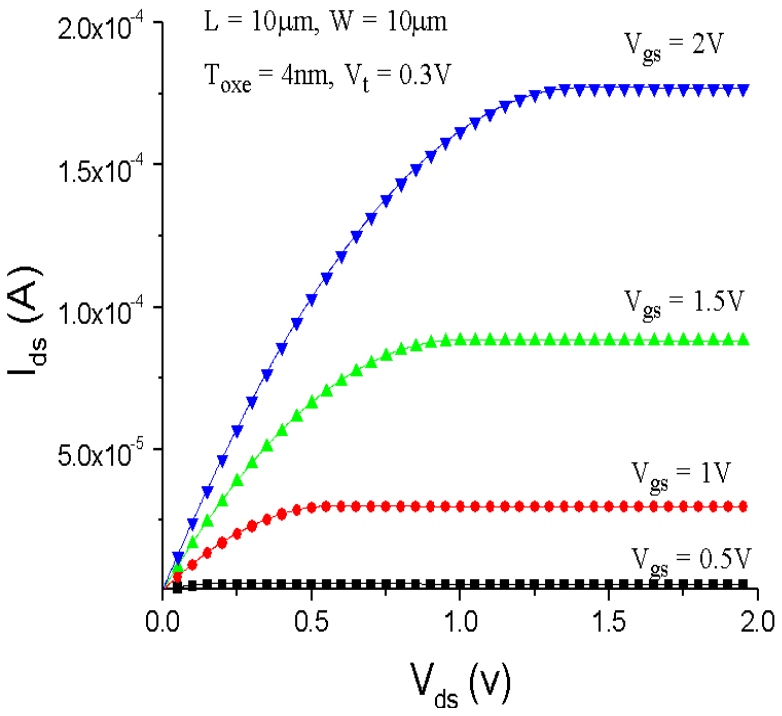
V_{dsat} : Drain Saturation Voltage



$$\frac{dI_{ds}}{dV_{ds}} = 0 = \frac{W}{L} C_{oxe} \mu_{ns} (V_{gs} - V_t - mV_{ds}) \Rightarrow \boxed{V_{dsat} = \frac{V_{gs} - V_t}{m}}$$


$V_{ds} = V_{dsat}$
$V_{ds} > V_{dsat}$
V_{cs}
(a)
$V_{ds} = V_{dsat}$
0
L
x
V_{cs}
(e)
V_{ds}
V_{dsat}
0
L
x
$|Q_{inv}| = C_{ox}(V_g - mV_{cs} - V_t)$
(b)
0
L
x
$|Q_{inv}|$
(f)
0
L
x
$I = \mu_n Q_{inv} dV_{cs}/dx$
(c)
I_{dsat}
0
L
x
$I = \mu_n Q_{inv} dV_{cs}/dx$
(g)
I_{dsat}
0
L
x
E_c
source
drain
(d)
E_c
source
$V_{ds} - V_{dsat}$
drain
(h)

Saturation Current and Transconductance



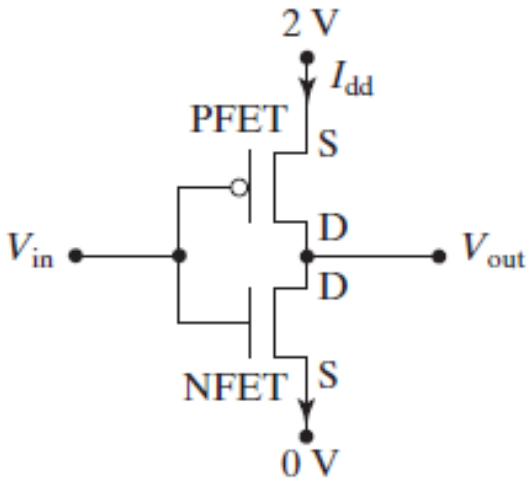
- linear region, saturation region

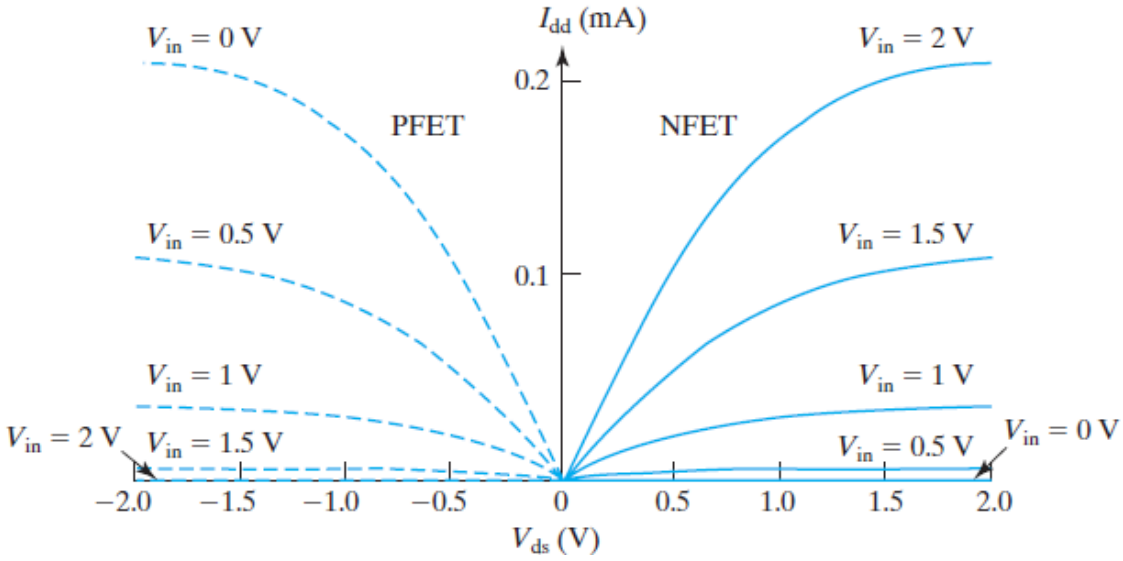
$$I_{dsat} = \frac{W}{2mL} C_{oxe} \mu_{ns} (V_{gs} - V_t)^2$$

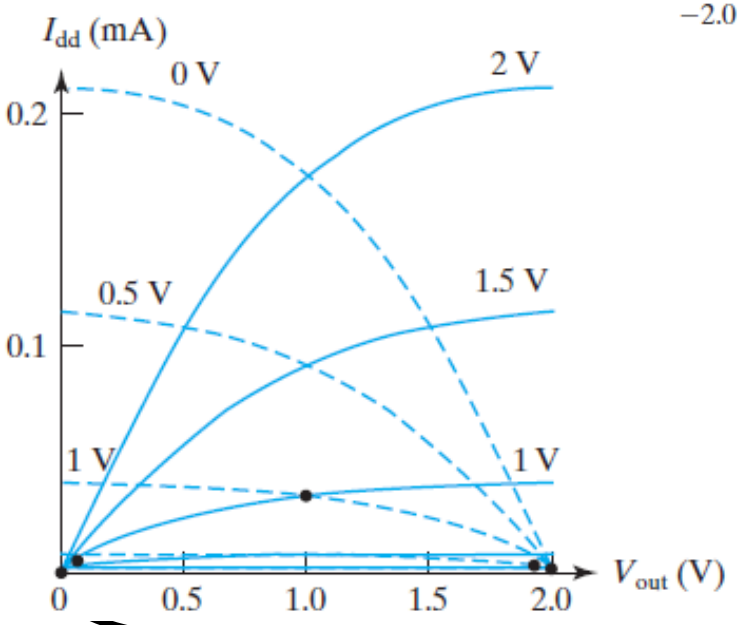
- transconductance: $g_m = dI_{ds}/dV_{gs}$

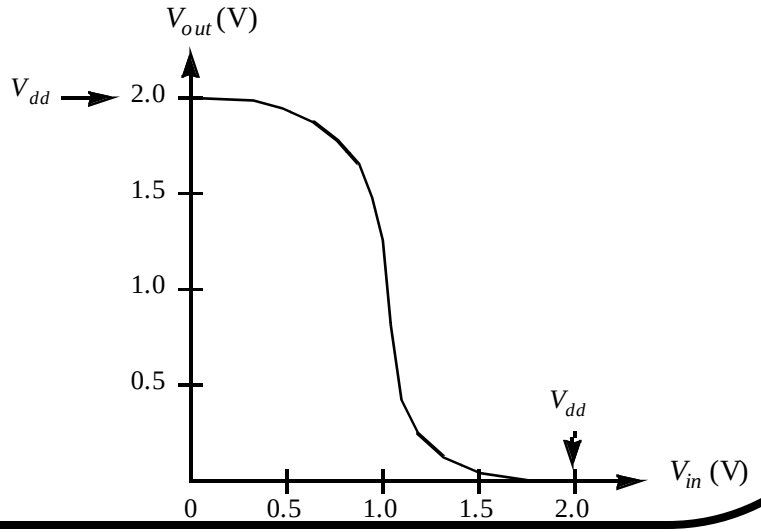
$$g_{msat} = \frac{W}{mL} C_{oxe} \mu_{ns} (V_{gs} - V_t)$$

6.7.1 CMOS Inverter--voltage transfer curve

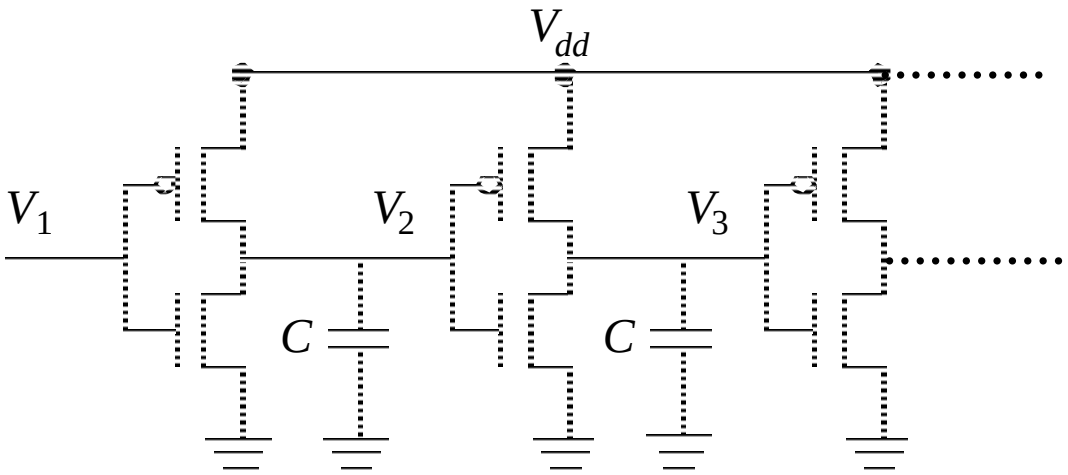


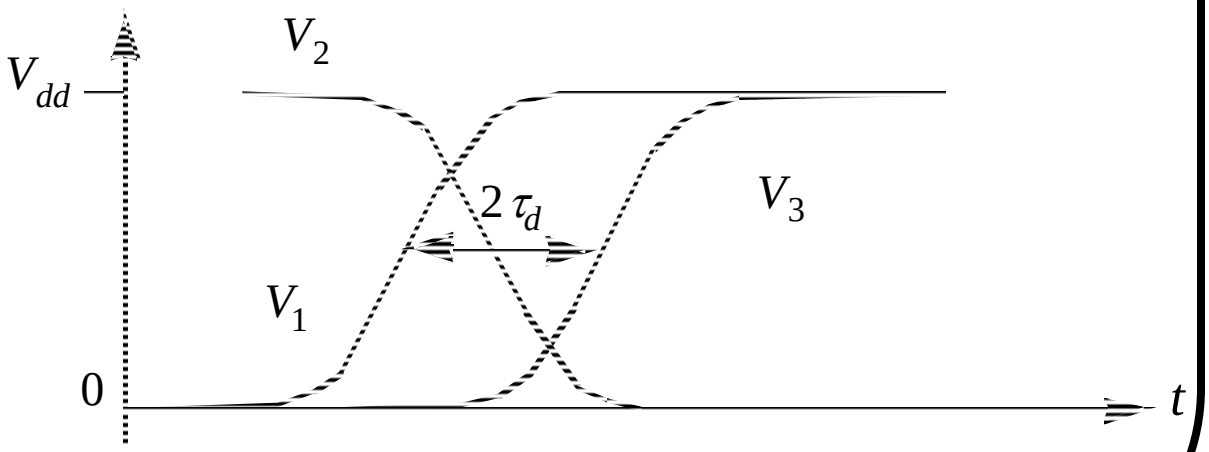






6.7.2 Inverter Speed – propagation delay





τ_d : propagation delay

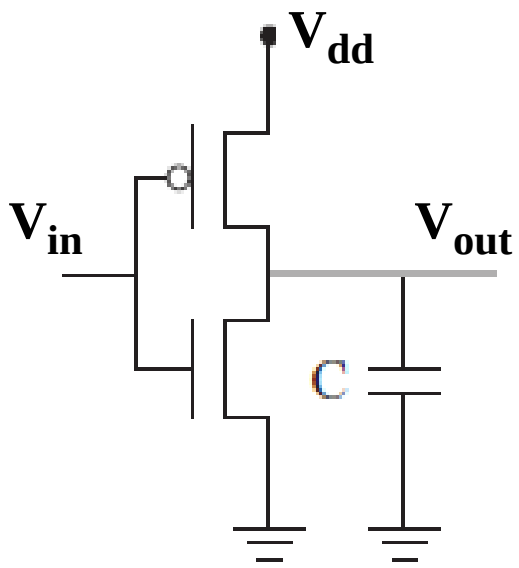
6.7.2 Inverter Speed - Impact of I_{on}

$$\tau_d \equiv \frac{1}{2}(pull-down\ delay\ +\ pull-up\ delay)$$

$$pull-up\ delay \approx \frac{CV_{dd}}{2I_{onP}}$$

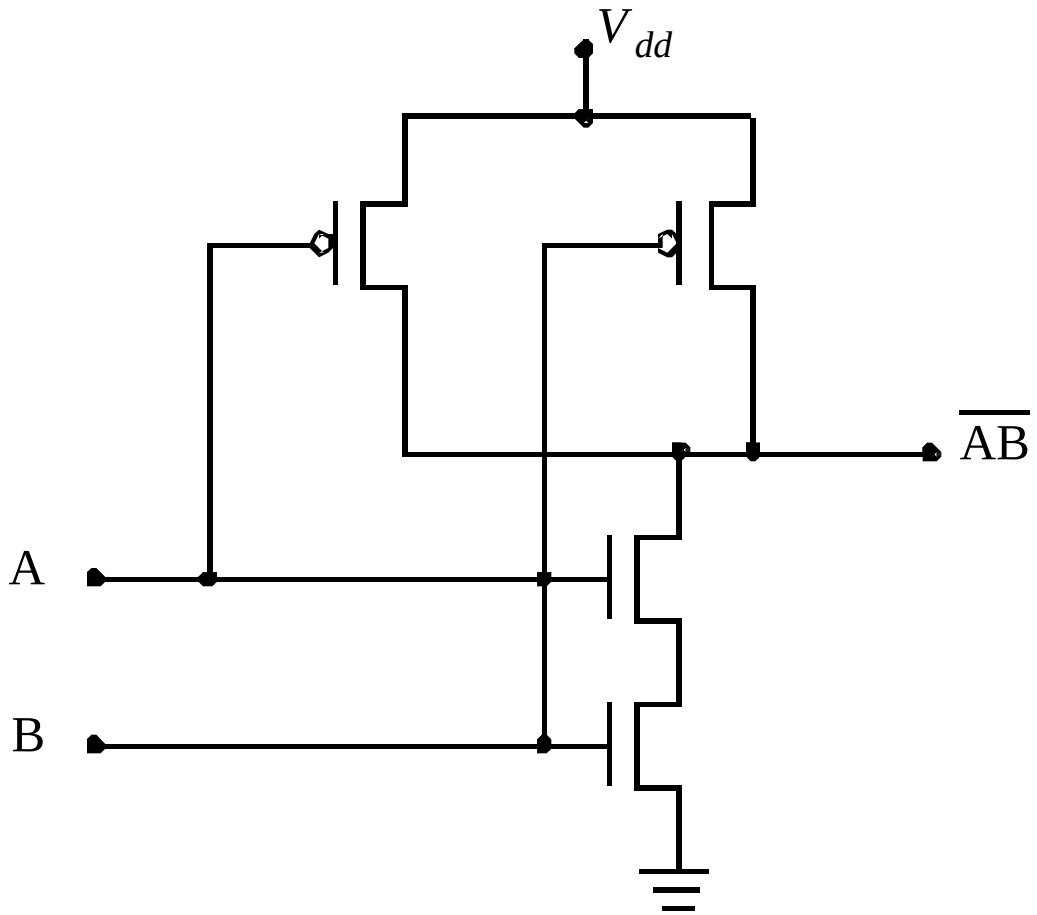
$$pull-down\ delay \approx \frac{CV_{dd}}{2I_{onN}}$$

$$\tau_d = \frac{CV_{dd}}{4}\left(\frac{1}{I_{onN}} + \frac{1}{I_{onP}}\right)$$



How can the speed of an inverter circuit be improved?

Logic Gates



This two-input NAND gate and many other logic gates are extensions of the inverter.

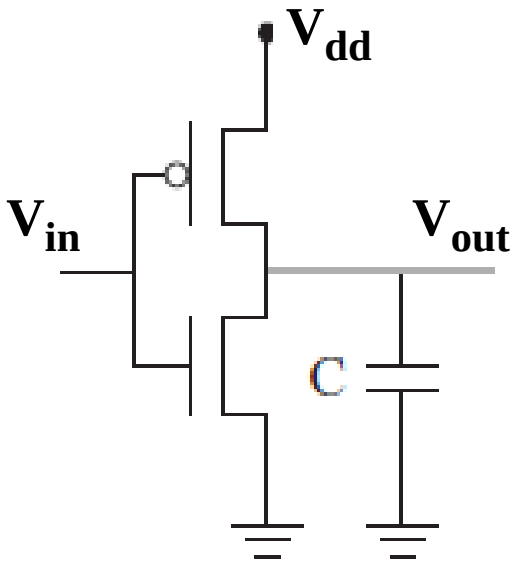
6.7.3 Power Consumption

$$P_{dynamic} = V_{dd} \times average\ current = k\,CV_{dd}^2\,f$$

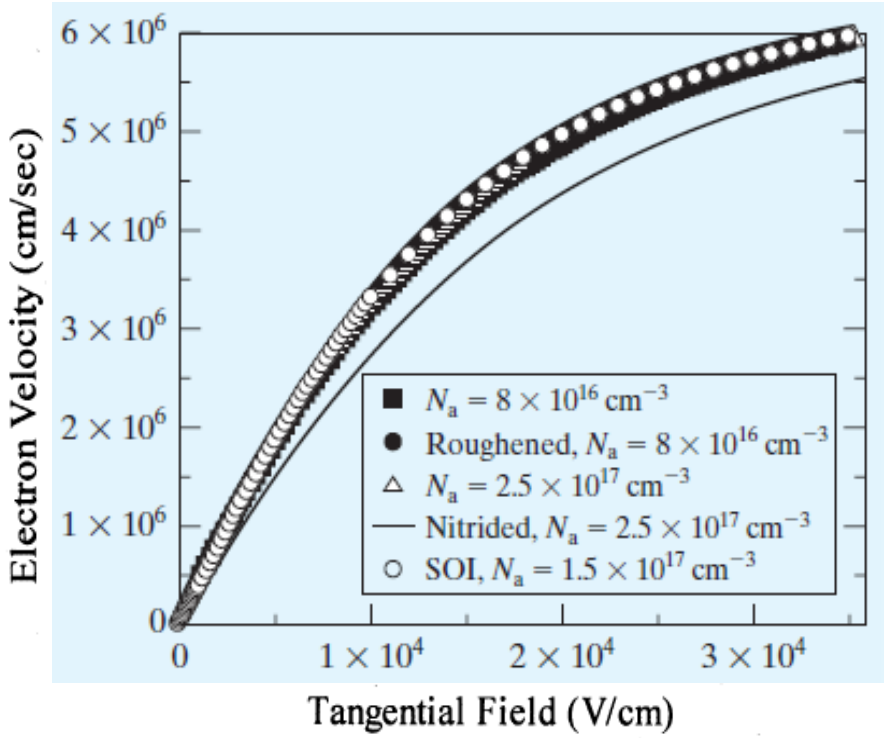
$$P_{static} = V_{dd}\,I_{off}$$

Total power consumption

$$P = P_{dynamic} + P_{static}$$



6.8 Velocity Saturation



$$v = \frac{\mu_{ns}\mathcal{E}}{1 + \dfrac{\mathcal{E}}{\mathcal{E}_{sat}}}$$

$$\mathcal{E} << \mathcal{E}_{sat} : v = \mu_{ns}\mathcal{E}$$

$$\mathcal{E} >> \mathcal{E}_{sat} : v = \mu_{ns}\mathcal{E}_{sat}$$

- Velocity saturation has large and deleterious effect on the I_{on} of MOSFETS

6.9 MOSFET IV Model with Velocity Saturation

$$I_{ds} = WQ_{inv}v$$

$$I_{ds} = WC_{oxe}(V_{gs} - mV_{cs} - V_t)\frac{\mu_{ns}dV_{cs}/dx}{1+\frac{dV_{cs}}{dx}/\mathcal{E}_{sat}}$$

$$\int_0^L I_{ds}dx = \int_0^{V_{ds}} [WC_{oxe}\mu_{ns}(V_{gs} - mV_{cs} - V_t) - I_{ds}/\mathcal{E}_{sat}]dV_{cs}$$

$$I_{ds}L = WC_{oxe}\mu_{ns}(V_{gs} - V_t - \frac{m}{2}V_{ds})V_{ds} - I_{ds}V_{ds}/\mathcal{E}_{sat}$$

6.9 MOSFET IV Model with Velocity Saturation

$$I_{ds} = \frac{\frac{W}{L} C_{oxe} \mu_{ns} \left(V_{gs} - V_t - \frac{m}{2} V_{ds}\right) V_{ds}}{1 + \frac{V_{ds}}{\mathcal{E}_{sat} L}}$$

$$\boxed{I_{ds} = \frac{long\text{-}channel\ I_{ds}}{1 + V_{ds} / \mathcal{E}_{sat} L}}$$

6.9 MOSFET IV Model with Velocity Saturation

Solving $\frac{dI_{ds}}{dV_{ds}} = 0,$

$$V_{dsat} = \frac{2(V_{gs} - V_t)/m}{1 + \sqrt{1 + 2(V_{gs} - V_t)/m\mathcal{E}_{sat}L}}$$

A simpler and more accurate V_{dsat} is:

$$\boxed{\frac{1}{V_{dsat}} = \frac{m}{V_{gs} - V_t} + \frac{1}{\mathcal{E}_{sat}L}}$$

$$\mathcal{E}_{sat} \equiv \frac{2v_{sat}}{\mu_{ns}}$$

EXAMPLE: Drain Saturation Voltage

Question: *At V_{gs} = 1.8 V, what is the V_{dsat} of an NFET with T_{oxe} = 3 nm, V_t = 0.25 V, and W_{dmax} = 45 nm for (a) L =10 μm, (b) L = 1 um, (c) L = 0.1 μm, and (d) L = 0.05 μm?*

Solution: *From V_{gs}, V_t, and T_{oxe}, μ_{ns} is 200 $cm^2V^{-1}s^{-1}$.*

$$\mathcal{E}_{sat} = 2v_{sat}/\mu_{ns} = 8 \times 10^4 \text{ V/cm}$$

$$m = 1 + 3T_{oxe}/W_{dmax} = 1.2$$

$$V_{dsat} = \left(\frac{m}{V_{gs} - V_t} + \frac{1}{\mathcal{E}_{sat} L} \right)^{-1}$$

EXAMPLE: Drain Saturation Voltage

$$V_{dsat} = \left(\frac{m}{V_{gs} - V_t} + \frac{1}{\mathcal{E}_{sat} L} \right)^{-1}$$

(a) L = 10 μm, $V_{dsat} = (1/1.3\text{V} + 1/80\text{V})^{-1} = 1.3$ V

(b) L = 1 μm, $V_{dsat} = (1/1.3\text{V} + 1/8\text{V})^{-1} = 1.1$ V

(c) L = 0.1 μm, $V_{dsat} = (1/1.3\text{V} + 1/.8\text{V})^{-1} = 0.5$ V

(d) L = 0.05 μm, $V_{dsat} = (1/1.3\text{V} + 1/.4\text{V})^{-1} = 0.3$ V

I_{dsat} with Velocity Saturation

Substituting V_{dsat} for V_{ds} in I_{ds} equation gives:

$$I_{dsat} = \frac{W}{2mL} C_{oxe} \mu_s \frac{(V_{gs} - V_t)^2}{1 + \frac{V_{gs} - V_t}{m\mathcal{E}_{sat} L}} = \frac{long\text{-}channel\ I_{dsat}}{1 + \frac{V_{gs} - V_t}{m\mathcal{E}_{sat} L}}$$

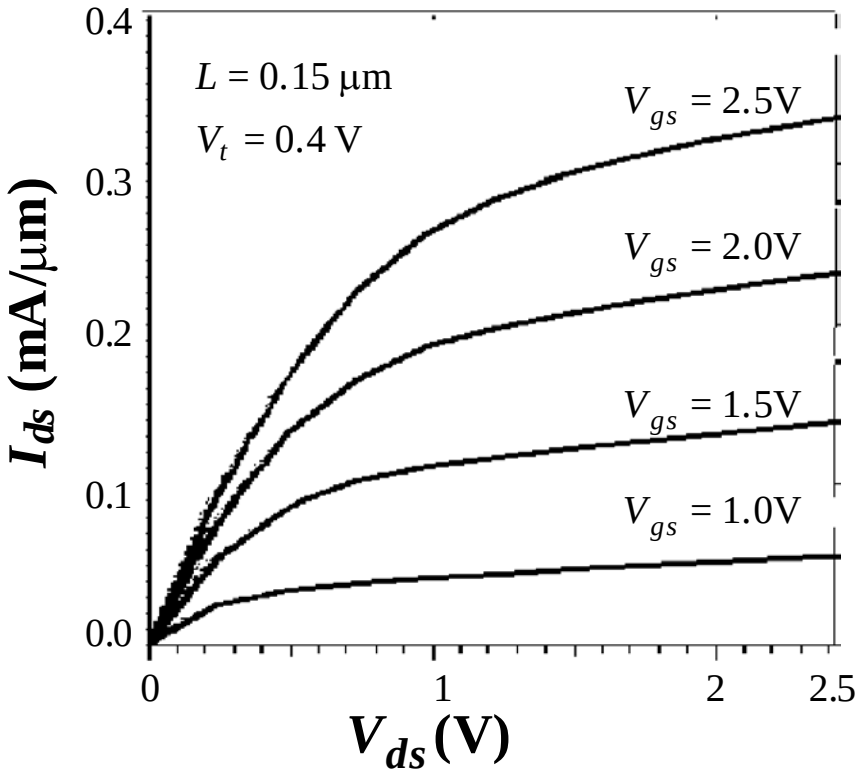
Very short channel case: $\mathcal{E}_{sat} L << V_{gs} - V_t$

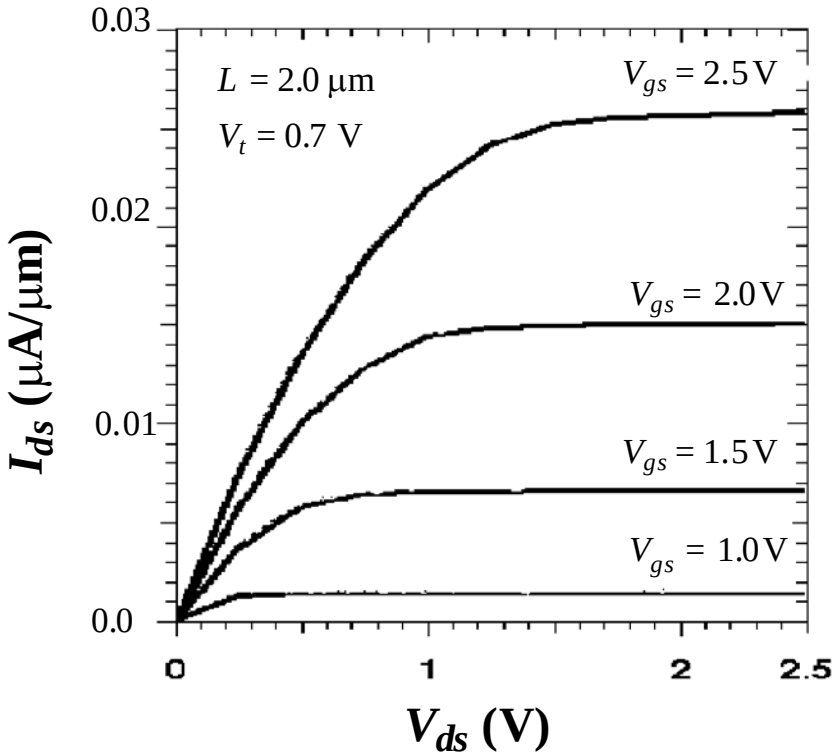
$$I_{dsat} = W v_{sat} C_{oxe} (V_{gs} - V_t - m\mathcal{E}_{sat} L)$$

$$I_{dsat} = W v_{sat} C_{oxe} (V_{gs} - V_t)$$

- I_{dsat} is proportional to $V_{gs} - V_t$ rather than $(V_{gs} - V_t)^2$, not as sensitive to L as $1/L$.

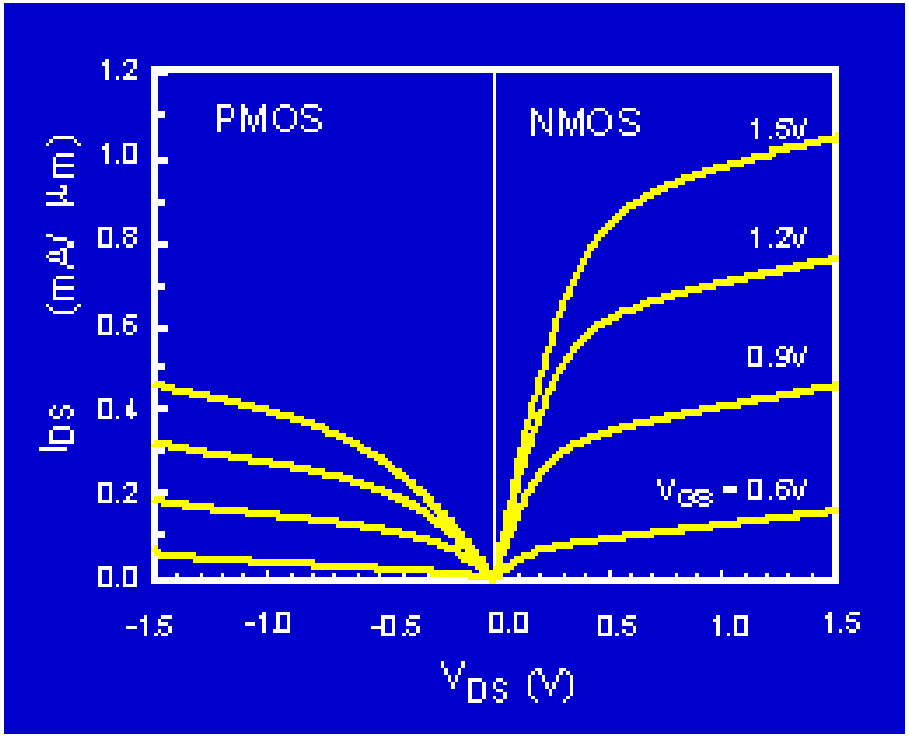
Measured MOSFET IV





What is the main difference between the V_g dependence of the long- and short-channel length IV curves?

PMOS and NMOS IV Characteristics



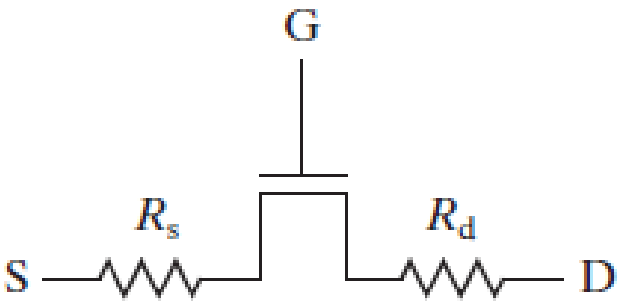
The PMOS IV is qualitatively similar to the NMOS IV, but the current is about half as large. How can we design a CMOS inverter so that its voltage transfer curve is symmetric?

6.9.1 Velocity Saturation vs. Pinch-Off

Current saturation : the carrier velocity reaches V_{sat} at the drain.

Instead of the **pinch-off region**, there is a **velocity saturation region** next to the drain where Q_{inv} is a constant (I_{dsat}/W_{vsat}).

6.10 Parasitic Source-Drain Resistance

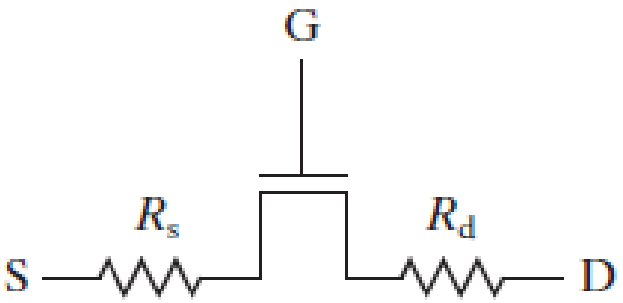


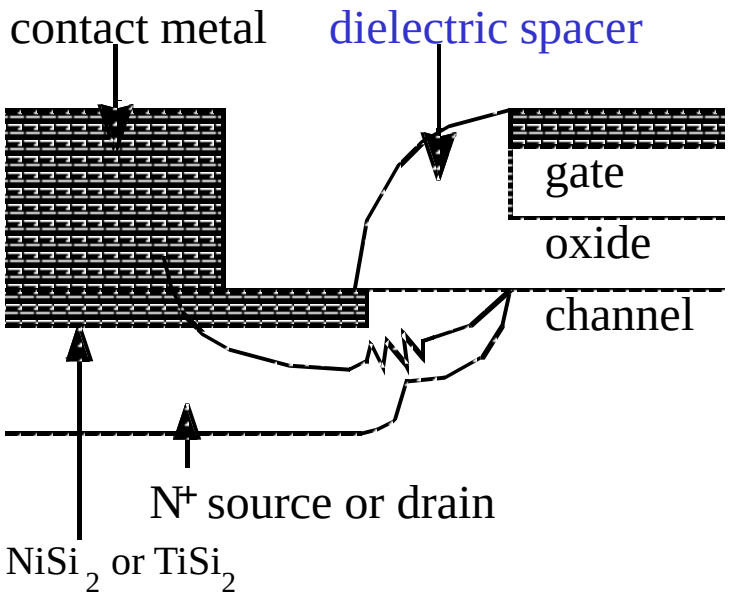
- If $I_{dsat0} \propto V_g - V_t$, $I_{dsat} = \dfrac{I_{dsat0}}{1 + \dfrac{I_{dsat0} R_s}{(V_{gs} - V_t)}}$

- I_{dsat} can be reduced by about 15% in a 0.1μm MOSFET. Effect is greater in shorter MOSFETs.

- $V_{dsat} = V_{dsat0} + I_{dsat}(R_s + R_d)$

SALICIDE (Self-Aligned Silicide) Source/Drain



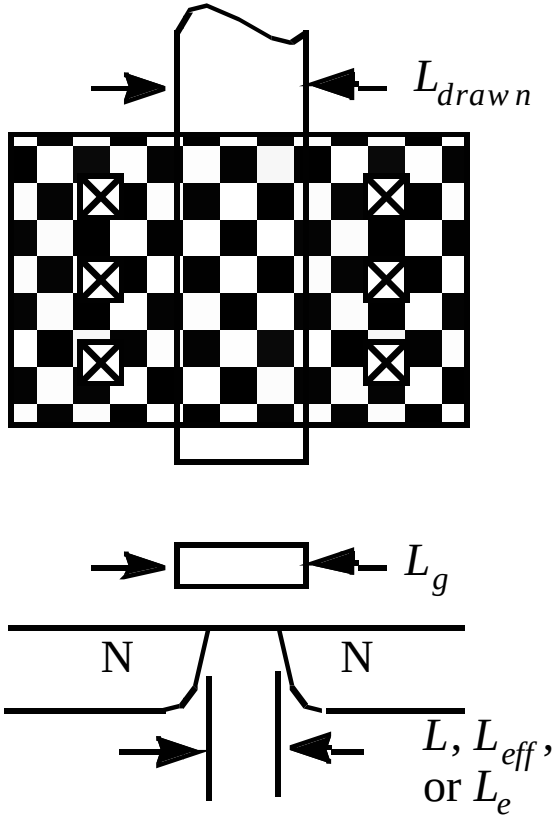


After the spacer is formed, a Ti or Mo film is deposited. Annealing causes the silicide to be formed over the source, drain, and gate. Unreacted metal (over the spacer) is removed by wet etching.

Question:

- What is the purpose of siliciding the source/drain/gate?
- What is self-aligned to what?

Definitions of Channel Length



$$L \equiv L_g - \Delta L$$

6.11 Extraction of the Series Resistance and the Effective Channel Length

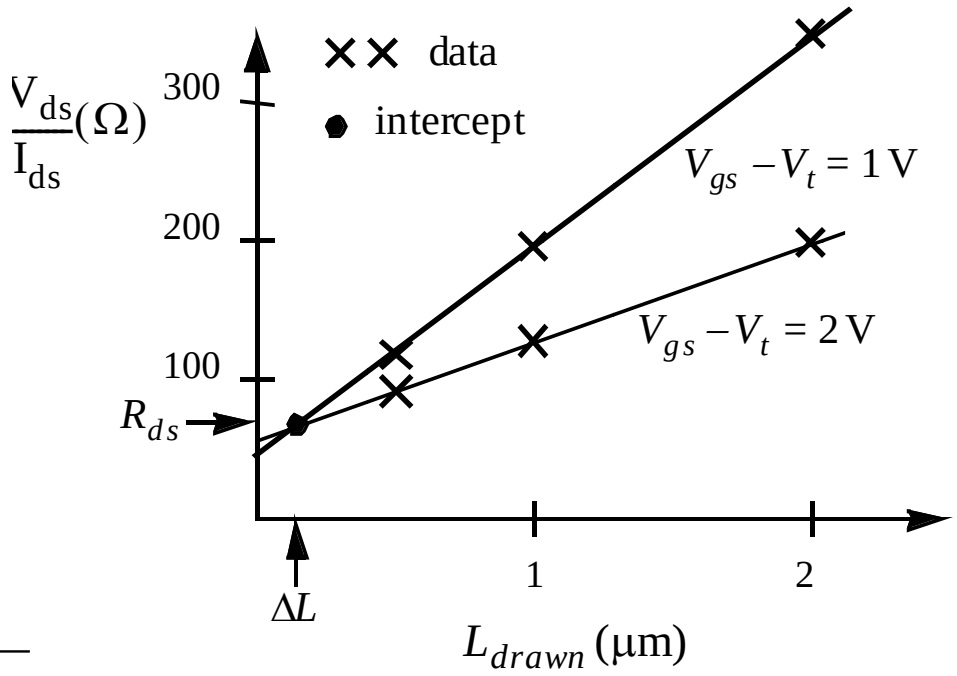
$$I_{ds} = \frac{WC_{oxe}\mu_s V_{ds}}{L_{drawn} - \Delta L}(V_{gs} - V_t)$$

$$V_{ds} = \frac{I_{ds}(L_{drawn} - \Delta L)}{WC_{oxe}(V_{gs} - V_t)\mu_s}$$

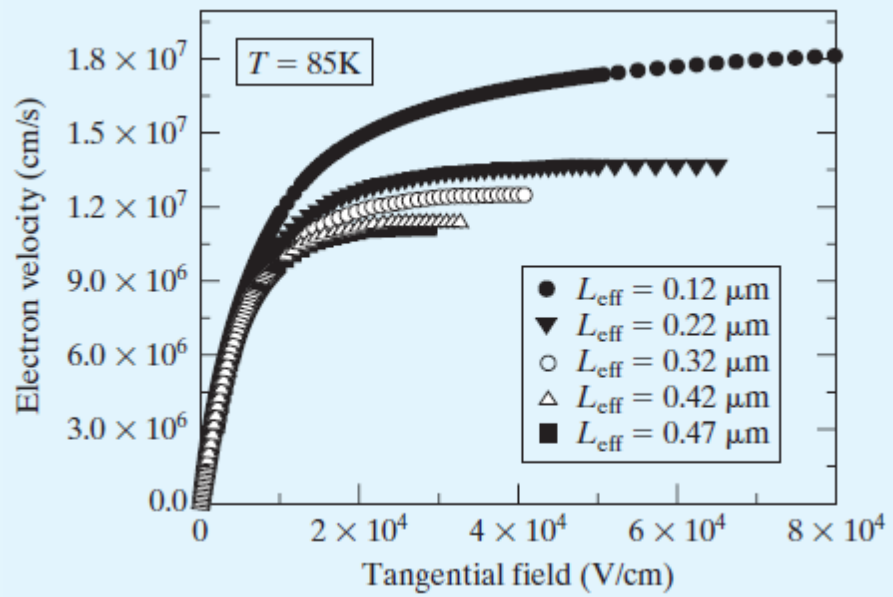
Include series resistance,

$R_{ds} \equiv R_d + R_s$,

$$\frac{V_{ds}}{I_{ds}} = R_{ds} + \frac{L_{drawn} - \Delta L}{WC_{oxe}(V_{gs} - V_t)\mu_s}$$

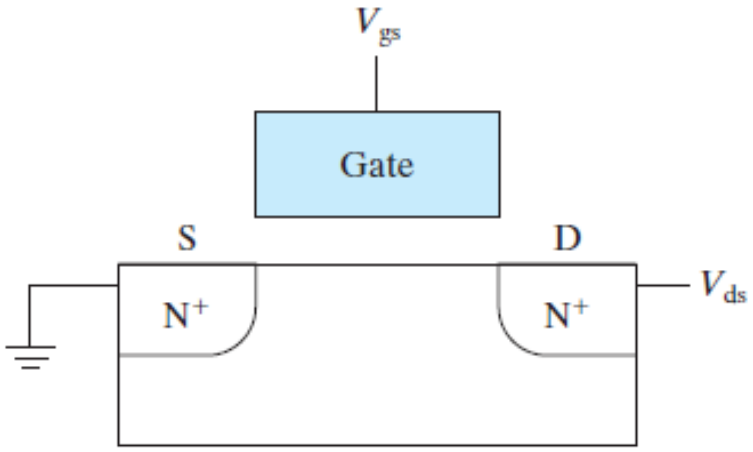


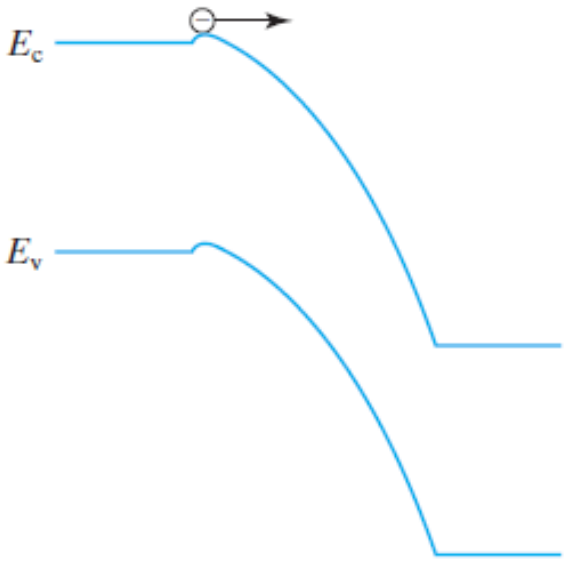
6.12 Velocity Overshoot



- Velocity saturation should not occur in very short MOSFETs.
- This velocity overshoot could lift the limit on Ids .
- *But…*

6.12 Source Velocity Limit





- Carrier velocity is limited by the thermal velocity with which they enter the channel from the source.

- $I_{dsat} = WBv_{thx}Q_{inv}$
  $= WBv_{thx}C_{oxe}(V_{gs} - V_t)$

- Similar to

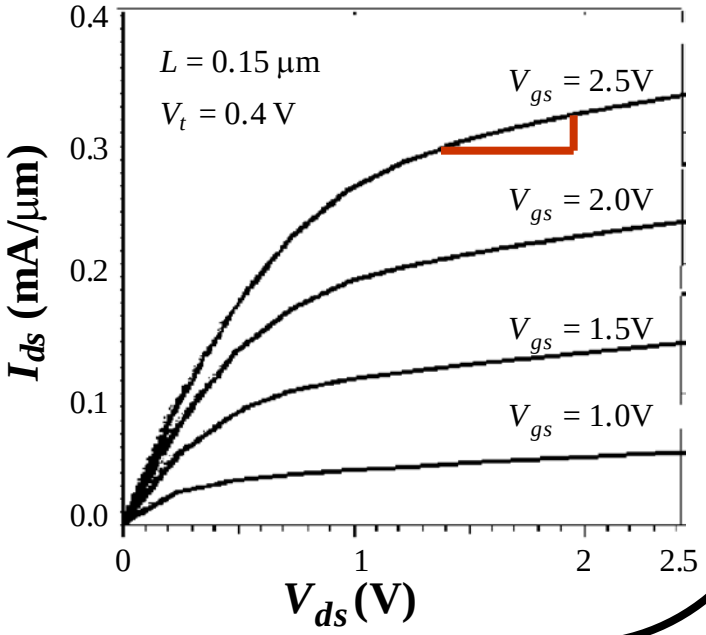
$$I_{dsat} = Wv_{sat}C_{oxe}(V_{gs} - V_t)$$

6.13 Output Conductance

- I_{dsat} does NOT saturate in the saturation region, especially in short channel devices!
- The slope of the I_{ds}-V_{ds} curve in the saturation region is called the **output conductance (g_{ds})**,

$$g_{ds} \equiv \frac{dI_{dsat}}{dV_{ds}}$$

- A smaller g_{ds} is desirable for a large voltage gain, which is beneficial to analog and digital circuit applications.



Example of an Amplifier

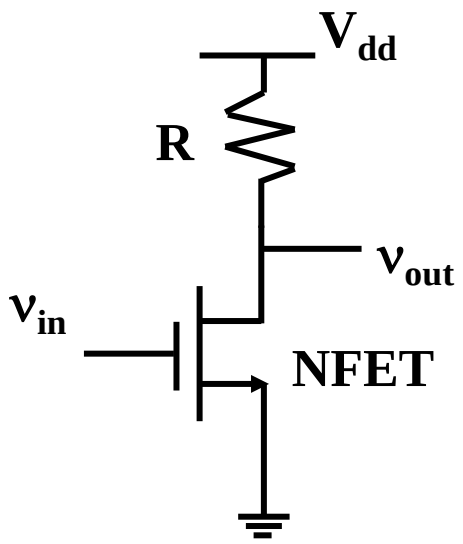
- The transistor operates in the saturation region. A ***small signal*** input, v_{in}, is applied.

$$i_{ds} = g_{msat} \cdot v_{gs} + g_{ds} \cdot v_{ds}$$

$$= g_{msat} \cdot v_{in} + g_{ds} \cdot v_{out}$$

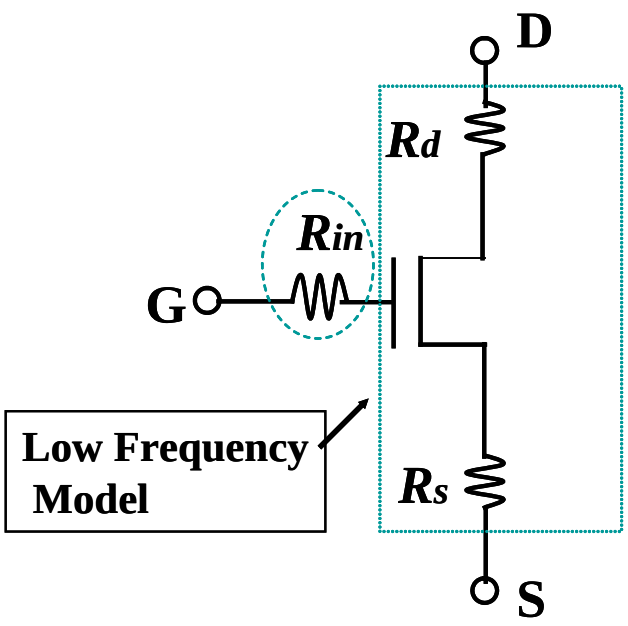
$$i_{ds} = -v_{out} / R \cdot$$

$$\Rightarrow \quad v_{out} = \frac{-g_{msat}}{(g_{ds} + 1/R)} \times v_{in}$$



- The voltage gain is $g_{msat}/(g_{ds} + 1/R)$.
- A smaller g_{ds} is desirable for large voltage gain.
- Maximum available gain (or intrinsic voltage gain) is g_{msat}/g_{ds}

6.14 High-Frequency Performance



High-frequency performance is limited by input R and/or C.

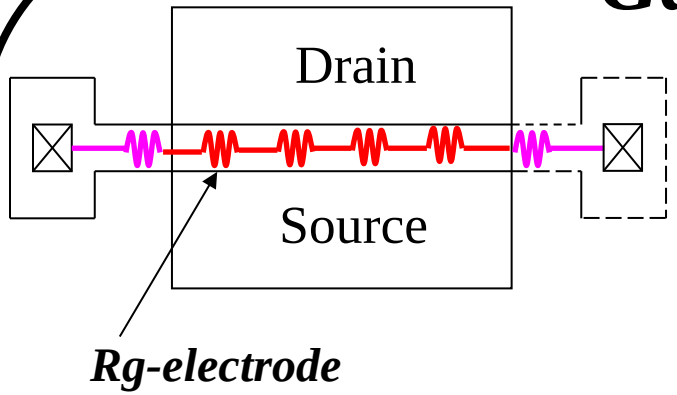
Cutoff frequency (f_T) : Frequency at which the output current becomes equal to the input current.

Maximum oscillation frequency (f_{max}) : Frequency at which the power gain drops to unity

$$R_{in} = R_{g-electrode} + R_{ii}$$

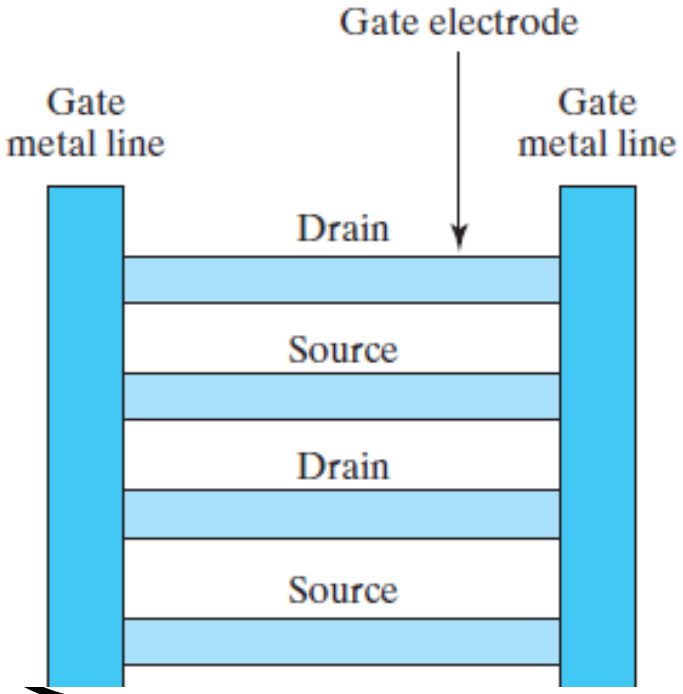
Gate-electrode resistance **Intrinsic input resistance**

Gate-Electrode Resistance



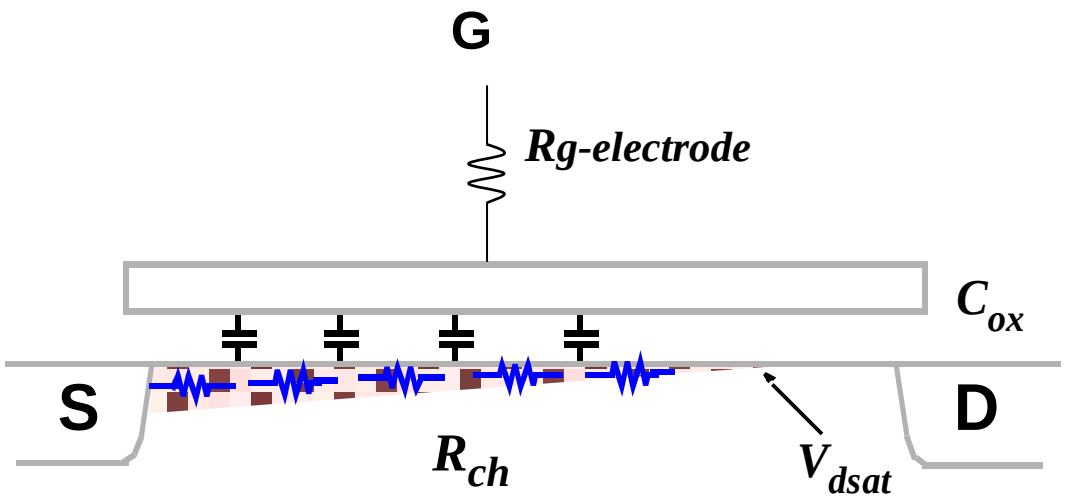
Multi-finger layout greatly reduces the gate electrode resistance

$$R_{g-electrode} = \rho W / 12 T_g L_g N_f^2$$



ρ : resistivity of gate material,
W_f : width of each gate finger,
T_g : gate thickness,
L_g : gate length,
N_f : number of fingers.

Intrinsic Input Resistance



$$R_{ii} = \kappa \int dR_{ch} = \kappa \frac{V_{ds}}{I_{ds}}$$

The gate capacitor current flows through R_{ch} to the source and ground.

6.15 MOSFET Noises

Noise : All that corrupts the signal

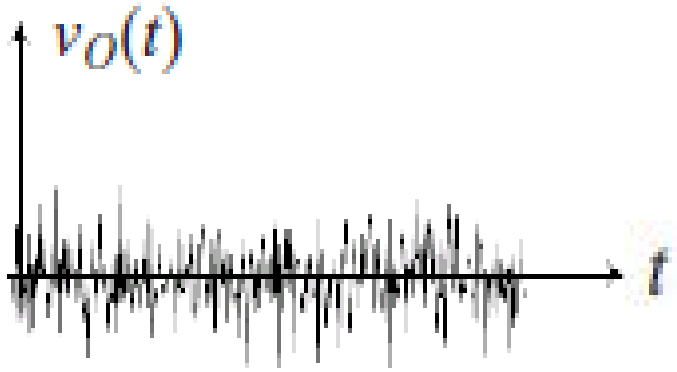
External noise:

- Inductive and capacitive interferences and cross talks created by wiring
- Needs to be controlled with shielding and circuit layout carefully

Fundamental noise:

- Noise inherent to the electronic devices.
- Due to the random behaviors of the electric carriers inside the device

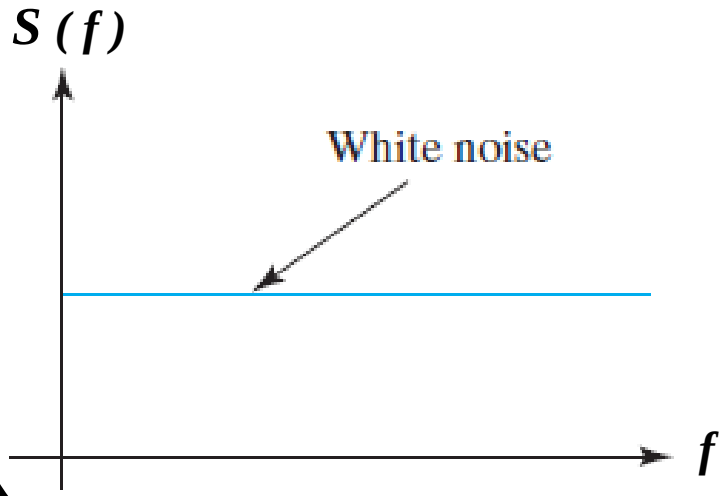
6.15.1 Thermal Noise of a Resistor



Thermal noise: caused by random thermal motion of the charge carriers

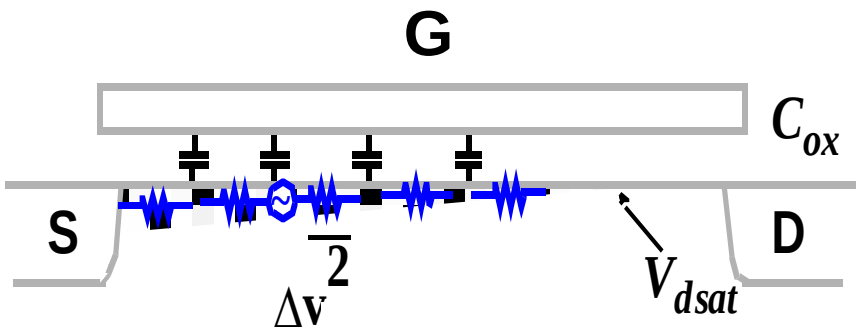
$$\overline{v_n^2} = 4kT\Delta f R = S_{v_n}\Delta f$$

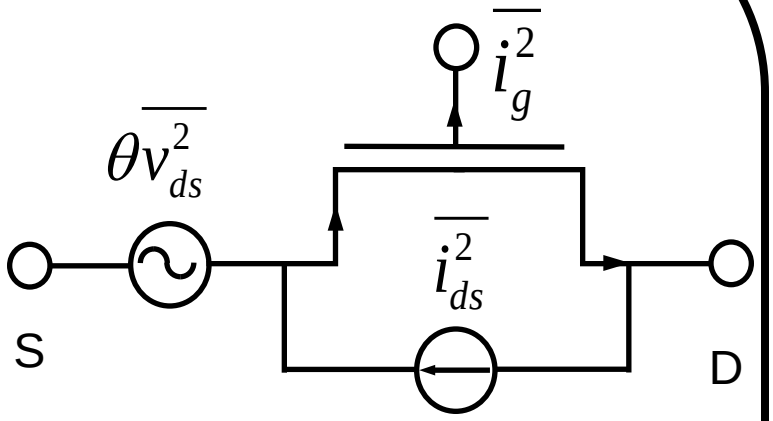
$$\overline{i_n^2} = 4kT\Delta f / R = S_{i_n}\Delta f$$



S **:** noise power density spectrum

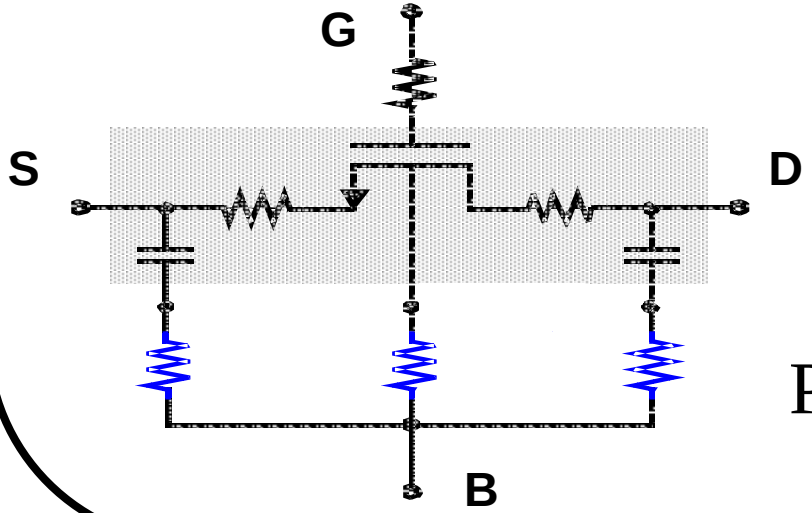
6.15.2 MOSFET Thermal Noise





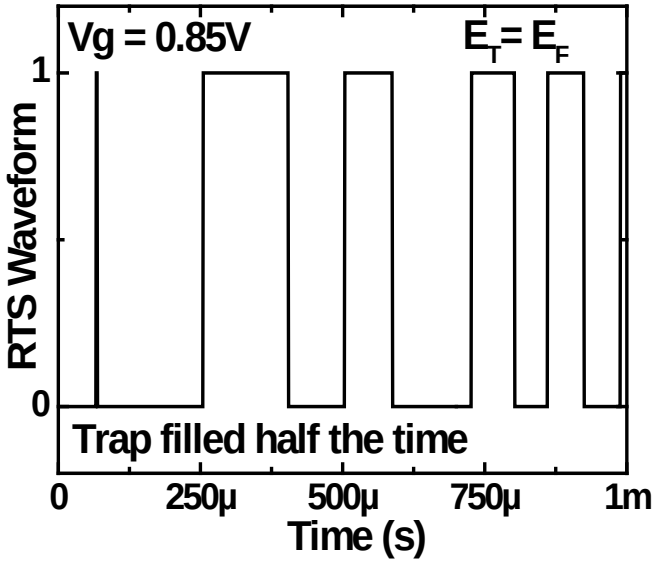
$$\overline{v_{ds}^2} = 4\gamma kT\Delta f \,/\, g_{ds}$$

$$\overline{i_{ds}^2} = 4\gamma kT\Delta f g_{ds}$$



Parasitic-resistance noise

6.15.3 MOSFET Flicker Noise

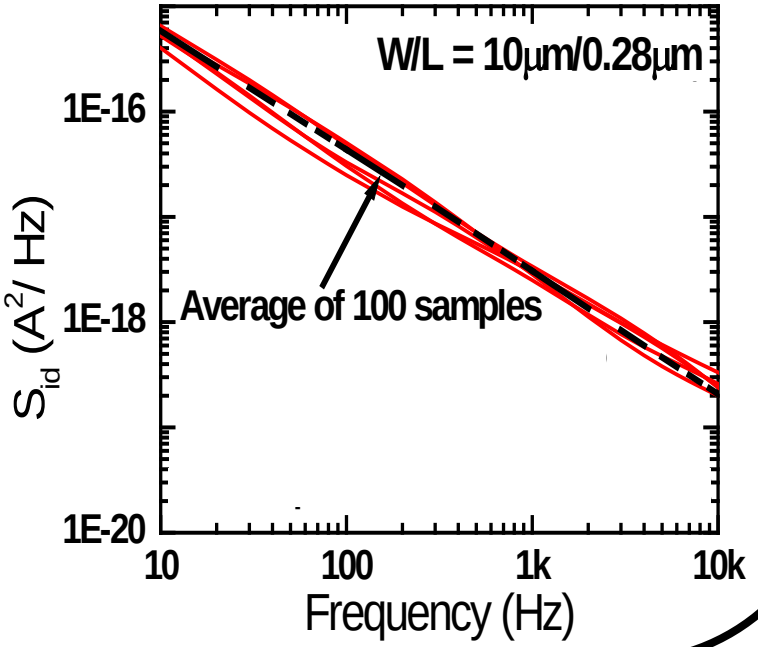


Charge trapping and releasing by a single oxide trap generate Random Telegraph Noise

Many traps produce a 1/f power density spectrum.

1/f noise

$$\overline{i_{ds}^2} = \frac{KF \cdot W}{fL^2 C_{ox}} \left(\frac{I_{ds}}{W}\right)^{AF} \cdot kT\Delta f$$



6.15.4 Signal to Noise Ratio, Noise Factor, Noise Figure

SNR: Signal power ÷ noise power.

Decibel or dB:10 times the base-10 logarithm of the noise power.

$$10 \times \log \frac{S}{N}$$

Noise factor: The ratio of the input SNR and output SNR.

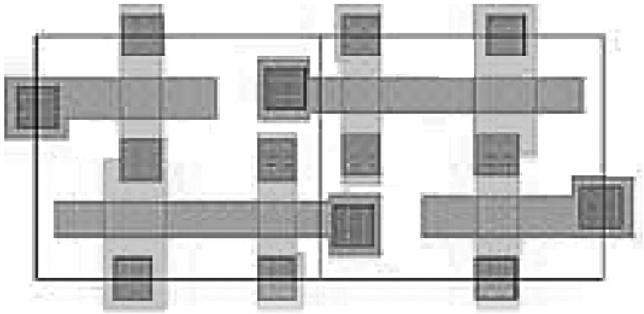
$$F = \frac{S_i / N_i}{S_0 / N_0}$$

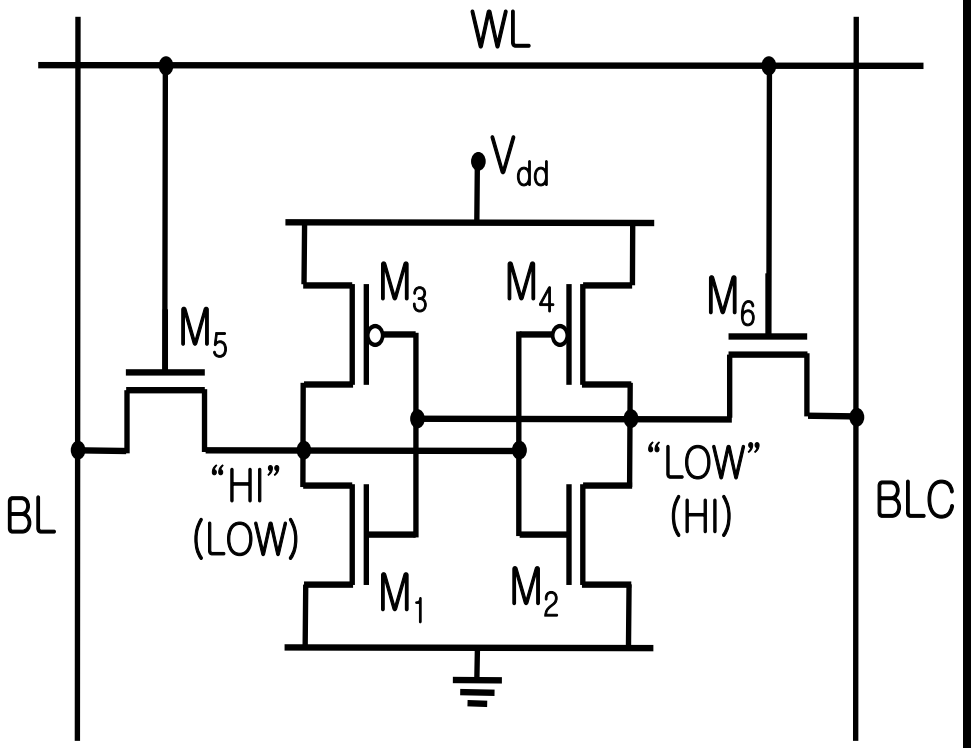
6.16 Memory Devices

	Keep data without power?	*Cell size and cost/bit*	*Rewrite cycles*	*Write-one-byte speed*	*Compatible with basic CMOS fabrication*	*Main applications*
SRAM	*No*	*Large*	*Unlimited*	*Fastest*	*Totally*	*Embedded in logic chips*
DRAM	*No*	*Small*	*Unlimited*	*Fast*	*Needs modification*	*Stand-alone main memory*
Flash memory (NVM)	*Yes*	*Smallest*	*Limited*	*Slow*	*Needs extensive modification*	*Nonvolatile data and code storage*

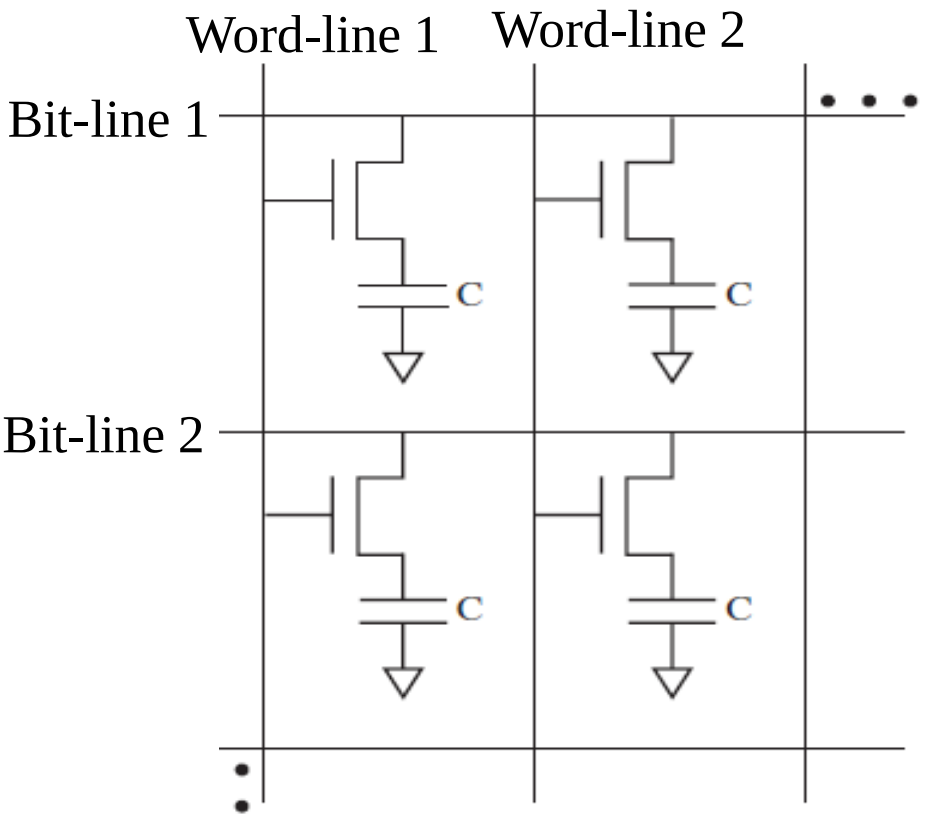
6.16.1 SRAM

>Fastest among all memories.
>Totally CMOS compatible.
>Cost per bit is the highest-- uses 6 transistors to store one bit of data.



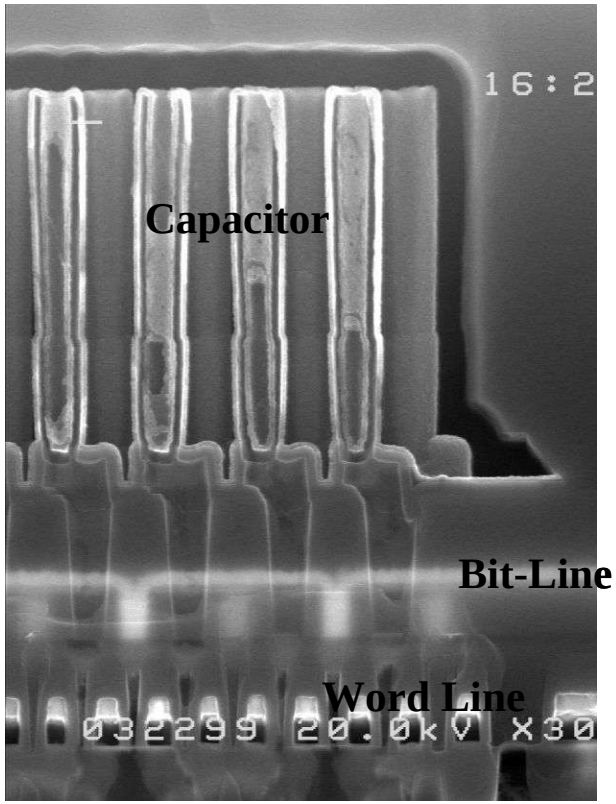


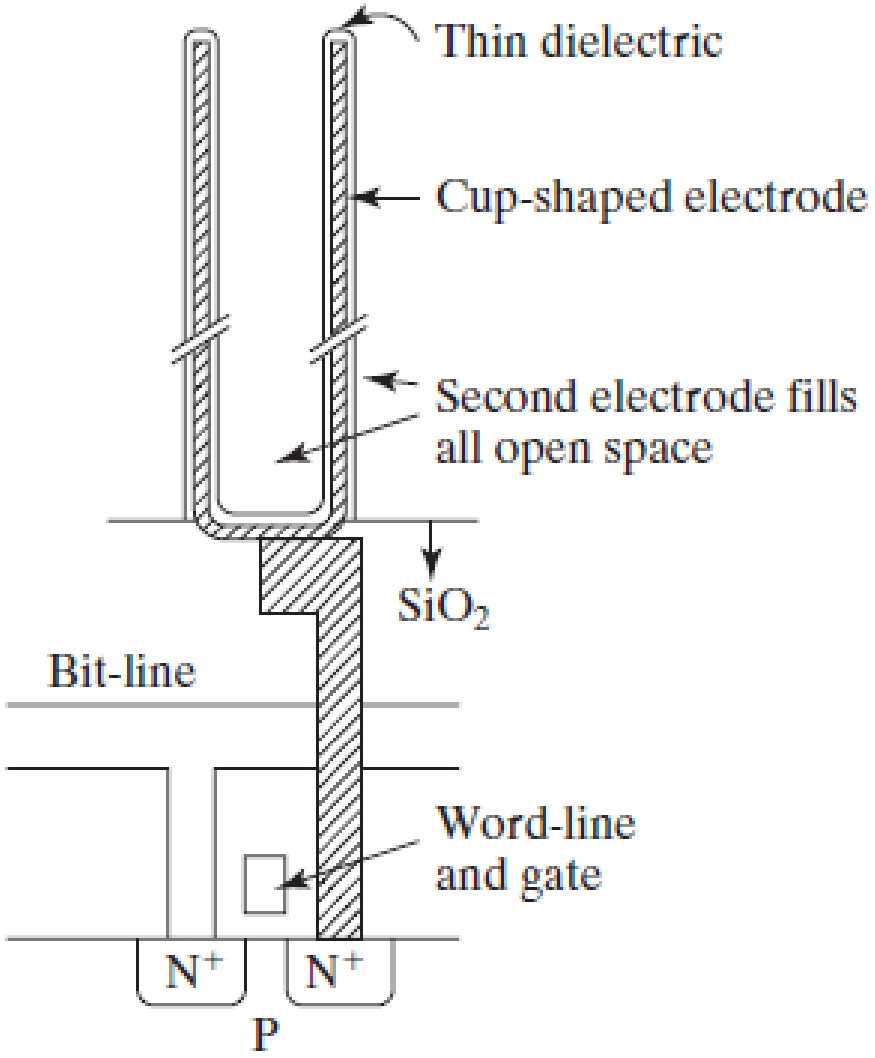
6.16.2 DRAM



- DRAM capacitor can only hold the data (charge) for a limited time because of leakage current.
- Needs refresh.
- Needs ~10fF C in a small and shrinking area -- for refresh time and error rate.

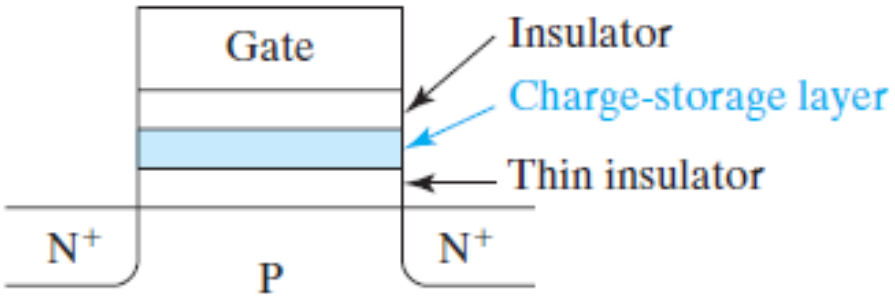
6.16.2 DRAM capacitor technology

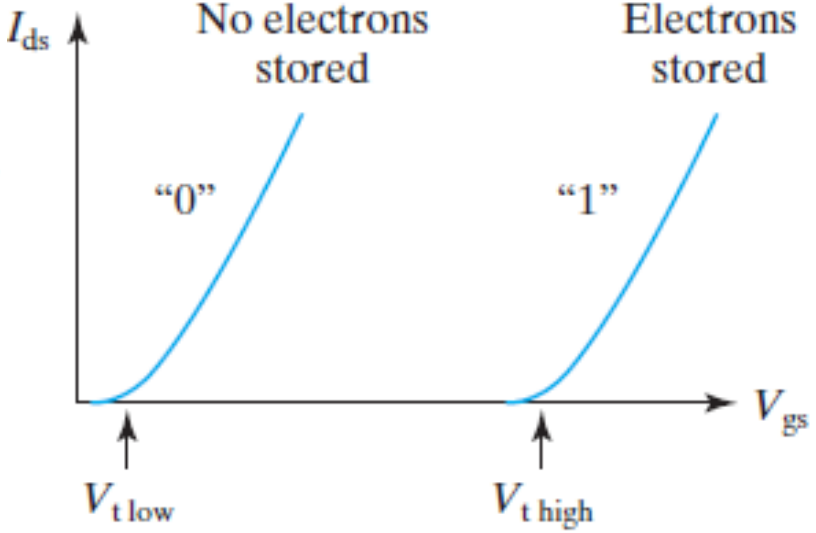


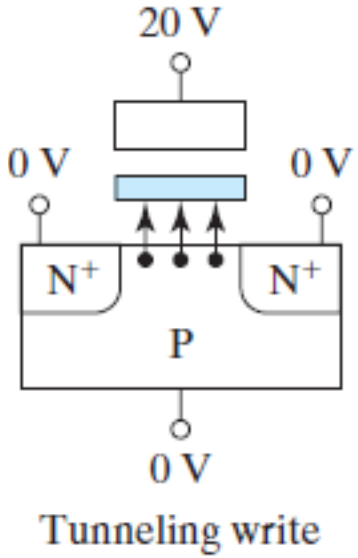


Stacked capacitor* and *Trench capacitor

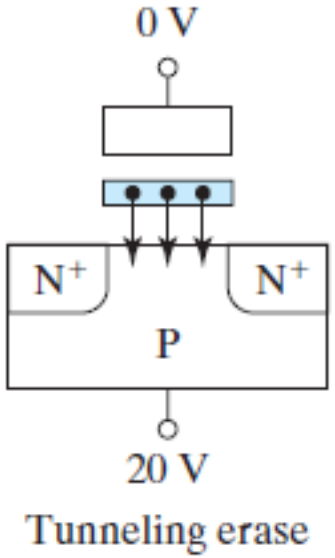
6.16.3 Nonvolatile (Flash) Memory



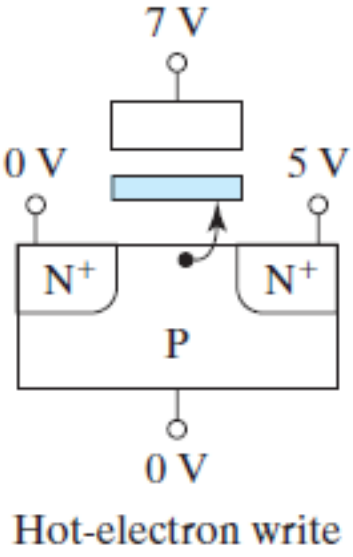




Tunneling write



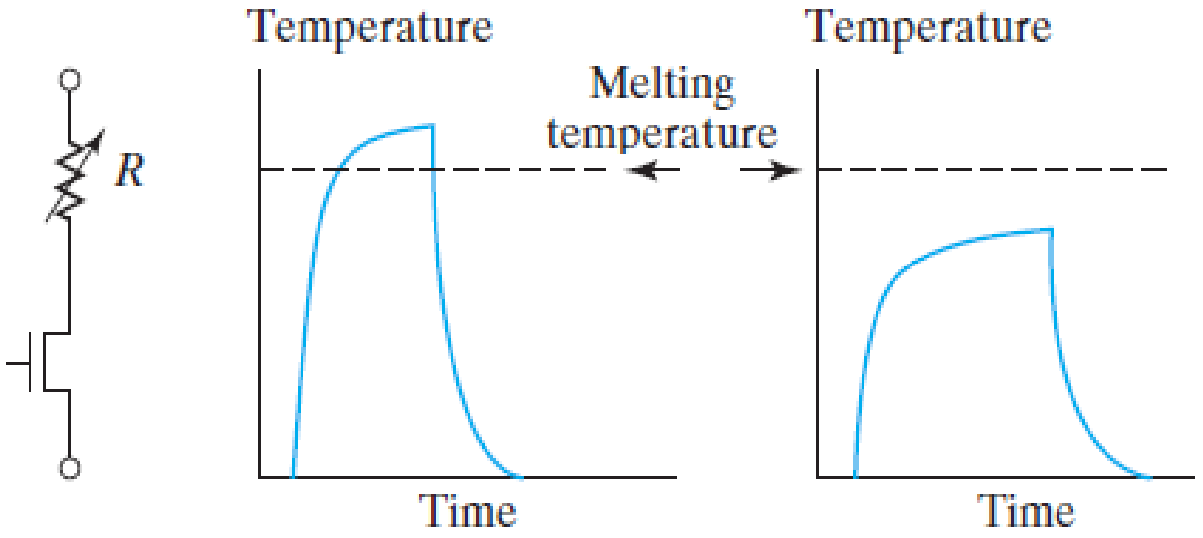
Tunneling erase



Hot-electron write

- **Floating gate (poly-Si)**
- **Charge trap (SONOS)**
- **Nanocrystal**

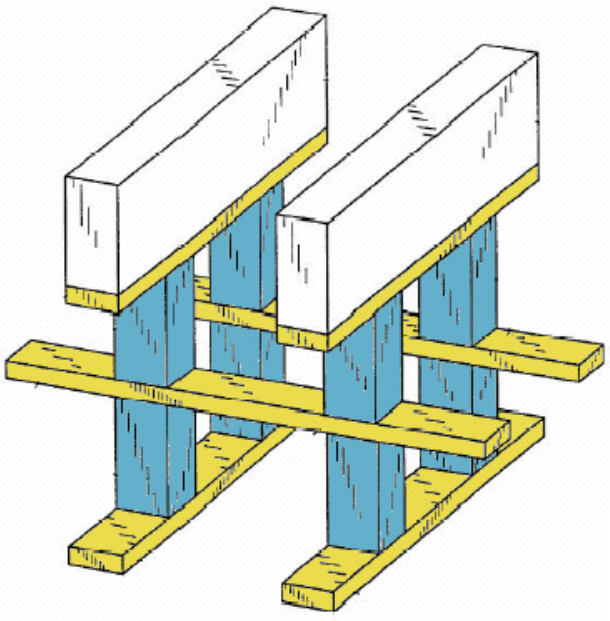
Phase Change Memory



Alloy of Ge, Sb, Te has high resistivity in amorphous phase and low resistivity in polycrystalline phase.

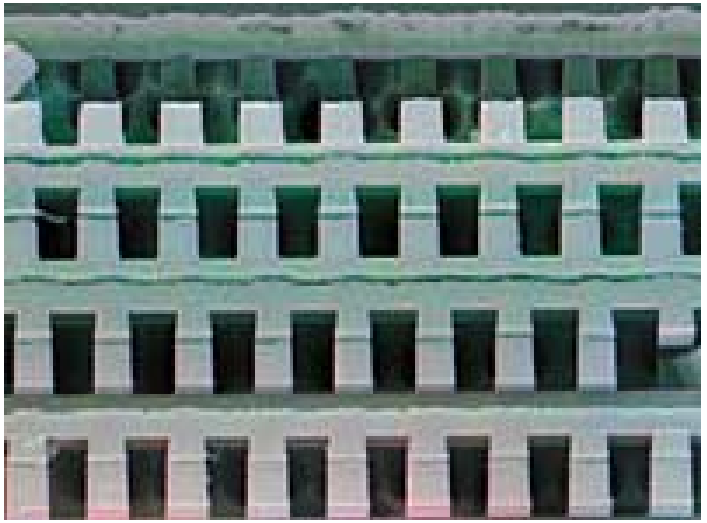
3D (Multi-layer) Memory

- Epitaxy from seed windows can produce Si layers.
- Ideally memory element is simple and does not need single-crystalline material.



Blue = Device

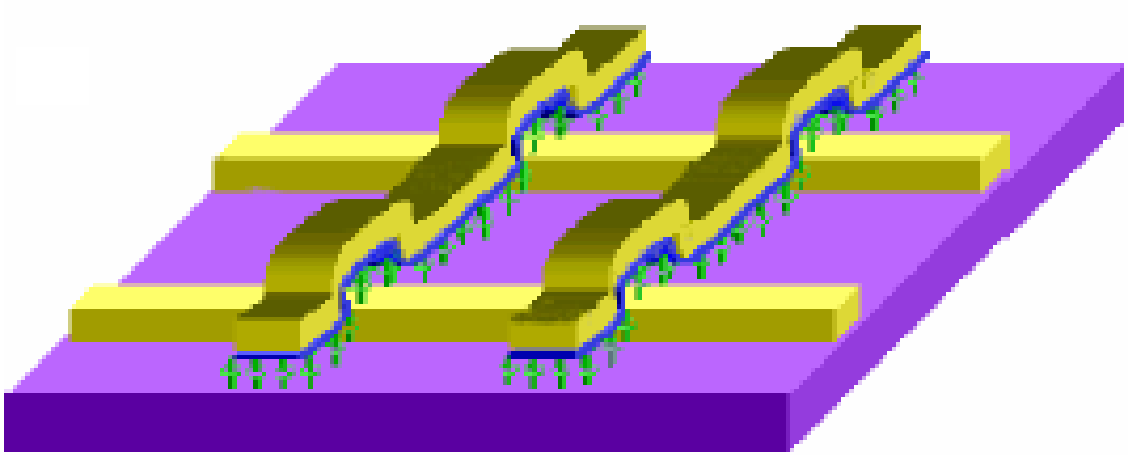
Yellow = Conductor



Resistive Memory (RRAM)

-- Organic, inorganic, metallic.. material

-- Future extension to 3-D



6.17 Chapter Summary

- ***propagation delay***

$$\tau_d \approx \frac{CV_{dd}}{4}\left(\frac{1}{I_{onN}} + \frac{1}{I_{onP}}\right)$$

- ***Power Consumption***

$$P = kCV_{dd}^2 f + V_{dd} I_{off}$$

- ***body effect***

$$\boxed{V_t(V_{sb}) = V_{t0} + \alpha V_{sb}}$$ for steep retrograde body doping

$$\alpha = 3T_{oxe} / W_{dmax}$$

6.17 Chapter Summary

- ***basic I_{ds} model***

$$I_{ds} = \frac{W}{L} C_{oxe} \mu_s \left(V_{gs} - V_t - \frac{m}{2} V_{ds}\right) V_{ds}$$

$$m = 1 + 3T_{oxe} / W_{dmax} \approx 1.2$$

• Small α and m are desirable. Therefore, small T_{oxe} is good. Ch.7 shows that large W_{dmax} is not acceptable.
• CMOS circuit speed is determined by CV_{dd}/I_{dsat}, and its power by $CV_{dd}^2 f + V_{dd} I_{off}$.

6.17 Chapter Summary

IV characteristics can be divided into a ***linear region*** and a ***saturation region.***

I_{ds} saturates at:

$$V_{dsat} = \frac{V_{gs} - V_t}{m}$$

$$I_{dsat} = \frac{W}{2mL} C_{oxe} \mu_s (V_{gs} - V_t)^2$$

transconductance:

$$g_{msat} = \frac{W}{mL} C_{oxe} \mu_s (V_{gs} - V_t)$$

Considering ***velocity saturation,***

$$V_{dsat} = \left(\frac{m}{V_{gs} - V_t} + \frac{1}{\mathcal{E}_{sat} L} \right)^{-1}$$

$$I_{dsat} = \frac{long\text{-}channel\ I_{dsat}}{1 + \frac{V_{gs} - V_t}{m\mathcal{E}_{sat} L}}$$

6.17 Chapter Summary

- At very small L $I_{dsat} = Wv_{sat}C_{oxe}(V_{gs} - V_t)$

- Velocity overshoot can lift v_{sat}, but source velocity limit sets a similar top over I_{dsat}.

$$I_{dsat} = WBv_{thx}C_{oxe}(V_{gs} - V_t)$$

- Intrinsic voltage gain is g_{msat}/g_{ds}

- High f_T and f_{MAX} need low $R_{in} = R_{g-electrode} + R_{ii}$

$$R_{ii} \propto \frac{V_{ds}}{I_{ds}} \qquad R_{g-electrode} \propto N_f^{\,2}$$

- Noise arises from the channel, gate, substrate thermal noises, and the flicker noise.

6.17 Chapter Summary

SRAM, DRAM, Nonvolatle memory

	Keep Data Without Power?	Cell Size and Cost/bit	Rewrite Cycles	Write-One-byte Speed	Compatible with Basic CMOS Manufacturing	Main Applications
SRAM	No	Large	Unlimited	Fast	Totally	Embedded in logic chips
DRAM	No	Small	Unlimited	Fast	Need modifications	Stand-alone chips and embedded
Flash memory	Yes	Smallest	Limited	Slow	Need extensive modifications	Nonvolatile storage stand-alone

Chapter 7 MOSFETs in ICs – Scaling, Leakage, and Other Topics

7.1 Technology Scaling - for Cost, Speed, and Power Consumption

YEAR	1992	1995	1997	1999	2001	2003	2005	2007
Technology Generation	0.5 μm	0.35 μm	0.25 μm	0.18 μm	0.13 μm	90 nm	65 nm	45 nm

•New technology node every two years or so. Defined by minimum line width-spacing average.

• Feature sizes are ~70% of previous node's.

• **Reduction of circuit area by 2 — good for cost and speed.**

International Technology Roadmap for Semiconductors

Year of Shipment	2003	2005	2007	2010	2013
Technology Node (nm)	90	65	45	32	22
Lg (nm) (HP/LSTP)	37/65	26/45	22/37	16/25	13/20
EOT_e(nm) (HP/LSTP)	1.9/2.8	1.8/2.5	1.2/1.9	0.9/1.6	0.9/1.4
VDD (HP/LSTP)	1.2/1.2	1.1/1.1	1.0/1.1	1.0/1.0	0.9/0.9
Ion,HP (μA/μm)	1100	1210	1500	1820	2200
Ioff,HP (μA/μm)	0.15	0.34	0.61	0.84	0.37
Ion,LSTP (μA/μm)	440	465	540	540	540
Ioff,LSTP (μA/μm)	1e-5	1e-5	3e-5	3e-5	2e-5

→ Strained Silicon

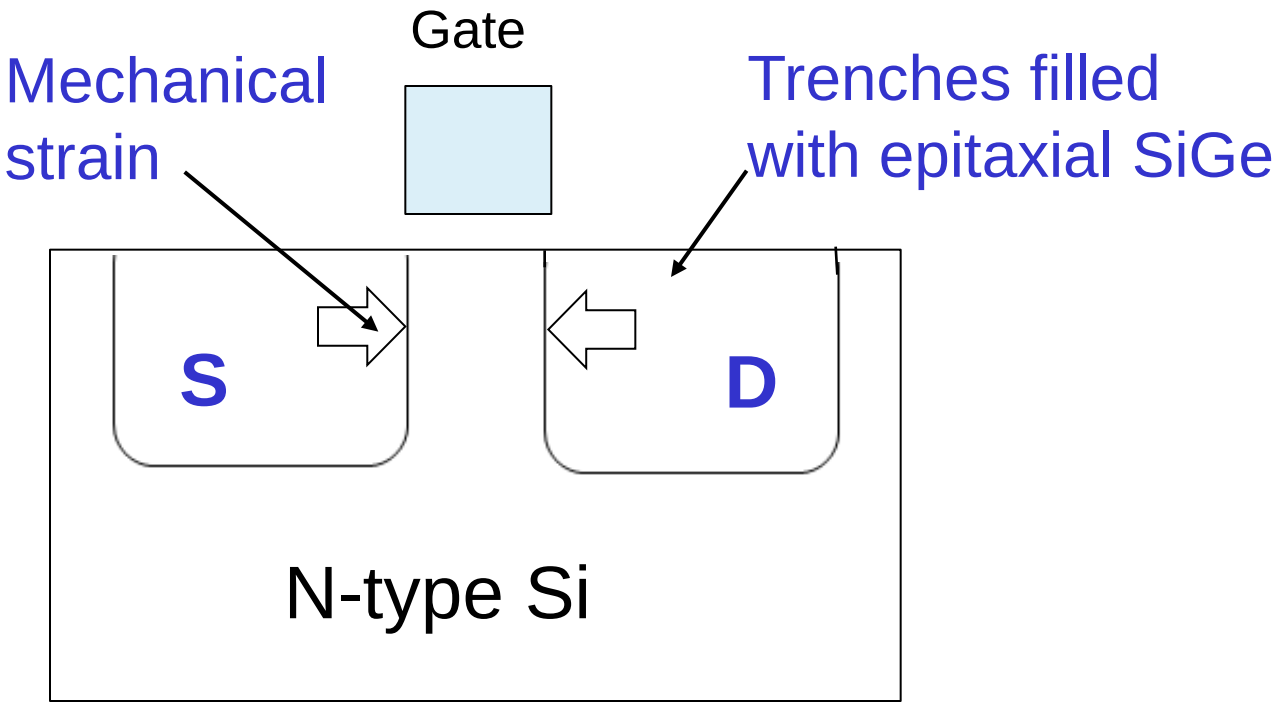
→ High-k/Metal-Gate

→ Wet Lithography

→ New Structure

- V_{dd} is reduced at each node to contain power consumption in spite of rising transistor density and frequency
- T_{ox} is reduced to raise I_{on} and retain good transistor behaviors
- HP: High performance; LSTP: Low stand-by power

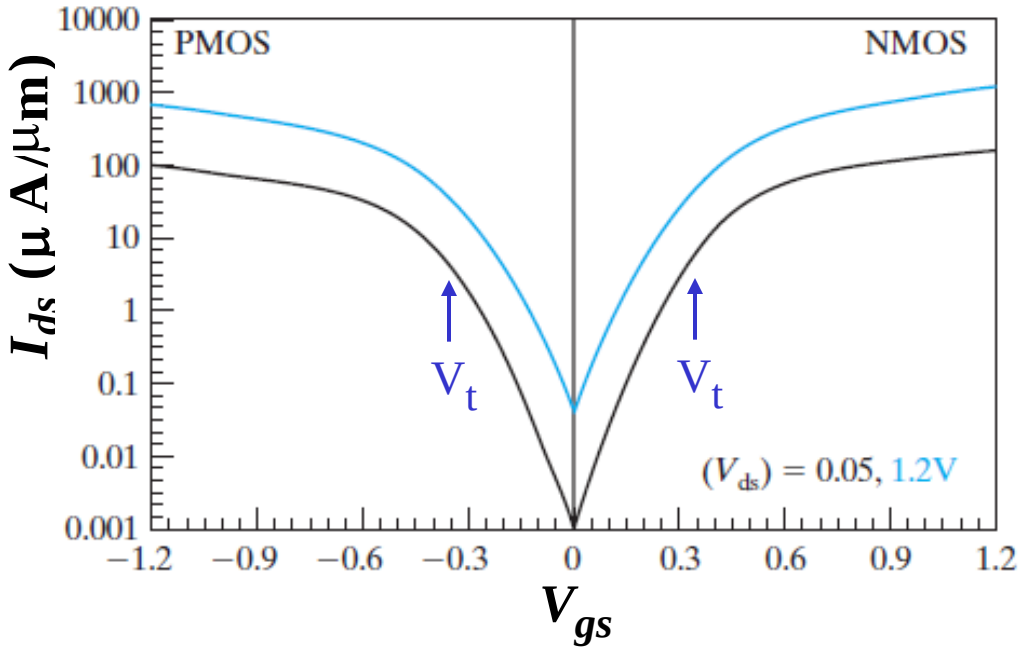
7.1.2 Strained Silicon: example of innovations



The electron and hole mobility can be raised by carefully designed mechanical strain.

7.2 Subthreshold Current

- The leakage current that flows at $V_g < V_t$ is called the subthreshold current.

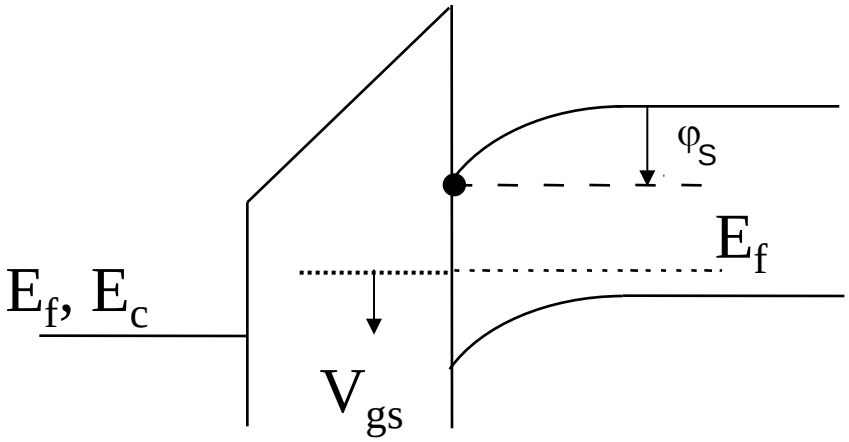


Intel, T. Ghani et al., IEDM 2003

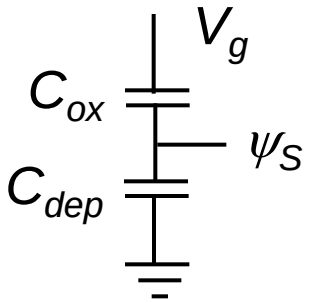
90nm technology.
Gate length: 45nm

- The current at $V_{gs}=0$ and $V_{ds}=V_{dd}$ is called I_{off}.

- Subthreshold current $\propto n_s$ (surface inversion carrier concentration)
- $n_s \propto e^{q\varphi s/kT}$



- φ_s varies with V_g through a capacitor network

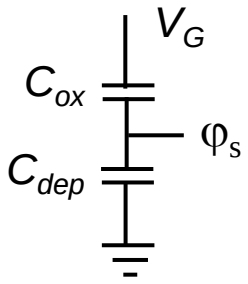


$$\frac{d\varphi_s}{dV_g} = \frac{C_{oxe}}{C_{oxe} + C_{dep}} = \frac{1}{\eta}$$

In subthreshold, $\varphi_s = constant + V_g/\eta$

Subthreshold Leakage Current

$$I_{ds} \propto n_s \propto e^{q\varphi_s / kT} \propto e^{q\left(\text{constant} + V_{gs}/\eta\right)/kT} \propto e^{qV_{gs}/\eta kT}$$



$$I_{ds} \propto e^{qV_{gs}/\eta kT}$$

$$\eta = 1 + \frac{C_{dep}}{C_{oxe}}$$

- Subthreshold current changes 10x for $\eta \cdot 60mV$ change in V_g.

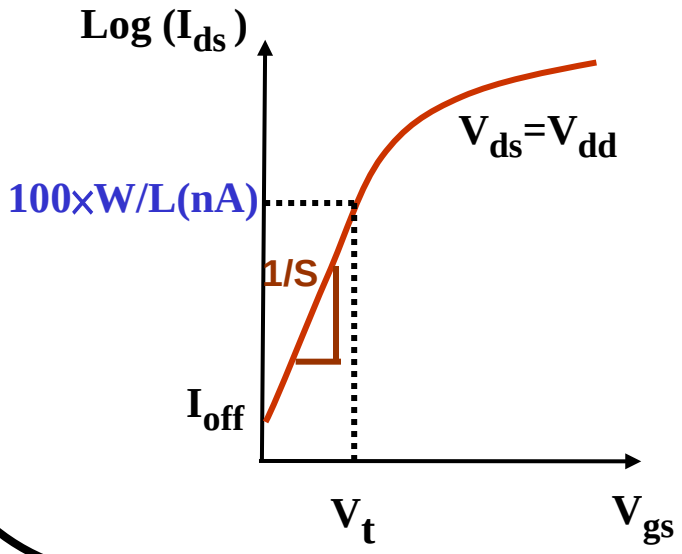
  Reminder: 60mV is (ln10)·kT/q

- **Subthreshold swing, S : the change in V_{gs} corresponding to 10x change in subthreshold current. $S = \eta \cdot 60mV$, typically 80-100mV**

Subthreshold Leakage Current

- Practical definition of V_t : the V_{gs} at which I_{ds}= 100nA×W/L

=> $$I_{subthreshold}\,(nA) \approx 100 \times \frac{W}{L} \times e^{q(V_g - V_t)/\eta kT} = 100 \times \frac{W}{L} \times 10^{(V_g - V_t)/S}$$



$$I_{off}\,(nA) = 100 \times \frac{W}{L} \times 10^{-V_t/S}$$

is determined only by V_t and subthreshold swing.

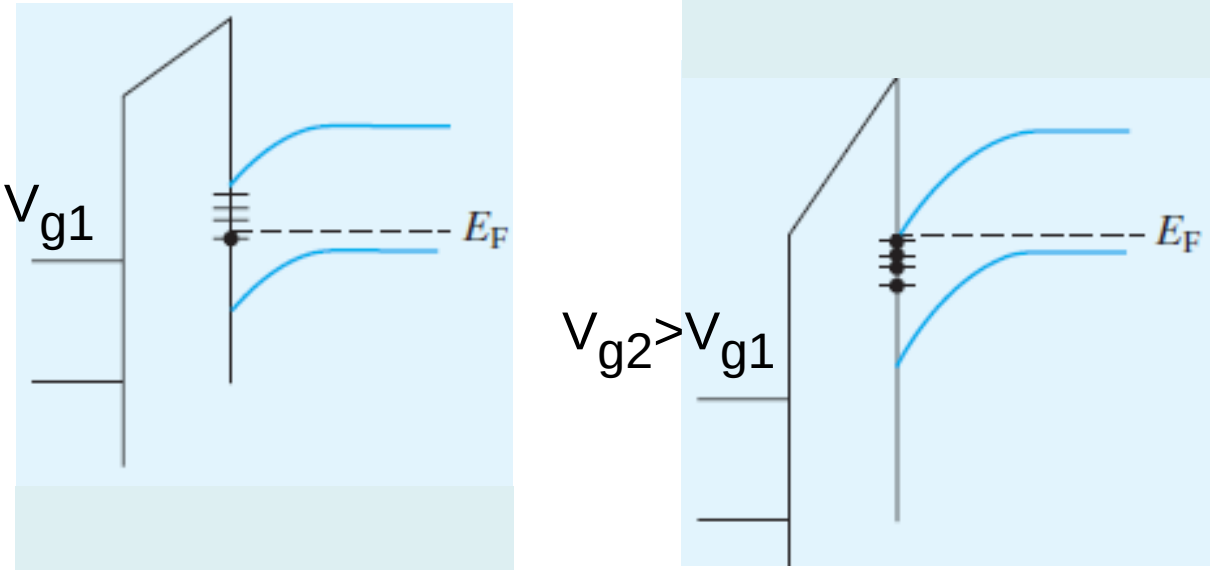
Subthreshold Swing

- Smaller S is desirable (lower I_{off} for a given V_t). Minimum possible value of S is 60mV/dec.
- How do we reduce swing?
  - Thinner T_{ox} => larger C_{oxe}
  - Lower substrate doping => smaller C_{dep}
  - Lower temperature

$$S = 60mV \cdot \left(1 + \frac{C_{dep}}{C_{oxe}}\right)$$

- Limitations
  - Thinner T_{ox} — oxide breakdown reliability or oxide leakage current
  - Lower substrate doping — doping is not a free parameter but set by V_t.

Effect of Interface States on Subthreshold Swing

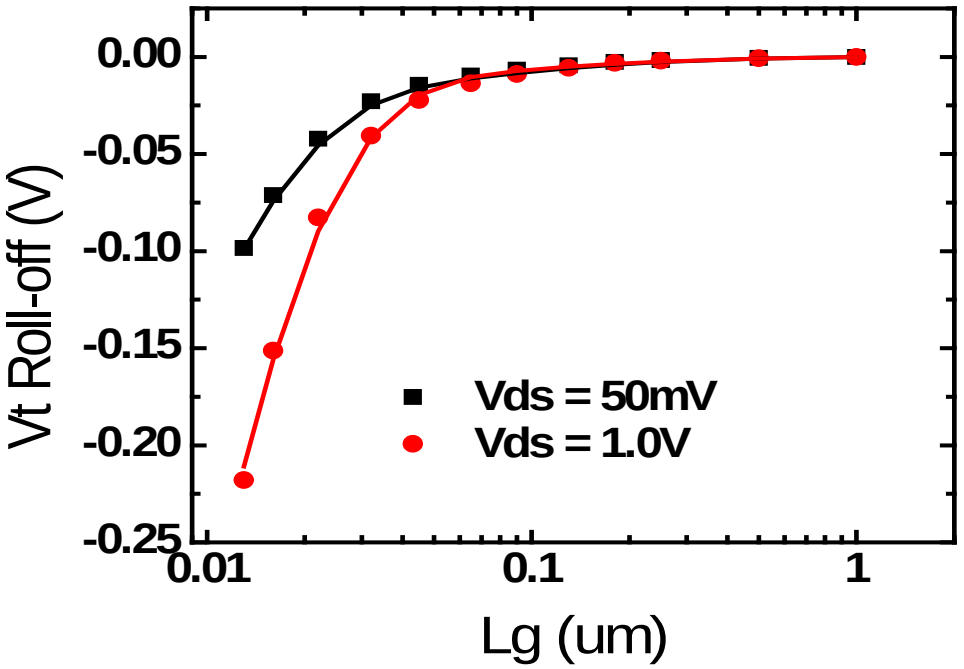


- Interface states may be filled by electrons or empty depending on its energy relative to E_F, i.e., depending on V_g.
- $dQ_{int}/d\varphi_s$ (number or interface state per eV-cm^2) presents another capacitance in parallel with C_{dep}

$$S = 60mV \cdot \left(1 + \frac{C_{dep} + dQ_{\text{int}} / d\phi_s}{C_{oxe}}\right)$$

7.3 V_t Roll-off

- V_t roll-off: V_t decreases with decreasing L_g.
- It determines the minimum acceptable L_g because I_{off} is too large if V_t becomes too small.

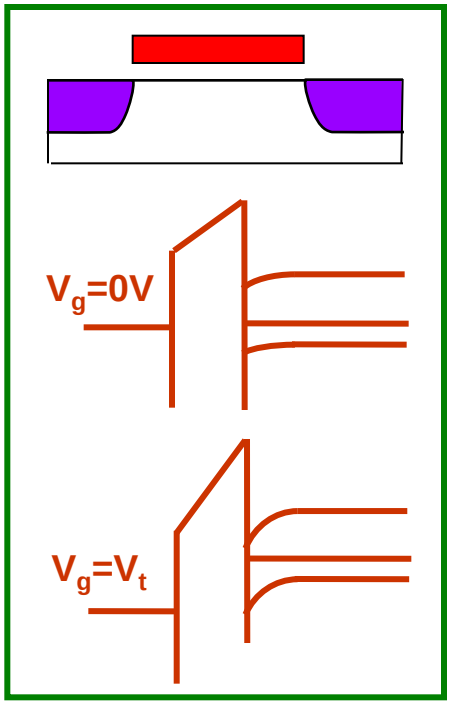


K. Goto et al., (Fujitsu) IEDM 2003 65nm technology. EOT=1.2nm, V_{dd}=1V

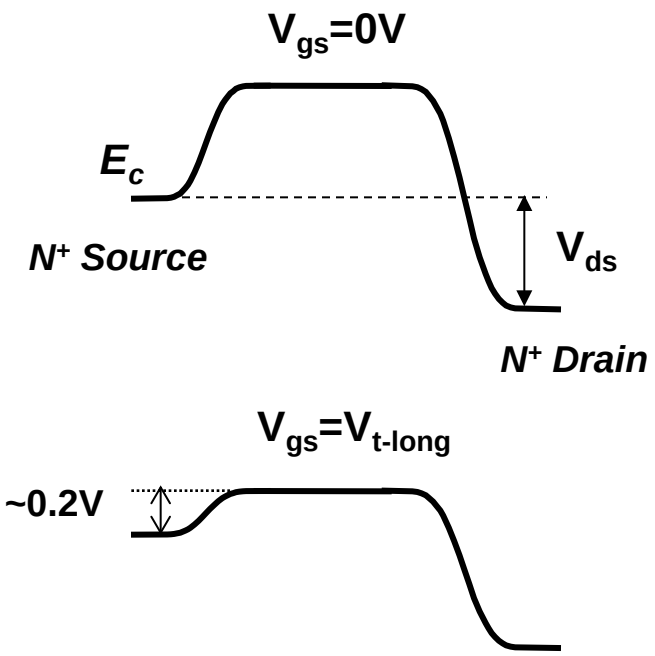
- ***Question: Why data is plotted against L_g, not L?***

  Answer: L is difficult to measure. L_g is. Also, L_g is the quantity that manufacturing engineers can control directly.

Why Does V_t Decrease with L?— Potential Barrier Concept



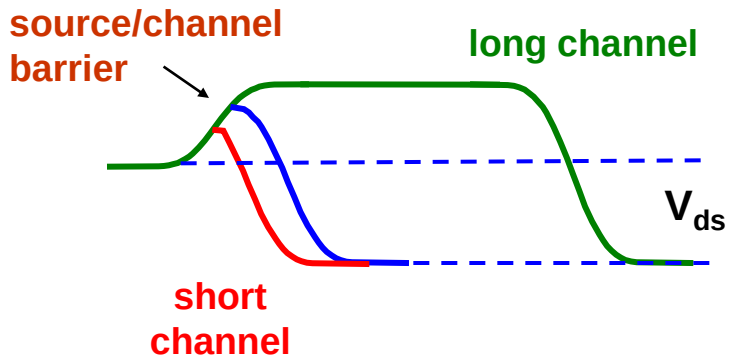
Long Channel



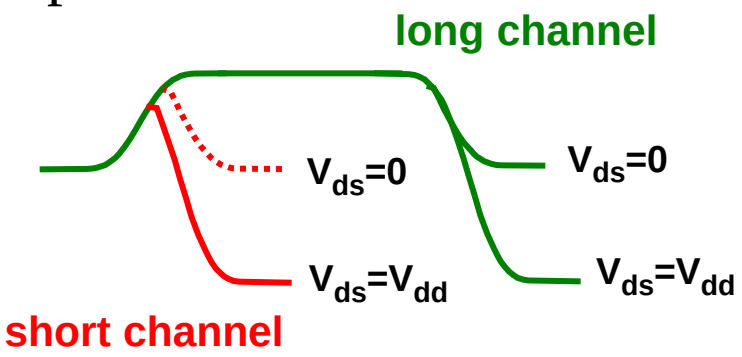
- When L is small, smaller V_g is needed to reduce the barrier to 0.2V, i.e. V_t is smaller.
- V_t roll-off is greater for shorter L

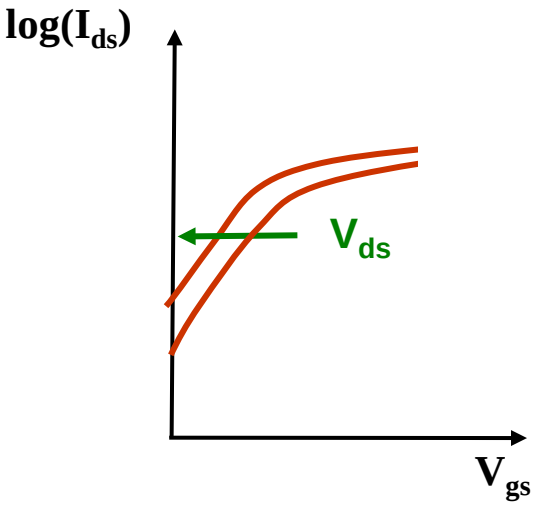
Energy-Band Diagram from Source to Drain

- L dependence

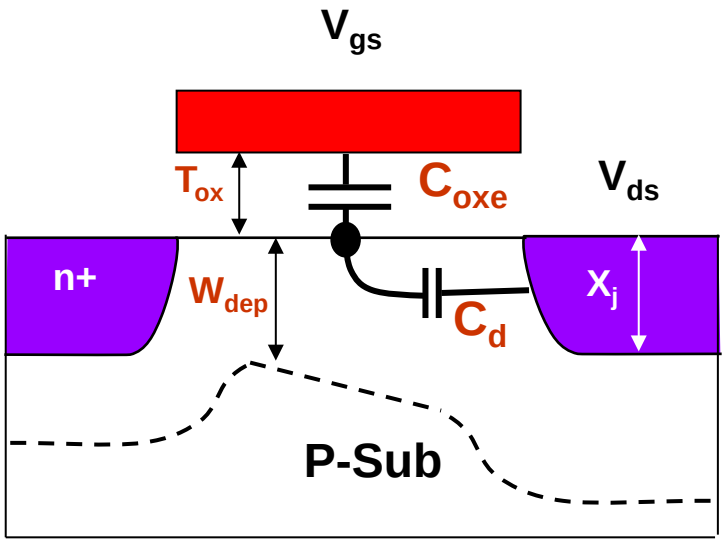


- V_{ds} dependence





V_t Roll-off – Simple Capacitance Model



V_{ds} helps V_{gs} to invert the surface, therefore

$$V_t = V_{t-long} - V_{ds} \cdot \frac{C_d}{C_{oxe}}$$

$$V_t = V_{t-long} - (V_{ds} + 0.4) \cdot \frac{C_d}{C_{oxe}}$$

Due to built-in potential between N^- channel and N^+ drain & source

As the channel length is reduced, drain to channel distance is reduced→ C_d increases

- 2-D Poisson Eq. solution shows that C_d is an exponential function of L.

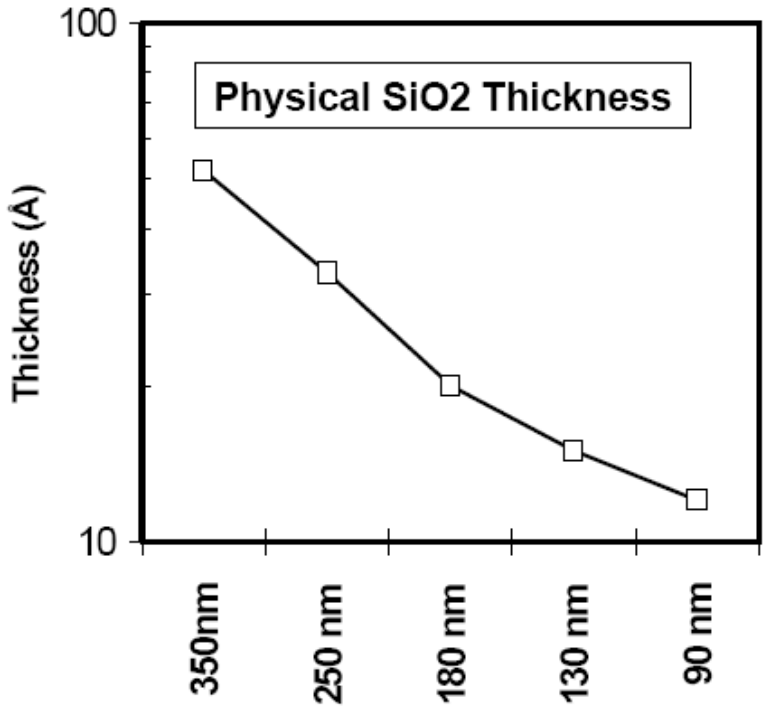
$$V_t = V_{t-long} - (V_{ds} + 0.4) \cdot e^{-L/l_d}$$

$$\text{where } l_d \approx \sqrt[3]{T_{ox} W_{dep} X_j}$$

- **Vertical dimensions (T_{ox}, W_{dep}, X_j) must be scaled to support L reduction**

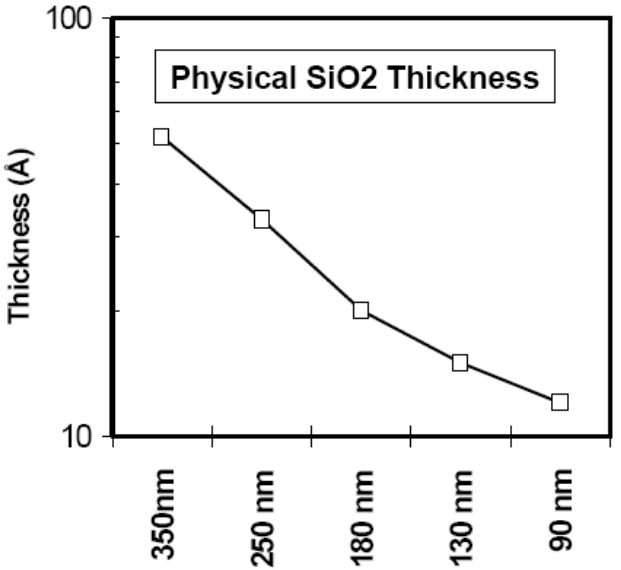
$$V_t = V_{t-long} - (V_{ds} + 0.4) \cdot e^{-L/l_d}$$

where $l_d \approx \sqrt[3]{T_{ox} W_{dep} X_j}$

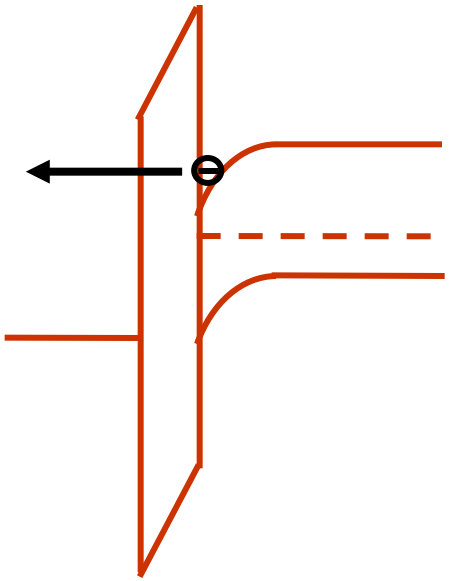


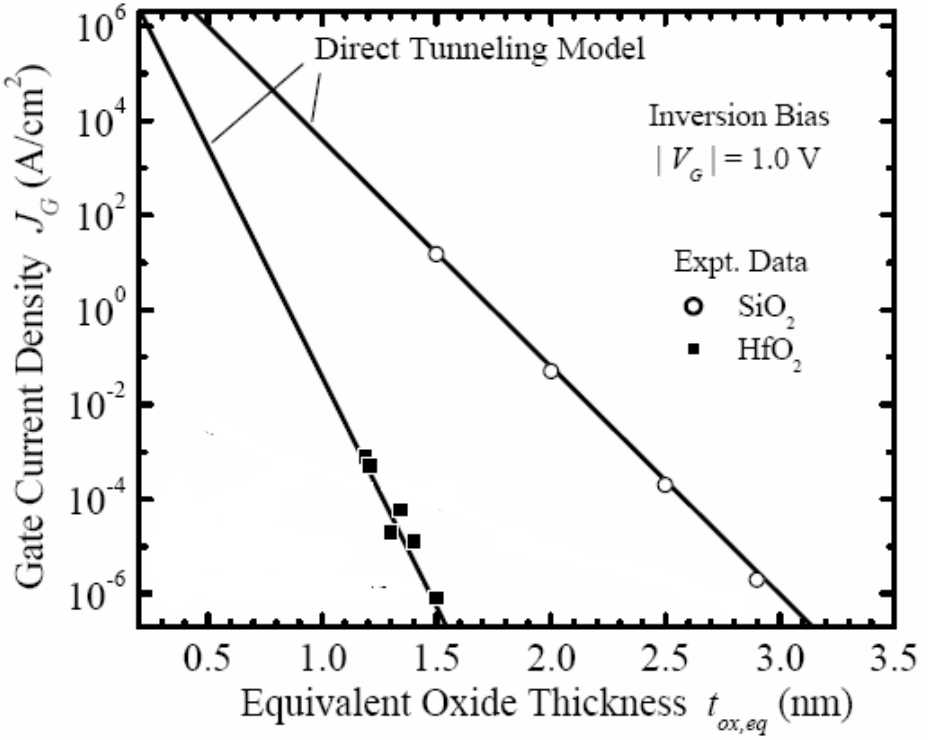
7.4 Reducing Gate-Insulator Electrical Thickness and Tunneling Leakage

- Oxide thickness has been reduced over the years from 300nm to 1.2nm.
- Why reduce oxide thickness?
  - Larger C_{ox} to raise I_{on}
  - Reduce subthreshold swing
  - Control V_t roll-off
- Thinner is better. However, if the oxide is too thin
  - Breakdown due to high field
  - Leakage current



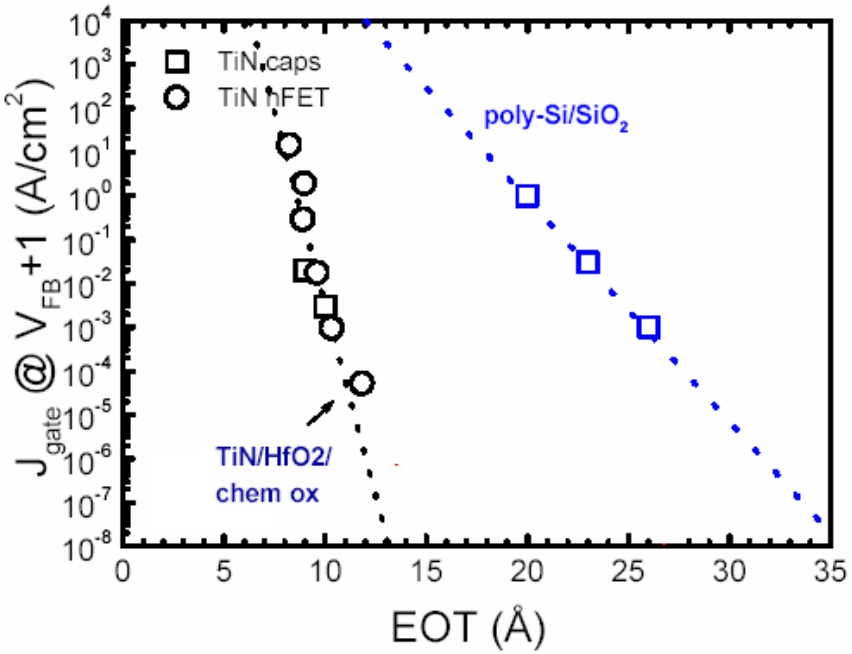
Gate Tunneling Leakage Current





- For SiO_2 films thinner than 1.5nm, tunneling leakage current has become the limiting factor.
- HfO_2 has several orders lower leakage for the same EOT.

Replacing SiO_2 with HfO_2---High-k Dielectric



(After W. Tsai et al., IEDM'03)

- HfO_2 has a relative dielectric constant (k) of ~24, six times large than that of SiO_2.
- For the same EOT, the HfO_2 film presents a much thicker (albeit a lower) tunneling barrier to the electrons and holes.
- Toxe can be further reduced by introducing metal-gate technology since the poly-depletion effect is eliminated.

Challenges of High-K Technology

- The difficulties of high-k dielectrics:
  - chemical reactions between them and the silicon substrate and gate,
  - lower surface mobility than the Si/SiO_2 system
  - too low a V_t for P-channel MOSFET (as if there is positive charge in the high-k dielectric).
  - long-term reliability

- A thin SiO_2 interfacial layer may be inserted between Si-substrate and high-k film.

Question: How can T_{inv} be reduced?

(Answer is in Sec. 7.4 text)

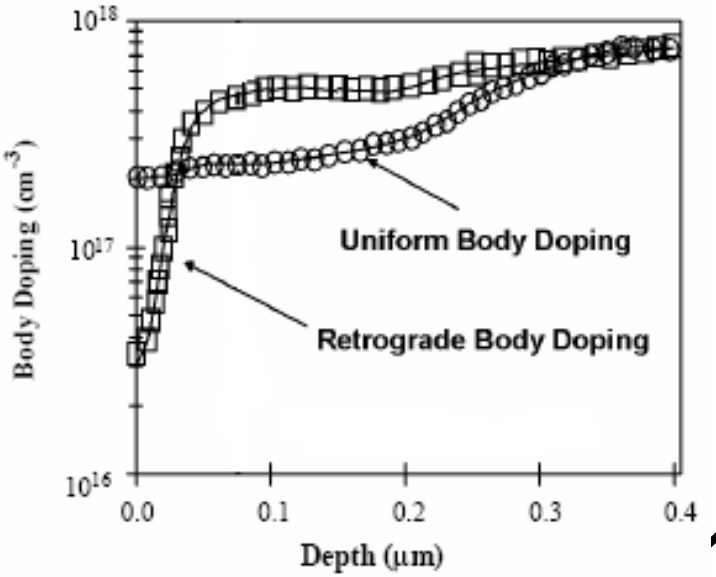
7.5 How to Reduce W_{dep}

- W_{dep} can be reduced by increasing N_{sub}

$$V_t = V_{fb} + \phi_{st} + \frac{\sqrt{qN_{sub}2\varepsilon_s\phi_{st}}}{C_{ox}} = V_{fb} + \phi_{st} + \frac{2\varepsilon_s\phi_{st}}{C_{ox}W_{dep}}$$

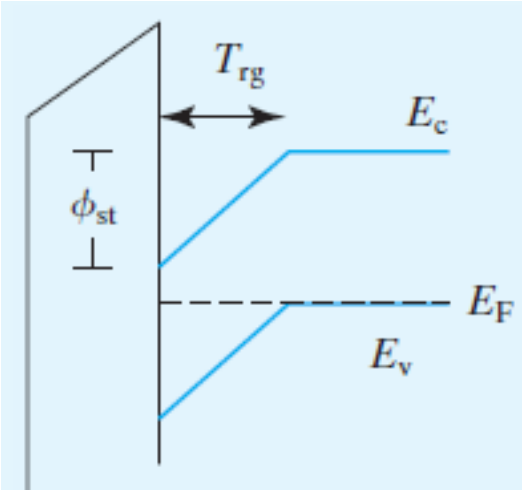
  - If N_{sub} is increased, C_{ox} has to be increased in order to keep V_t the same.
  - W_{dep} can be reduced in proportion to T_{ox}.

- Or use retrograde doping with very thin lightly doped surface layer
  - Also, less impurity scattering in the inversion layer → higher mobility



7.5 Ideal Retrograde Doping Profile

- Assume the body is heavily doped with an undoped layer, T_{rg} thick, at the surface.



$$V_{ox} = T_{ox}\mathscr{E}_{ox} = \phi_{st}\frac{\varepsilon_s T_{ox}}{\varepsilon_{ox} T_{rg}}$$

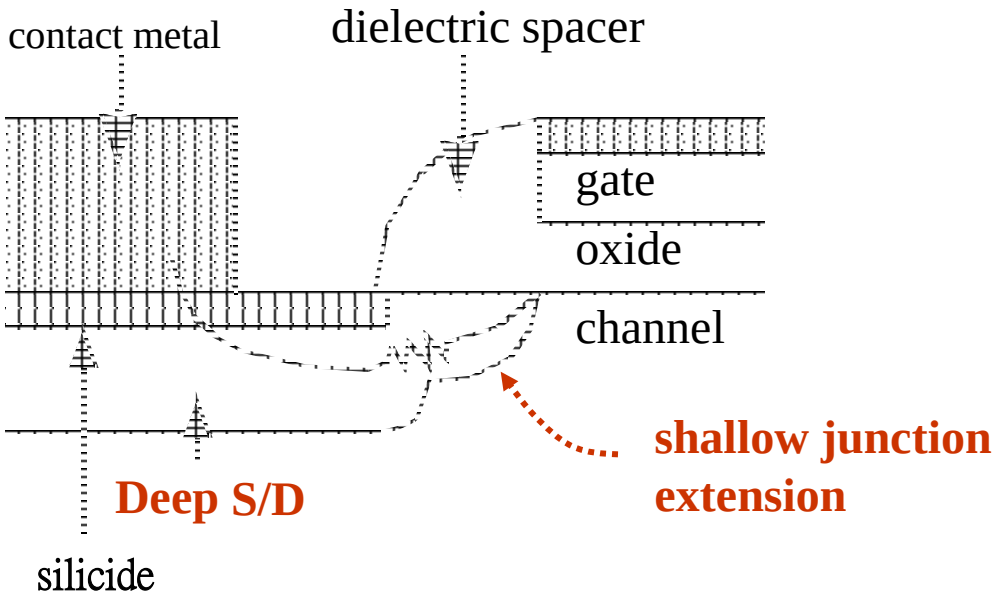
$$V_t = V_{fb} + \phi_{st}\left(1 + \frac{\varepsilon_s T_{ox}}{\varepsilon_{ox} T_{rg}}\right)$$

- Compared with uniformly doped body

$$V_t = V_{fb} + \phi_{st}(1 + \frac{2\varepsilon_s T_{ox}}{\varepsilon_{ox} W_{dep}})$$

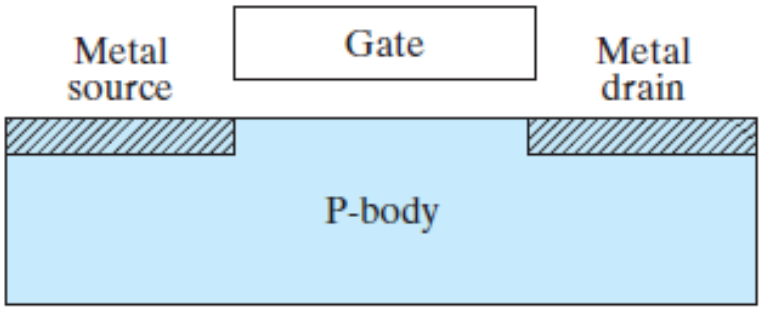
•Ideal retrograde doping yields a depletion region width (T_{rg}) half as thick as W_{dep} of a uniform doped body.

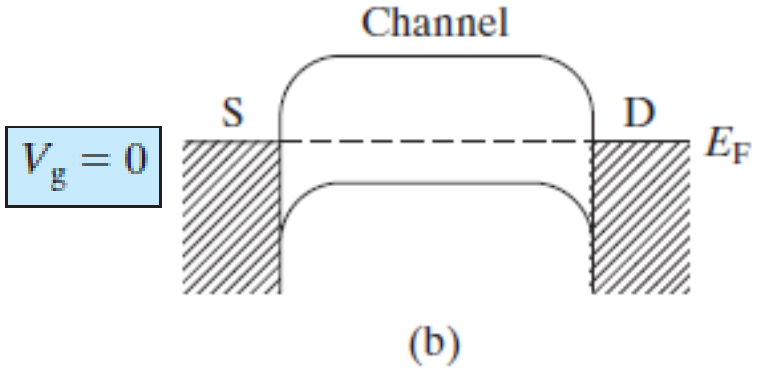
7.6 Shallow Junction and Metal Source/Drain



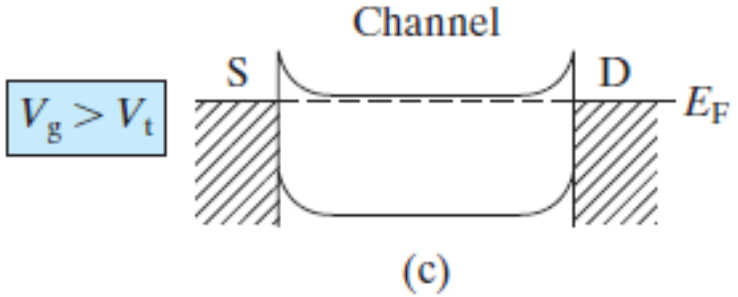
- The shallow junction extension helps to control V_t roll-off.
- Shallow junction and light doping combine to produce an undesirable parasitic resistance that reduces the precious $I_{on.}$
- Theoretically, metal S/D can be used as a very shallow "junction".

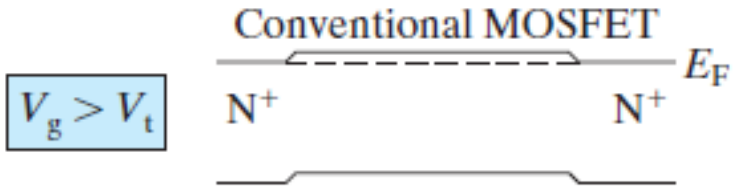
7.6.1 MOSFET with Metal Source/Drain



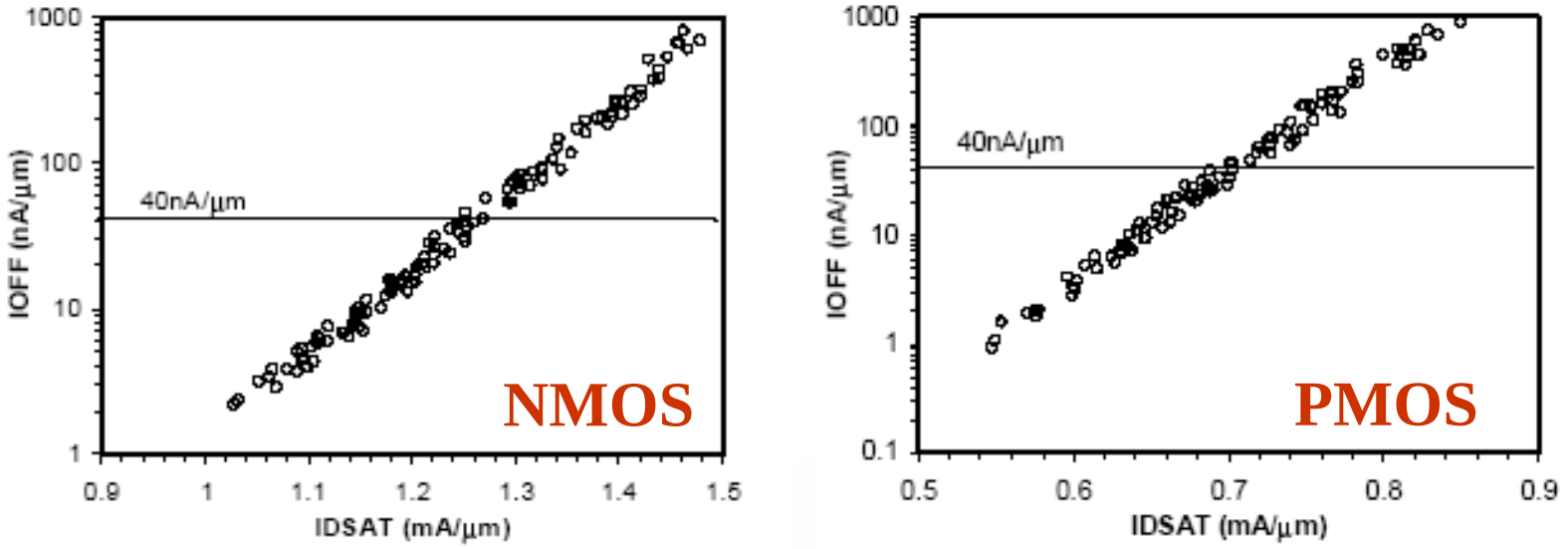


To unleash the potentials of Schottky S/D MOSFET, a low-ϕ_{Bn} Schottky junction is needed for NFETs and low-ϕ_{Bp} for PFET.





7.7 Variations and Design for Manufacturing



Intel, T. Ghani et al., IEDM 2003

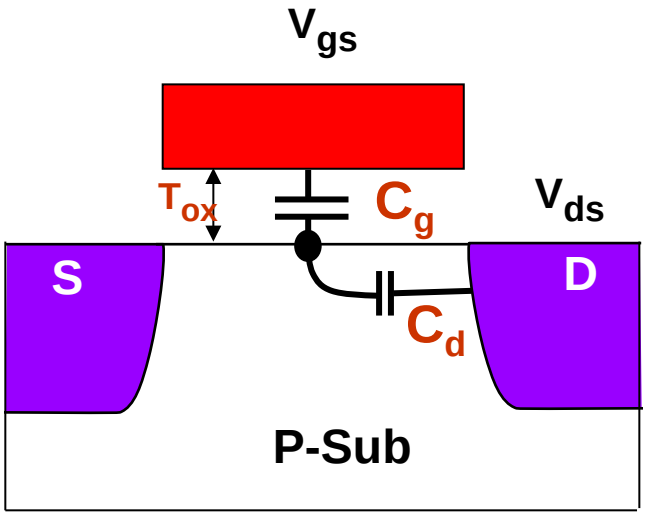
- Higher I_{on} goes hand-in-hand with larger I_{off} -- think L, V_t, T_{ox}, V_{dd}.
- Figure shows spread in I_{on} (and I_{off}) produced by intentional difference in L_g and unintentional manufacturing variatons in L_g and other parameters.

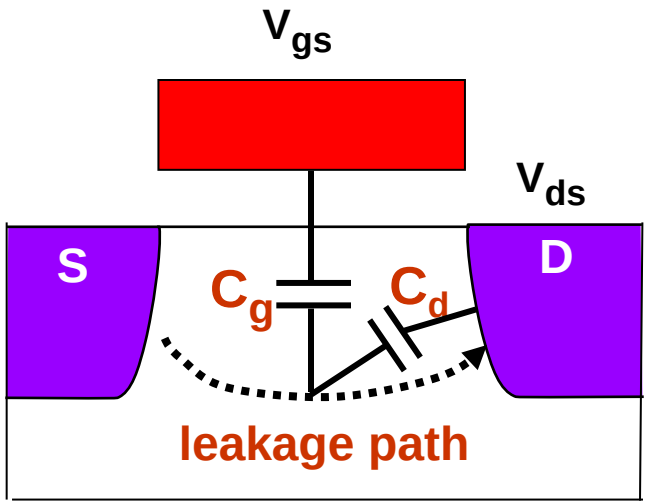
Variation Tolerant Circuit Design

- Multiple V_t
  - Lower V_t is used only in the blocks that need speed
- Multiple V_{dd}
  - Higher V_{dd} is used only in the blocks that need speed
- Substrate (well) bias
  - Only some circuit blocks need to operate at high speed.
  - Can use reverse well bias to raise the V_t for the rest.
  - This techniques can also reduce the chip-to-chip and block-to-block variations with intelligent control circuitry.
  - Would like larger body effect than conventional MOSFET.

7.8 Ultra-Thin-Body SOI and Multigate MOSFETs

- Reducing T_{ox} gives the gate excellent control of Si surface potential.
- But, the drain could still have more control than the gate along sub-surface leakage current paths. (Right figure.)

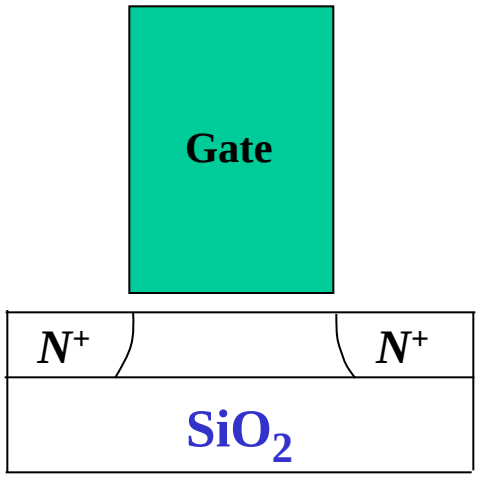


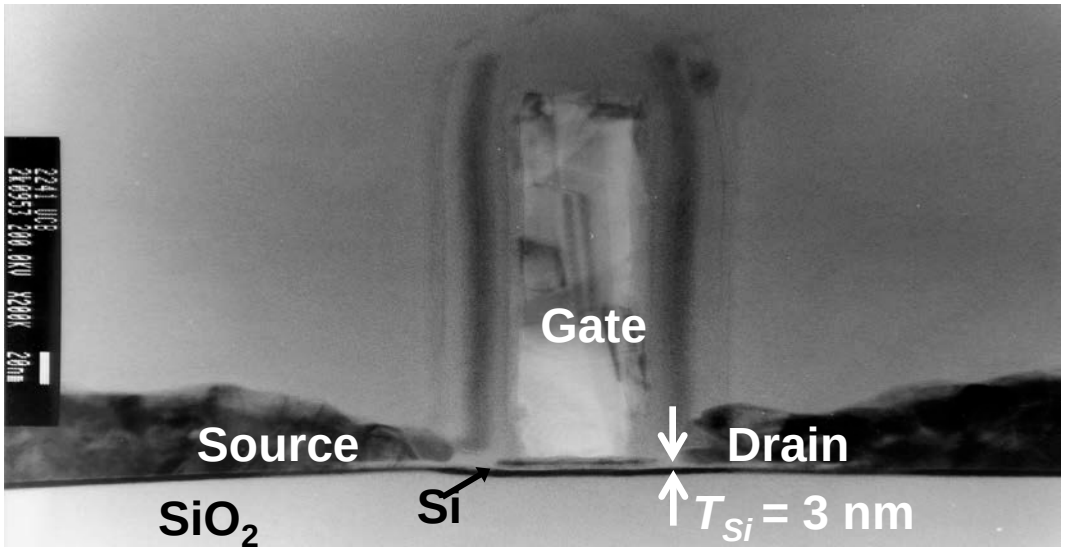


7.8.1 Ultra-Thin-Body MOSFET and SOI

- UTB MOSFET built on ultra thin silicon film on an insulator (SiO_2).
- Since the silicon film is very thin, perhaps less than 10nm, no leakage path is very far from the gate.

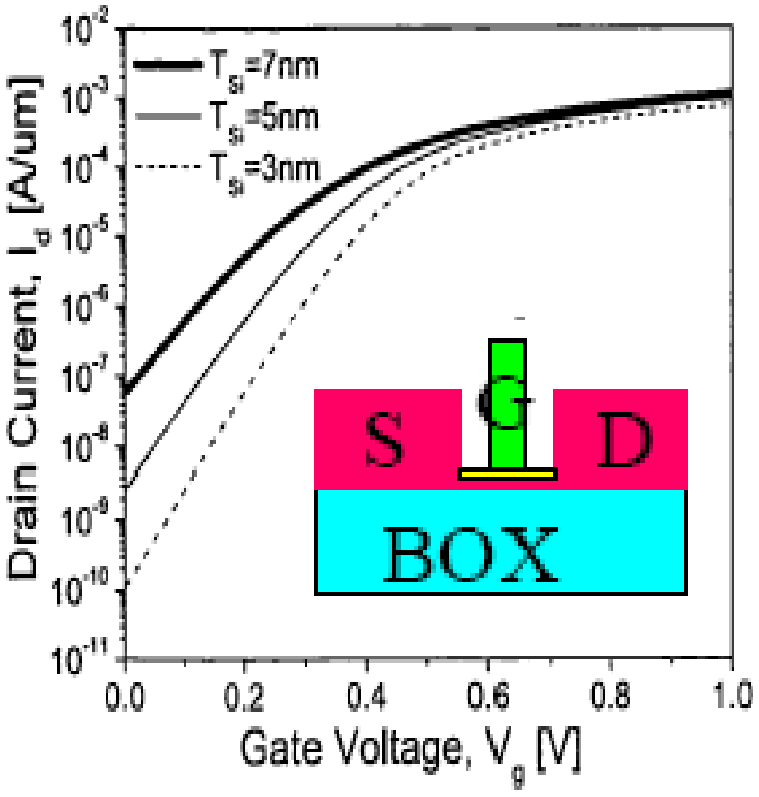
Electron Micrograph of UTB MOSFET



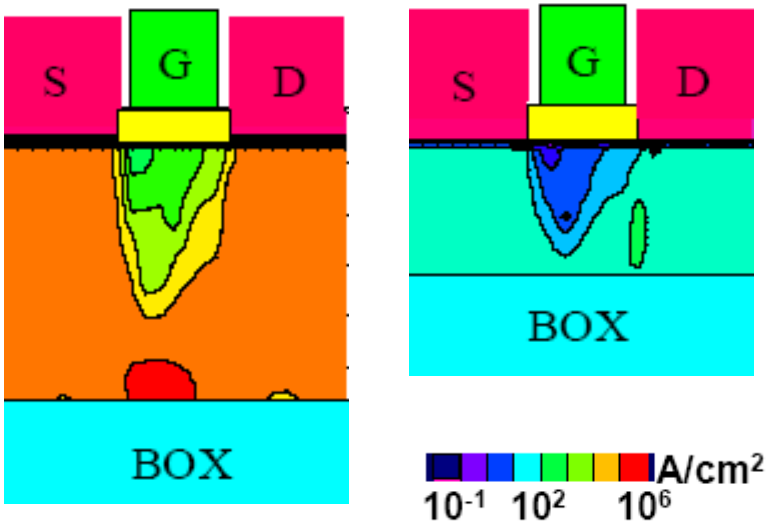


Ultra-Thin-Body MOSFET

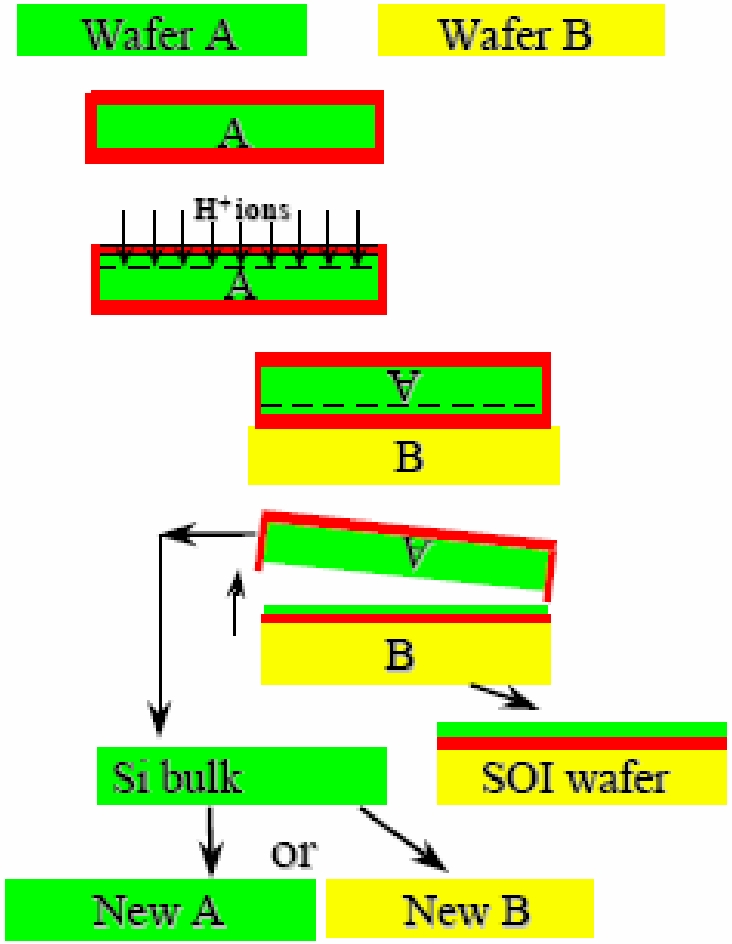
- The subthreshold leakage is reduced as the silicon film is made thinner.



T_{ox}=1.5nm, N_{sub}=1e15cm^{-3},
V_{dd}=1V, V_{gs}=0

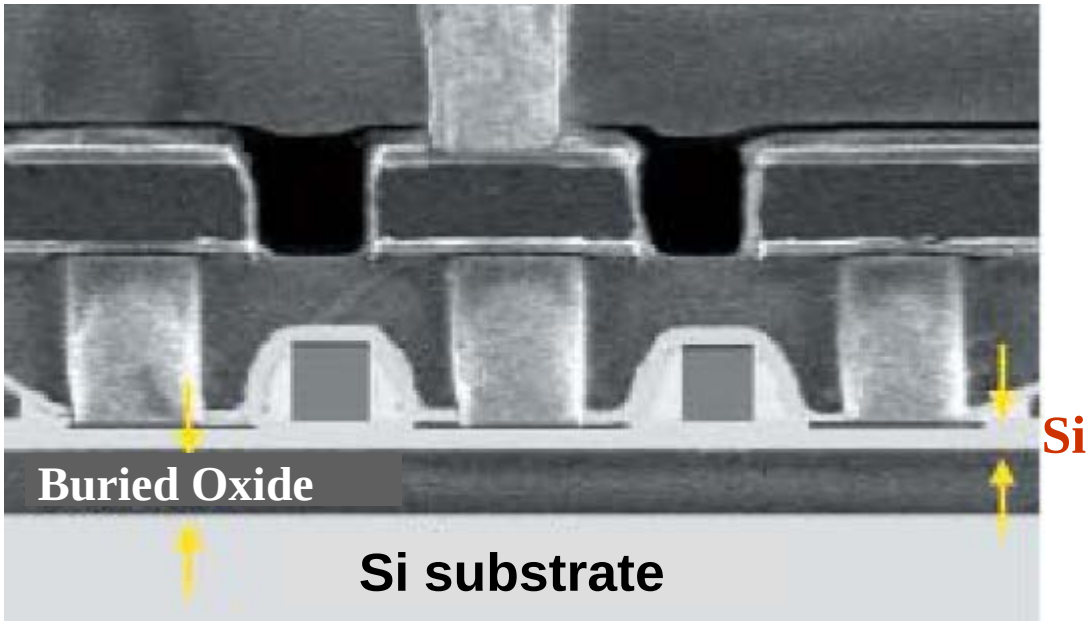


Producing Silicon-on-Insulator (SOI) Substrates



- Initial Silicon wafer A and B
- Oxidize wafer A to grow SiO2
- Implant hydrogen into wafer A
- Place wafer A, upside down, over wafer B.
- A low temperature annealing causes the two wafers to fuse together.
- Apply another annealing step to for H_2 bubbles and split wafer A.
- Polish the surface and the SOI wafer is ready for use.
- Wafer A can be reused.

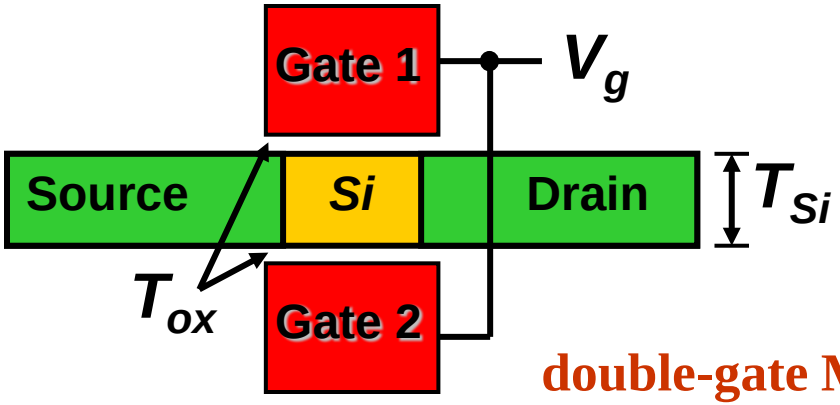
Cross-Section of SOI Circuits



- Due to the high cost of SOI wafers, only some microprocessors, which command high prices and compete on speed, have embraced this technology.
- In order to benefit from the UTB concept, Si film thickness must be agreesively reduced to ~ Lg/4

7.8.2 Multi-gate MOSFET and FinFET

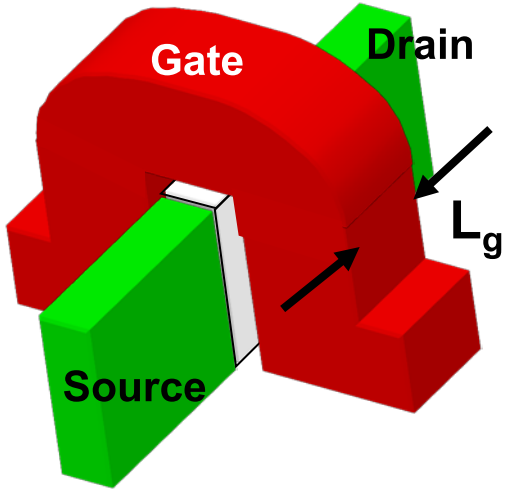
- The second way of eliminating deep leakage paths is to provide gate control from more than one side of the channel.
- The Si film is very thin so that no leakage path is far from one of the gates.
- Because there are more than one gates, the structure may be called **multi-gate MOSFET**.

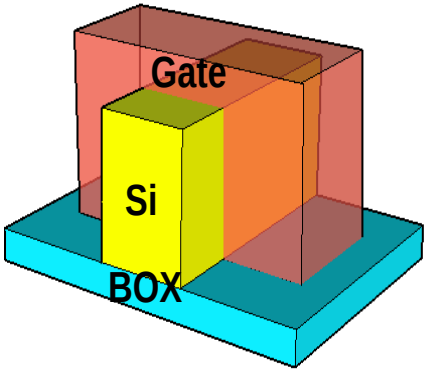


FinFET

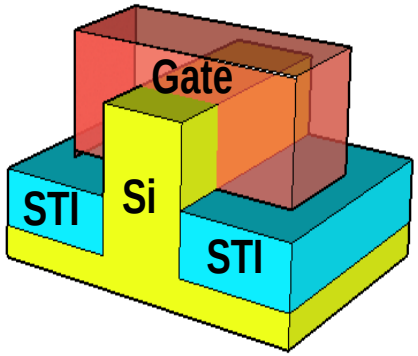
- One multi-gate structure, called **FinFET,** is particularly attractive for its simplicity of fabrication.
- The channel consists of the two vertical surfaces and the top surface of the fin.
- Question: What is the channel width, W?

Answer: The sum of twice the fin height and the width of the fin.



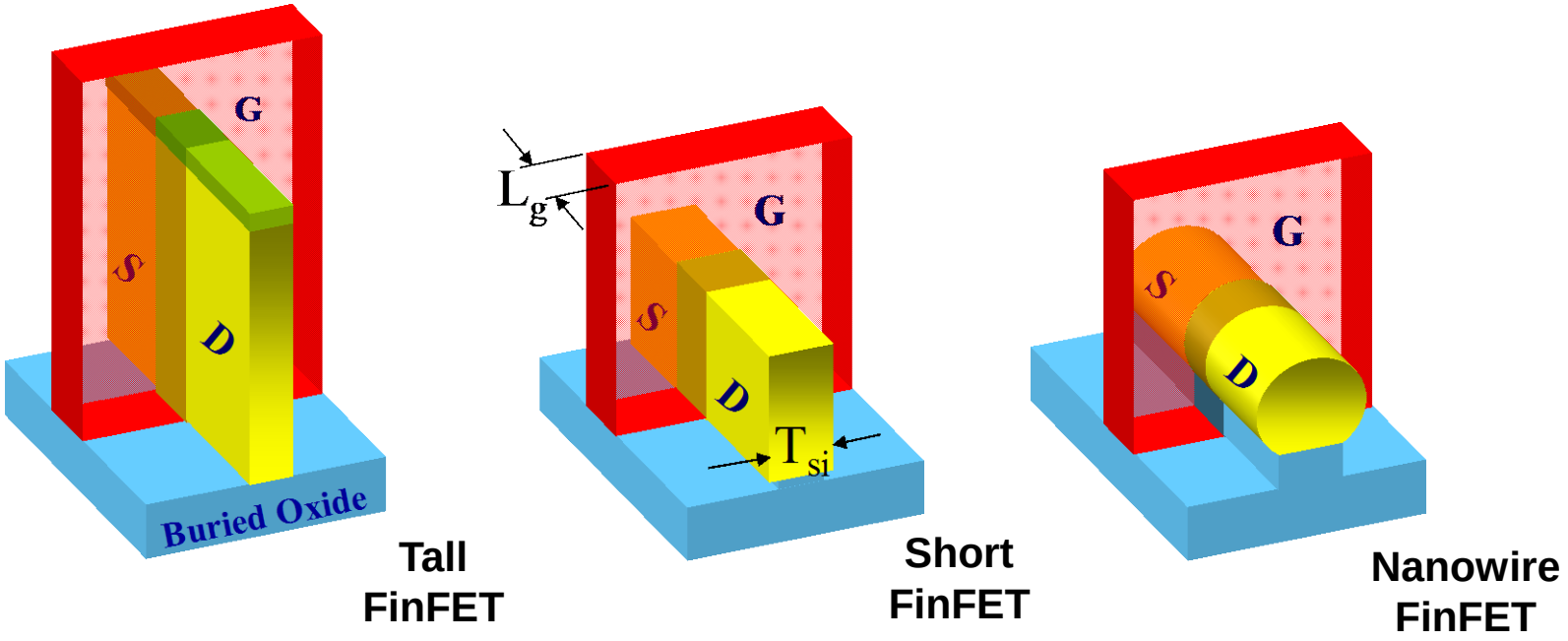


SOI FinFET



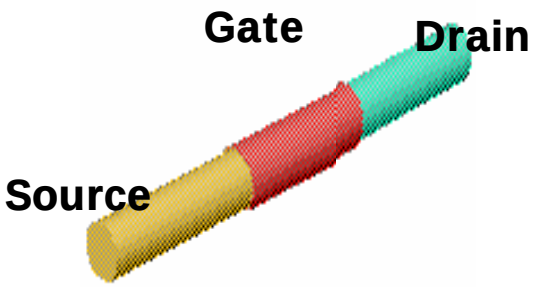
Bulk FinFET

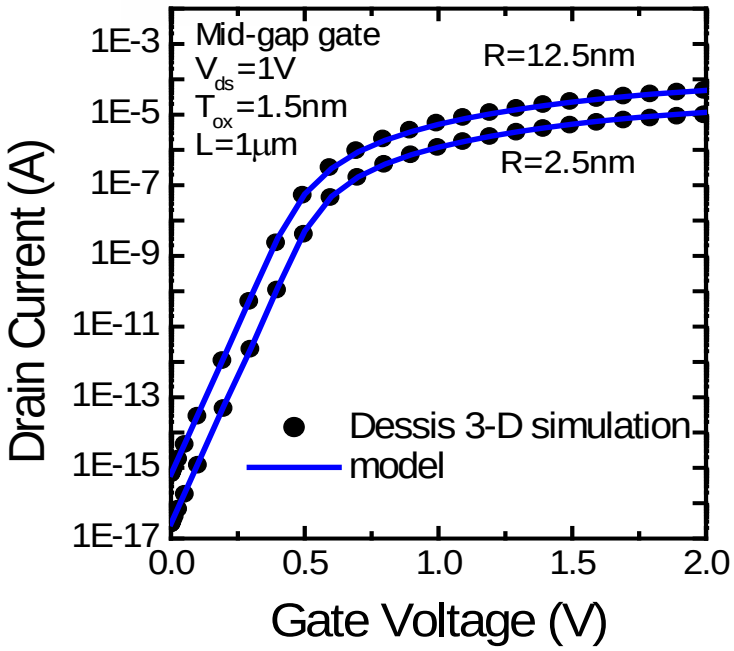
Variations of FinFET

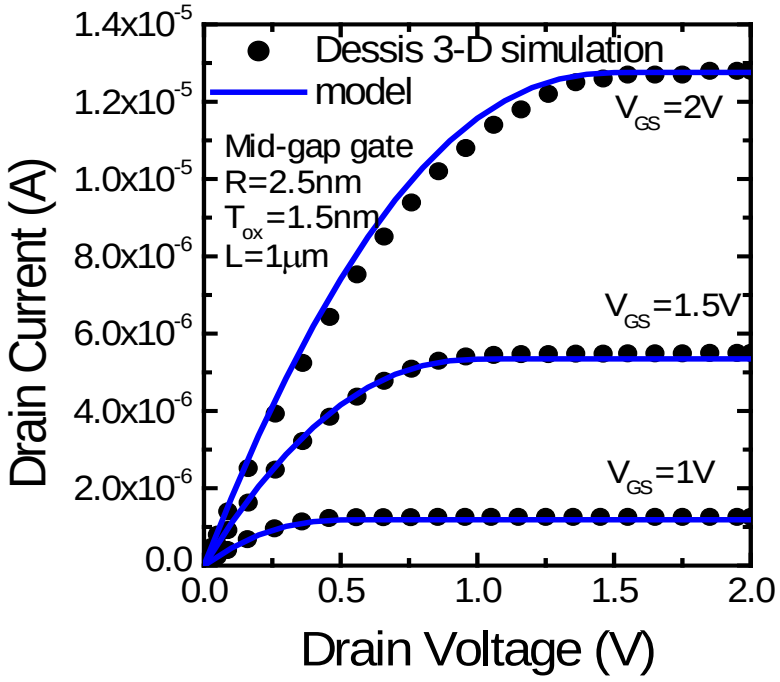


- **Tall FinFET** has the advantage of providing a large W and therefore large I_{on} while occupying a small footprint.
- **Short FinFET** has the advantage of less challenging lithography and etching.
- **Nanowire FinFET** gives the gate even more control over the silicon wire by surrounding it.

I-V of a Nanowire "Multi-Gate" MOSFET







7.9 Output Conductance

What Parameters Determine the g_{ds} ?

$$g_{ds} \equiv \frac{dI_{dsat}}{dV_{ds}} = \frac{dI_{dsat}}{dV_t} \cdot \frac{dV_t}{dV_{ds}}$$

$$\frac{dI_{dsat}}{dV_t} = \frac{-dI_{dsat}}{dV_{gs}} = -g_{msat} \quad \textbf{and} \quad \frac{dV_T}{dV_{ds}} = e^{-L/l_d}$$

I_{dsat} *is a function of* V_{gs}-V_t

(From Eq. 7.3.3, $V_t = V_{t-long} - V_{ds} \cdot e^{L/l_d}$*)*

$$\Rightarrow \quad g_{ds} = g_{msat} \times e^{-L/l_d}$$

$$\Rightarrow \quad \text{Max voltage gain } (R \to \infty) = \frac{g_{msat}}{g_{ds}} = e^{L/l_d}$$

- A larger L or smaller l_d , i.e. smaller Tox, Wdep, Xj, can increase the maximum voltage gain.
- The cause is "V_t dependence on V_{ds}"in short channel transistors.

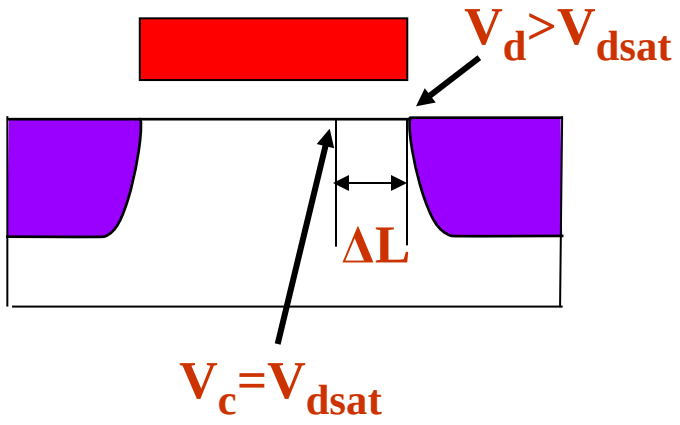
Channel Length Modulation

• For large L and V_{ds} close to V_{dsat}, another mechanism may dominate g_{ds}. That is **channel length modulation**.

•V_{ds}-V_{dsat}, is dissipated over a short distance next to drain, causing the "channel length" to decrease. More with increasing V_{ds}.

$$g_{ds} = \frac{l_d \cdot I_{dsat}}{L(V_{ds} - V_{dsat})}$$

$$l_d \approx \sqrt[3]{T_{ox} W_{dep} X_j}$$

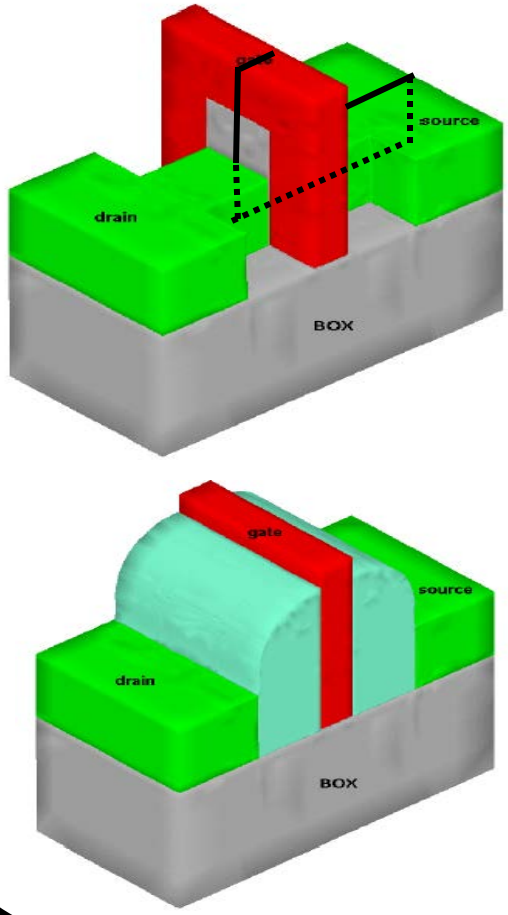


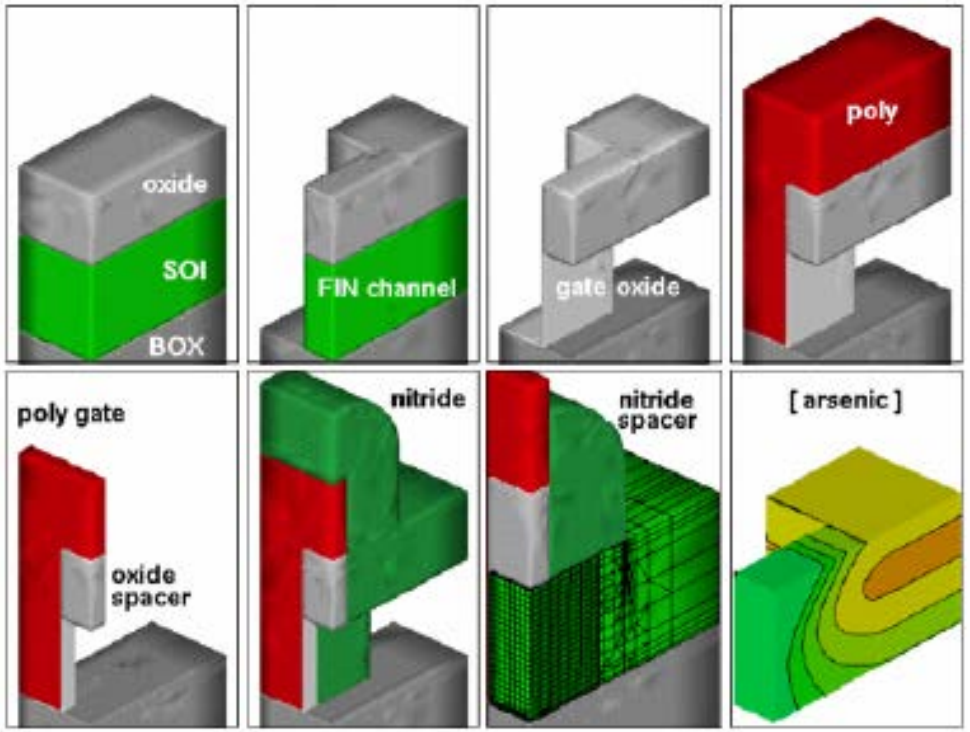
7.10 Device and Process Simulation

- **Device Simulation**
  - Commercially available computer simulation tools can solve all the equations presented in this book simultaneously with few or no approximations.
  - Device simulation provides quick feedback about device design before long and expensive fabrication.
- **Process Simulation**
  - Inputs to process simulation: lithography mask pattern, implantation dose and energy, temperatures and times for oxidization and annealing steps, etc.
  - The process simulator generates a 2-D or 3-D structures with all the deposited or grown and etched thin films and doped regions.
  - This output may be fed into a device simulator as input together with applied voltages.

Example of Process Simulation

- **FinFET Process**



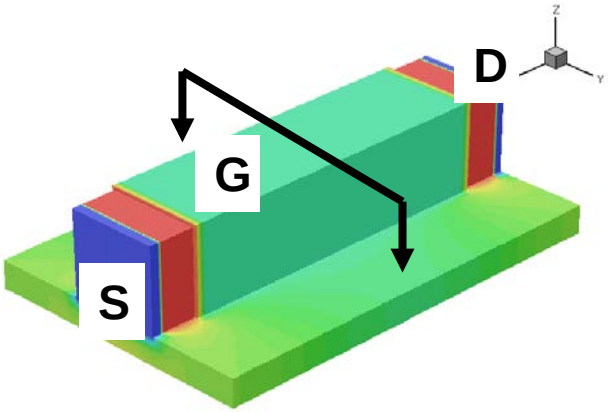


The small figures only show 1/4 of the complete FinFET--the quarter farthest from the viewer.

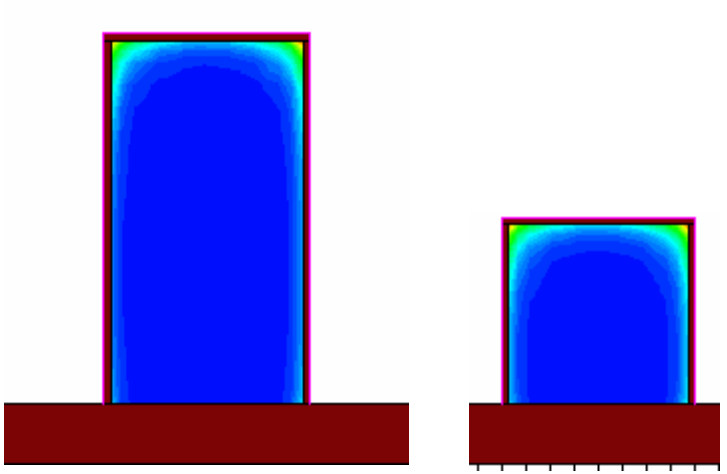
Manual, Taurus Process, Synoposys Inc.

Example of Device Simulation---

Density of Inversion Charge in the Cross-Section of a FinFET Body



C.-H. Lin et al., 2005 SRC TECHCON



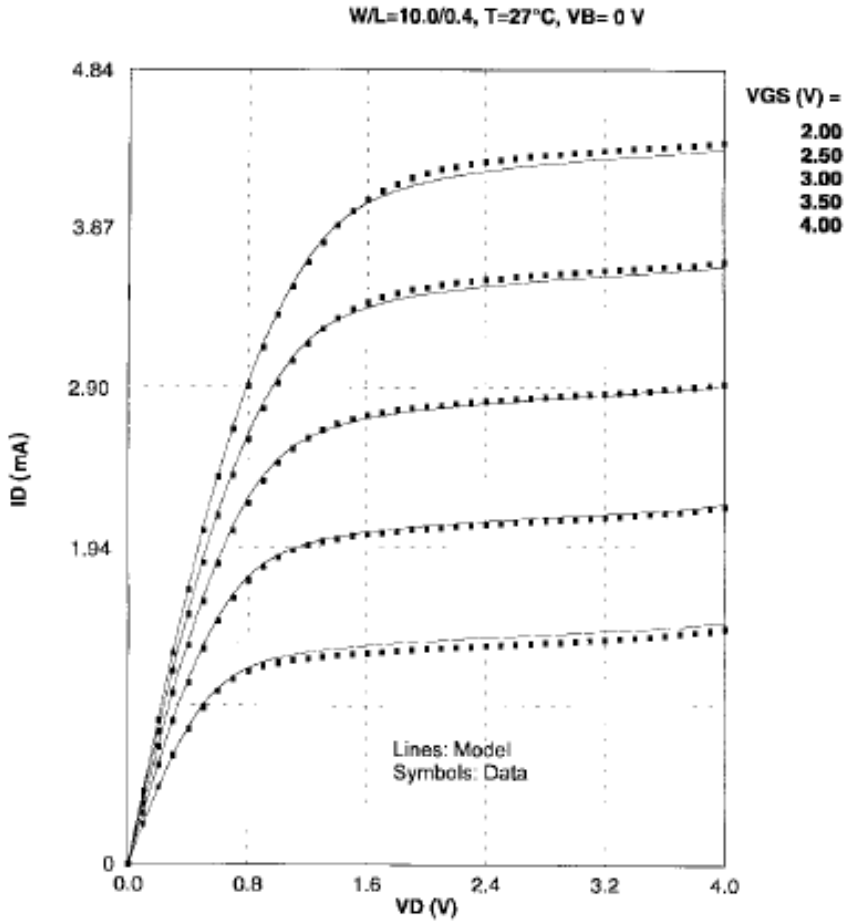
Tall FinFET **Short FinFET**

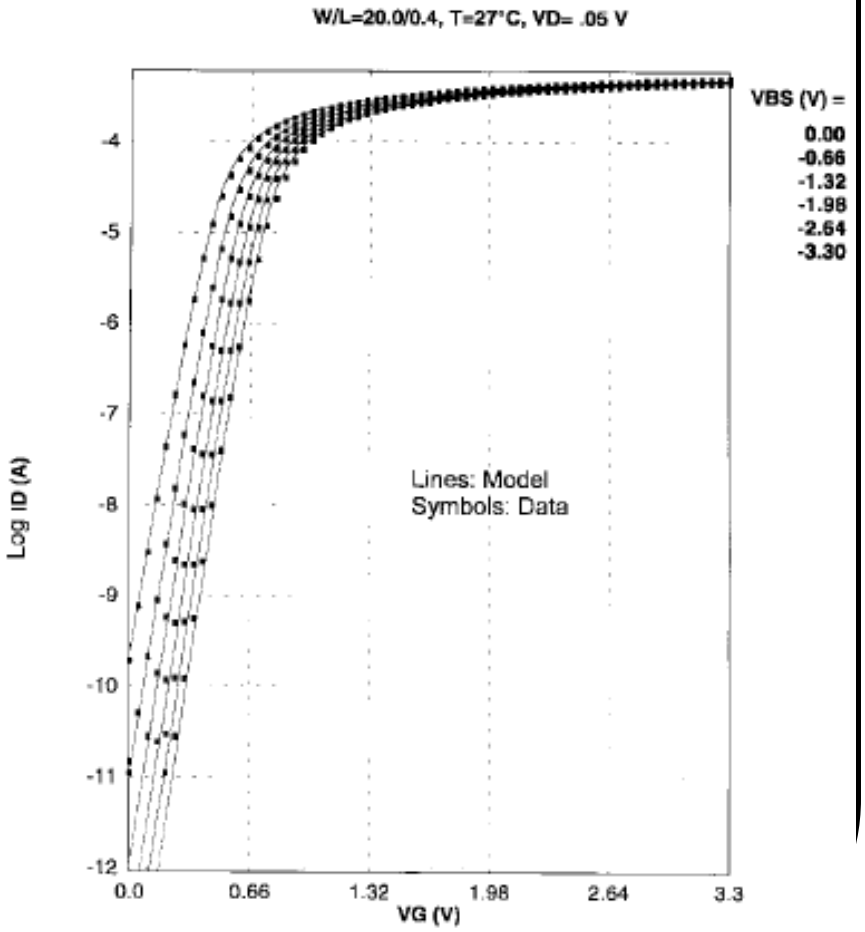
- The inversion layer has a significant thickness (T_{ch}).
- There are more more subthreshold inversion electrons at the corners.

7.11 MOSFET Compact Modeling for Circuit Simulation

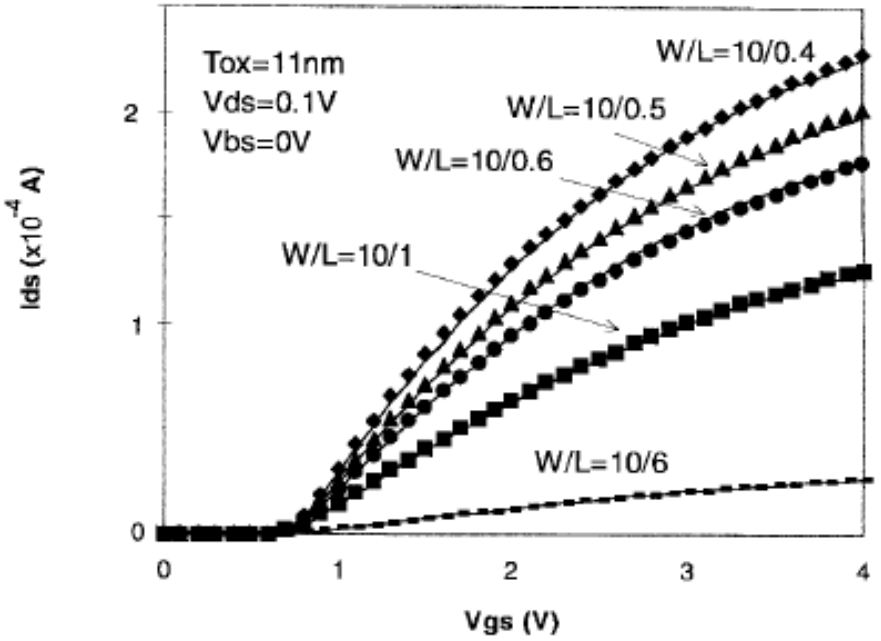
- For circuit simulation, MOSFETs are modeled with analytical equations.
- Device model is the link between technology/manufacturing and design/product. The other link is design rules.
- Circuits are designed A. through circuit simulations or B. using cell libraries that have been carefully designed beforehand using circuit simulations.
- **BSIM** is the first industry standard MOSFET model. It contains all the models presented in these chapters and more.

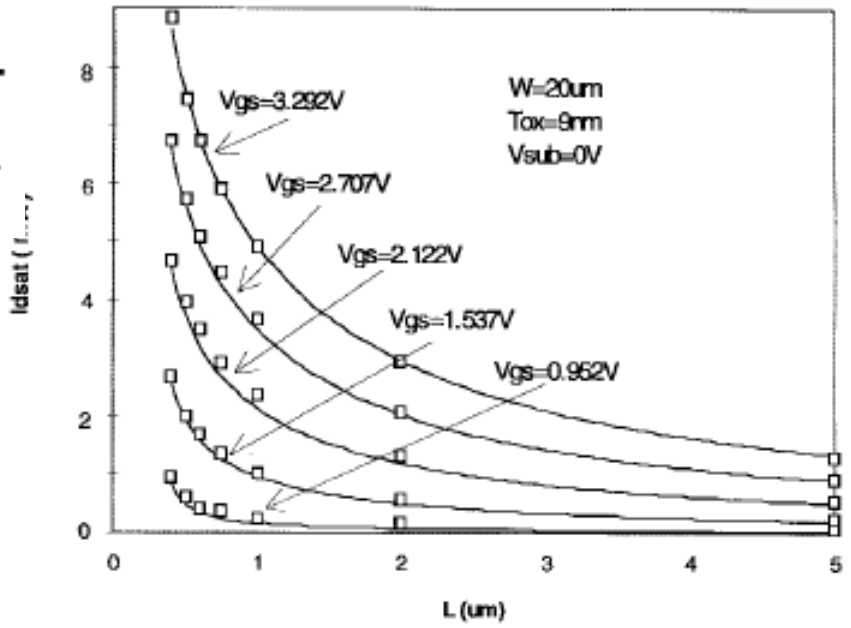
Examples of BSIM Model Results



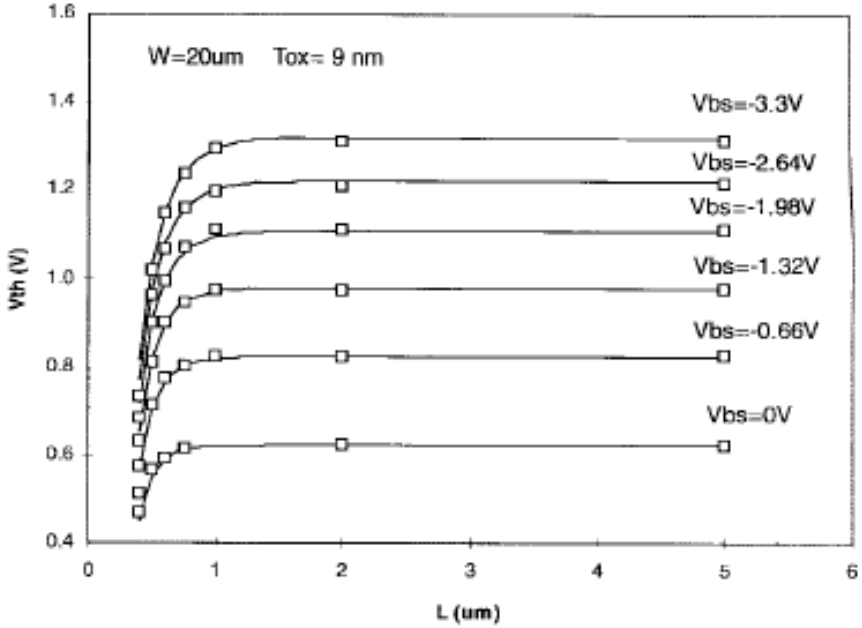


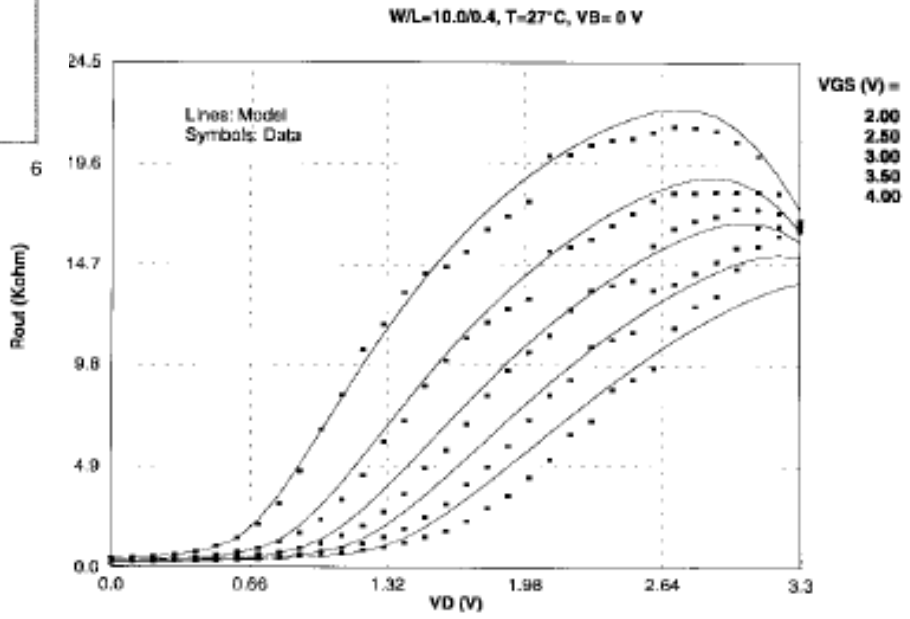
Example of BSIM Model Results





Example of BSIM Model Results





7.12 Chapter Summary

The major component of I_{off} is the *subthreshold current*

$$I_{off}(nA) = 100 \cdot \frac{W}{L} \cdot e^{-qV_t/\eta kT} = 100 \cdot \frac{W}{L} \cdot 10^{-V_t/S}$$

V_t decreases with L, a fact known as V_t *roll-off*, caused by drain-induced barrier lowering (DIBL).

$$V_t = V_{t-long} - (V_{ds} + 0.4) \cdot e^{-L/l_d}$$

$$l_d \propto \sqrt[3]{T_{oxe} W_{dep} X_j}$$

Output conductance of short channel transistors

$$g_{ds} = g_{msat} \times e^{-L/l_d}$$

Chapter 8 Bipolar Junction Transistors

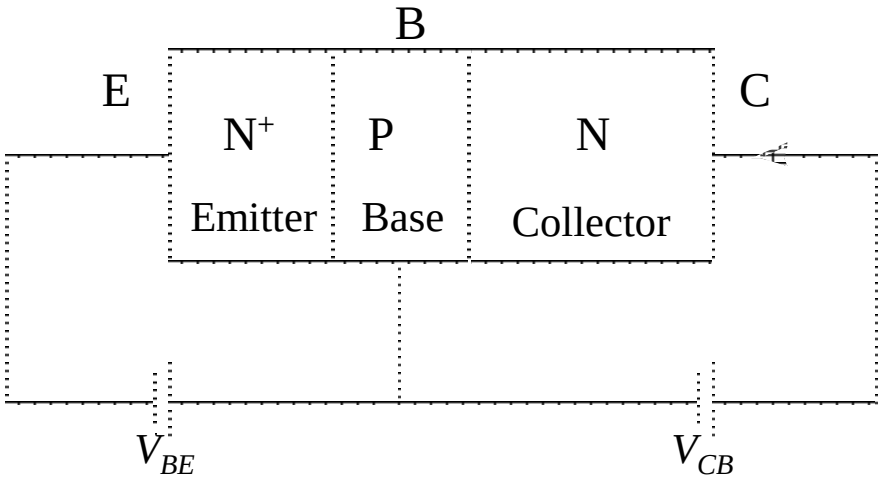
- Since 1970, the high density and low-power advantage of the MOS technology steadily eroded the BJT's early dominance.

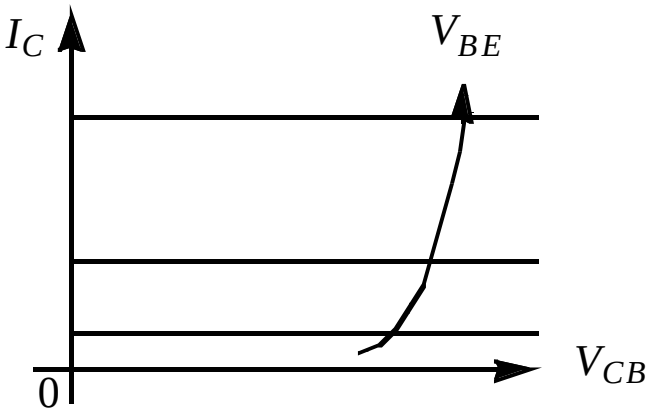
- BJTs are still preferred in some high-frequency and analog applications because of their high speed and high power output.

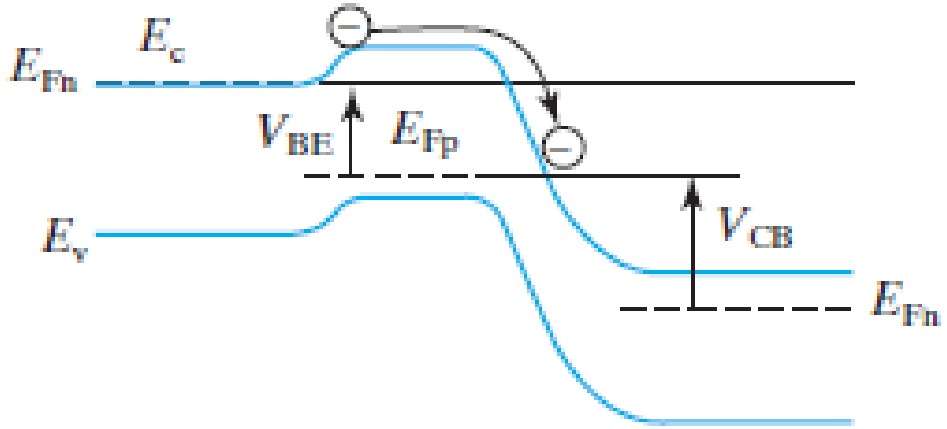
Question: What is the meaning of "bipolar" ?

8.1 Introduction to the BJT

NPN BJT:

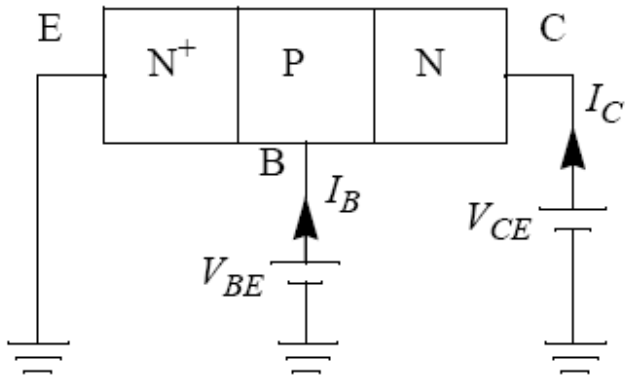


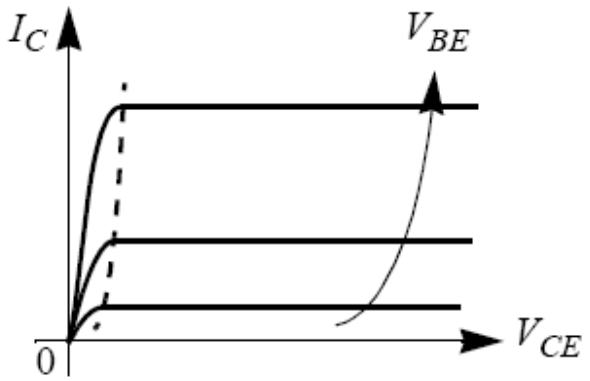


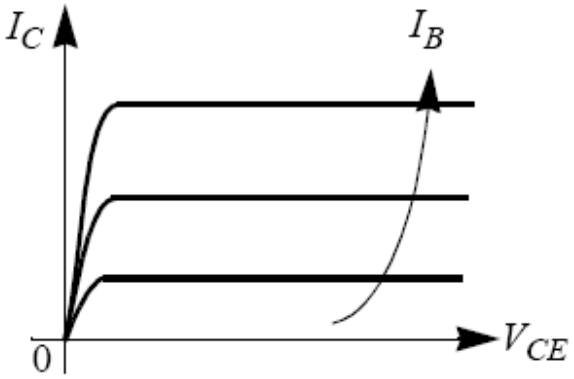


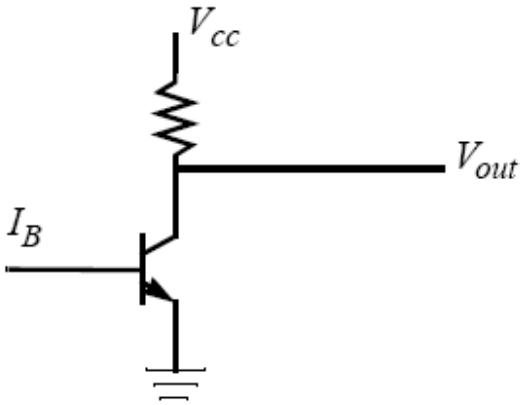
I_C is an exponential function of forward V_{BE} and independent of reverse V_{CB}.

Common-Emitter Configuration



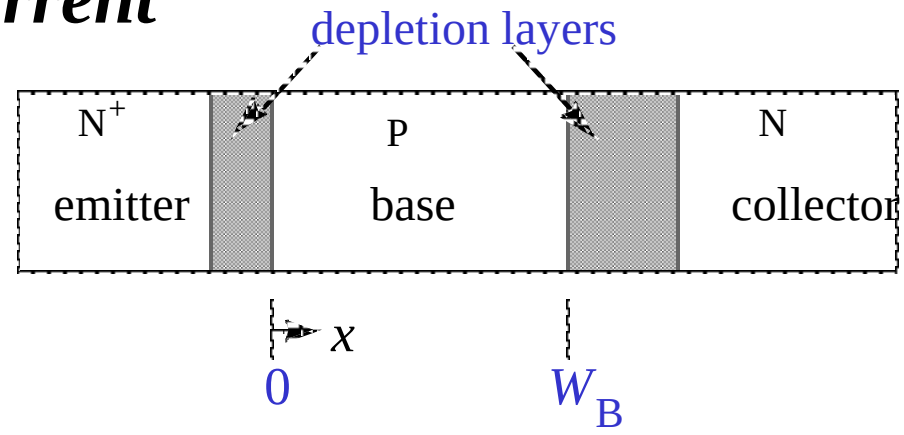






Question: Why is I_B often preferred as a parameter over V_{BE}?

8.2 Collector Current



$$\frac{d^2 n'}{dx^2} = \frac{n'}{L_B^2}$$

$$L_B \equiv \sqrt{\tau_B D_B}$$

τ_B : base recombination lifetime

D_B : base minority carrier (electron) diffusion constant

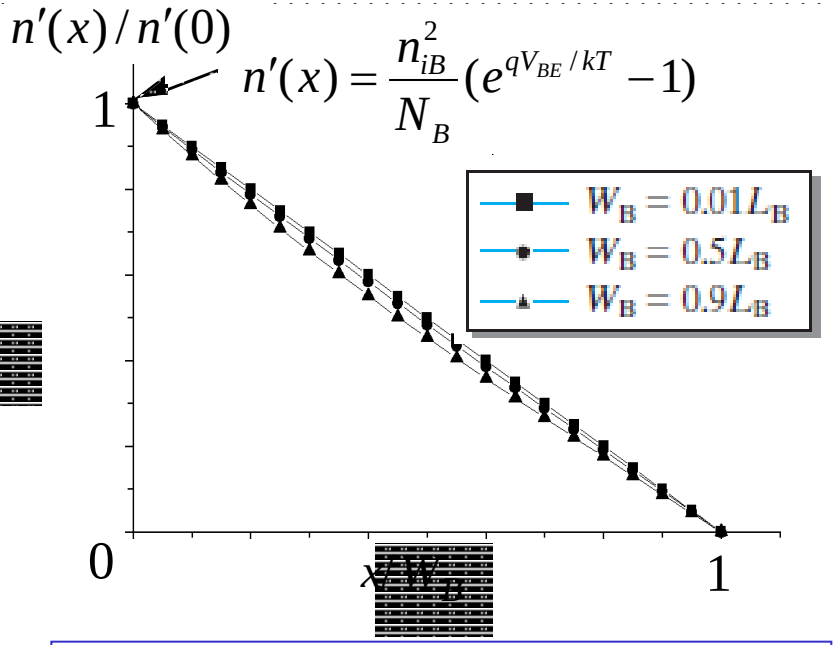
Boundary conditions :

$$n'(0) = n_{B0}(e^{qV_{BE}/kT} - 1)$$

$$n'(W_B) = n_{B0}(e^{qV_{BC}/kT} - 1) \approx -n_{B0} \approx 0$$

8.2 Collector Current

$$n'(x) = n_{B0}(e^{qV_{BE}/kT} - 1)\frac{\sinh\left(\frac{W_B - x}{L_B}\right)}{\sinh(W_B / L_B)}$$



$$n'(x) = n'(0)(1 - x/W_B)$$

$$= \frac{n_{iB}^2}{N_B}(e^{qV_{BE}/kT} - 1)(1 - x/W_B)$$

$$I_C = \left| A_E q D_B \frac{dn}{dx} \right|$$

$$= A_E q \frac{D_B}{W_B}\frac{n_{iB}^2}{N_B}(e^{qV_{BE}/kT} - 1)$$

$$I_C = I_S(e^{qV_{BE}/kT} - 1)$$

It can be shown

$$I_C = A_E \frac{qn_i^2}{G_B}(e^{qV_{BE}/kT} - 1)$$

$$G_B \equiv \int_0^{W_B} \frac{n_i^2}{n_{iB}^2}\frac{p}{D_B}dx$$

G_B (s·cm^4) is the ***base Gummel number***

8.2.1 High Level Injection Effect

- At low-level injection, inverse slope is 60 mV/decade
- High-level injection effect :

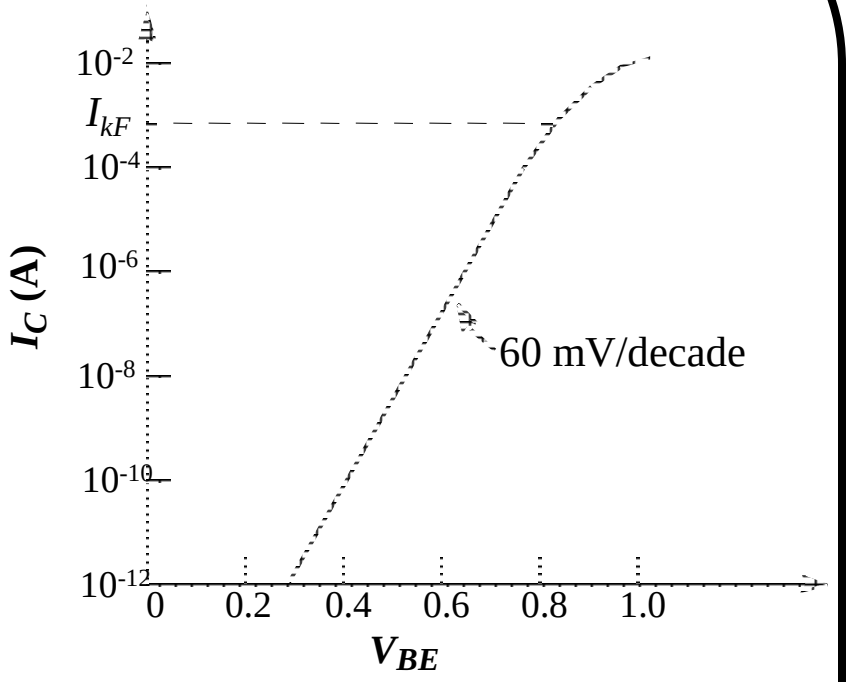
At large V_{BE}, $n' = p' >> N_B$

$$n' = p' = n = p$$

$$np = n_i^2 e^{q(E_{Fn}-E_{Fp})/kT} = n_i^2 e^{qV_{BE}/kT}$$

$$\therefore n \approx p \approx n_i e^{qV_{BE}/2kT}$$

$$G_B \propto p = n_i e^{qV_{BE}/2kT}$$

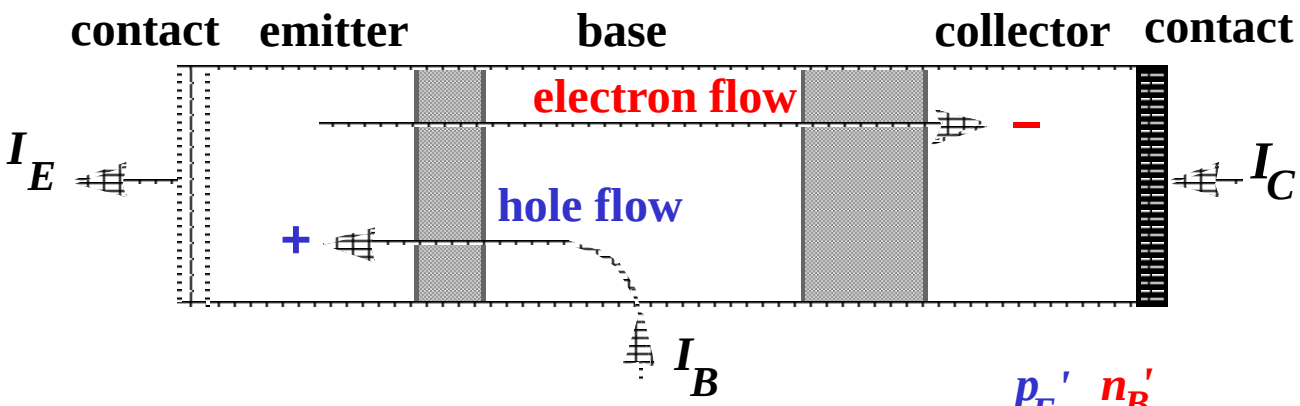


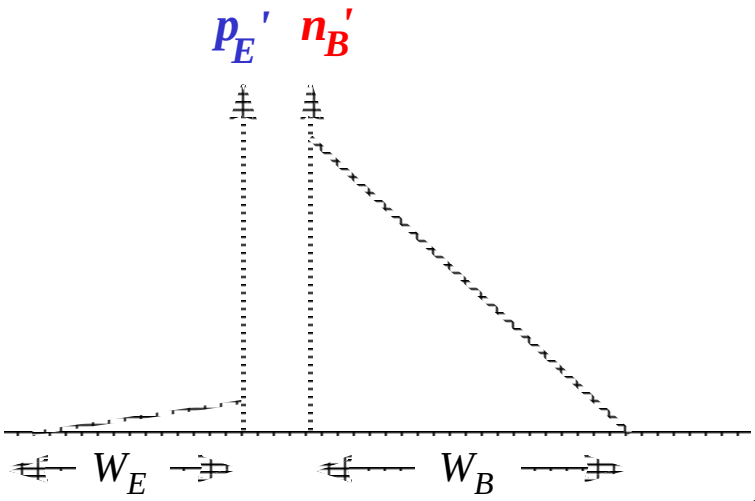
$$\boxed{I_C \propto n_i e^{qV_{BE}/2kT}}$$

When $p > N_B$, inverse slope is 120mV/decade.

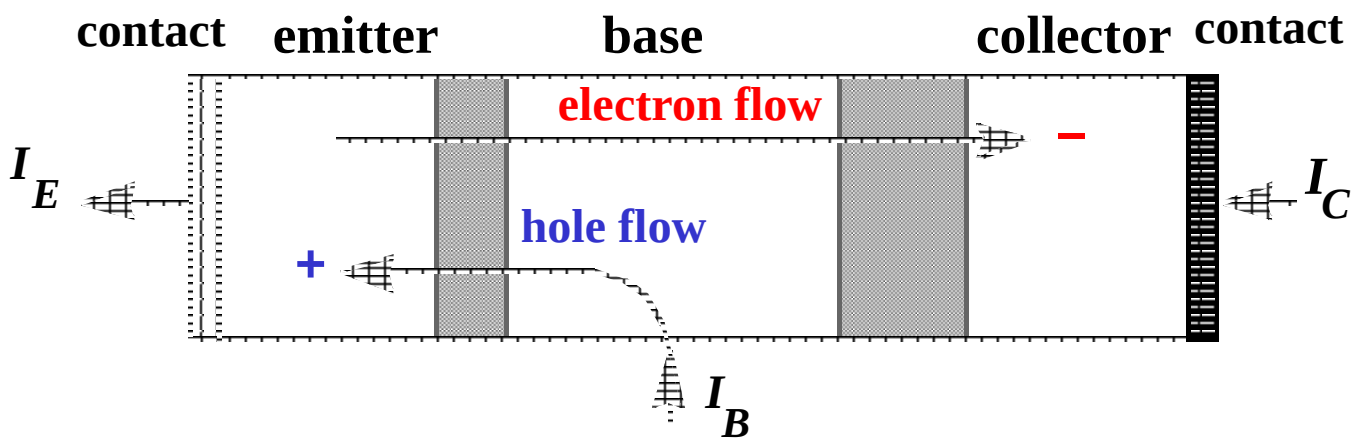
8.3 Base Current

Some holes are injected from the P-type base into the N^+ emitter. The holes are provided by the base current, I_B.





8.3 Base Current



$$I_B = A_E \frac{qn_i^2}{G_E}(e^{qV_{BE}/kT} - 1)$$

For a uniform emitter,

$$G_E \equiv \int_0^{W_E} \frac{n_i^2}{n_{iE}^2} \frac{n}{D_E} dx$$

$$I_B = A_E q \frac{D_E n_{iE}^2}{W_E N_E}(e^{qV_{BE}/kT} - 1)$$

Is a large I_B desirable? Why?

8.4 Current Gain

Common-emitter current gain, β_F :

$$\beta_F \equiv \frac{I_C}{I_B}$$

Common-base current gain:

$$I_C = \alpha_F I_E$$

$$\alpha_F \equiv \frac{I_C}{I_E} = \frac{I_C}{I_B + I_C} = \frac{I_C / I_B}{1 + I_C / I_B} = \frac{\beta_F}{1 + \beta_F}$$

It can be shown that $\beta_F = \dfrac{\alpha_F}{1 - \alpha_F}$

$$\beta_F = \frac{G_E}{G_B} = \frac{D_B W_E N_E n_{iB}^2}{D_E W_B N_B n_{iE}^2}$$

How can β_F be maximized?

EXAMPLE: Current Gain

A BJT has I_C = 1 mA and I_B = 10 μA. What are I_E, β_F and α_F?

Solution:

$$I_E = I_C + I_B = 1\,\text{mA} + 10\,\mu\text{A} = 1.01\,\text{mA}$$

$$\beta_F = I_C / I_B = 1\,\text{mA} / 10\,\mu\text{A} = 100$$

$$\alpha_F = I_C / I_E = 1\,\text{mA} / 1.01\,\text{mA} = 0.9901$$

We can confirm

$$\alpha_F = \frac{\beta_F}{1+\beta_F} \quad \text{and} \quad \beta_F = \frac{\alpha_F}{1-\alpha_F}$$

8.4.1 Emitter Bandgap Narrowing

$$\beta \propto \frac{N_E}{N_B}\frac{n_{iB}^2}{n_{iE}^2}$$

To raise β_F, N_E is typically very large. Unfortunately, large N_E makes $n_{iE}^2 > n_i^2$ (heavy doping effect).

$$n_i^2 = N_C N_V e^{-E_g/kT}$$

Since n_i is related to E_g, this effect is also known as band-gap narrowing.

$$n_{iE}^2 = n_i^2 e^{\Delta E_{gE}/kT}$$

ΔE_{gE} is negligible for $N_E < 10^{18}$ cm^{-3}, is 50 meV at 10^{19}cm^{-3}, 95 meV at 10^{20}cm^{-3}, and 140 meV at 10^{21} cm^{-3}.

Emitter bandgap narrowing makes it difficult to raise β_F by doping the emitter very heavily.

8.4.2 Narrow-Bandgap Base and Heterojuncion BJT

$$\beta \propto \frac{N_E}{N_B}\frac{n_{iB}^2}{n_{iE}^2}$$

To further elevate β_F , we can raise n_{iB} by using an epitaxial $Si_{1-\eta}Ge_{\eta}$ base.

With $\eta = 0.2$, E_{gB} is reduced by 0.1eV and n_{iE}^2 by 30x.

EXAMPLE: Emitter Bandgap Narrowing and SiGe Base

Assume $D_B = 3D_E$, $W_E = 3W_B$, $N_B = 10^{18}$ cm^{-3}, and $n_{iB}^2 = n_i^2$. What is β_F for (a) $N_E = 10^{19}$ cm^{-3}, (b) $N_E = 10^{20}$ cm^{-3}, and (c) $N_E = 10^{20}$ cm^{-3} and a SiGe base with $\Delta E_{gB} = 60$ meV ?

(a) At $N_E = 10^{19}$ cm^{-3}, $\Delta E_{gE} \approx 50$ meV,

$$n_{iE}^2 = n_i^2 e^{\Delta E_{gE}/kT} = n_i^2 e^{50\,\text{meV}/26\,\text{meV}} = n_i^2 e^{1.92} = 6.8 n_i^2$$

$$\beta_F = \frac{D_B W_E}{D_E W_B} \cdot \frac{N_E n_i^2}{N_B n_{iE}^2} = \frac{9 \cdot 10^{19} \cdot n_i^2}{10^{18} \cdot 6.8 n_i^2} = 13$$

(b) At $N_E = 10^{20}$ cm^{-3}, $\Delta E_{gE} \approx 95$ meV

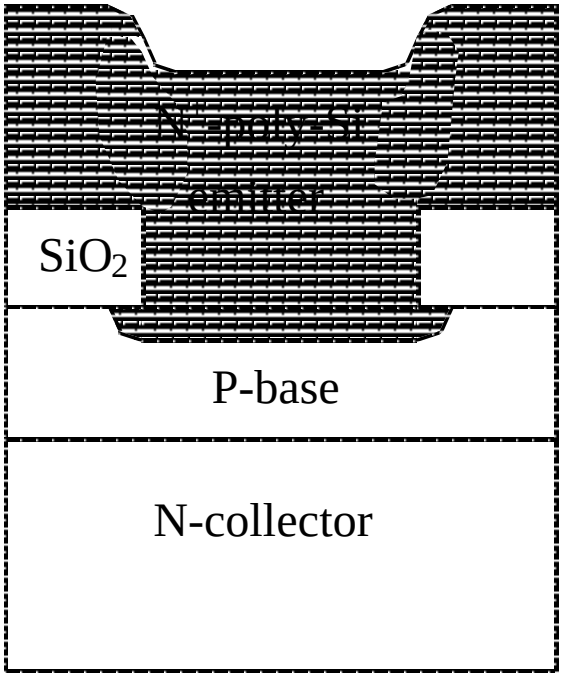
$$n_{iE}^2 = 38 n_i^2 \qquad \beta_F = 24$$

(c) $$n_{iB}^2 = n_i^2 e^{\Delta E_{gB}/kT} = n_i^2 e^{60\,\text{meV}/26\,\text{meV}} = 10 n_i^2 \qquad \beta_F = 237$$

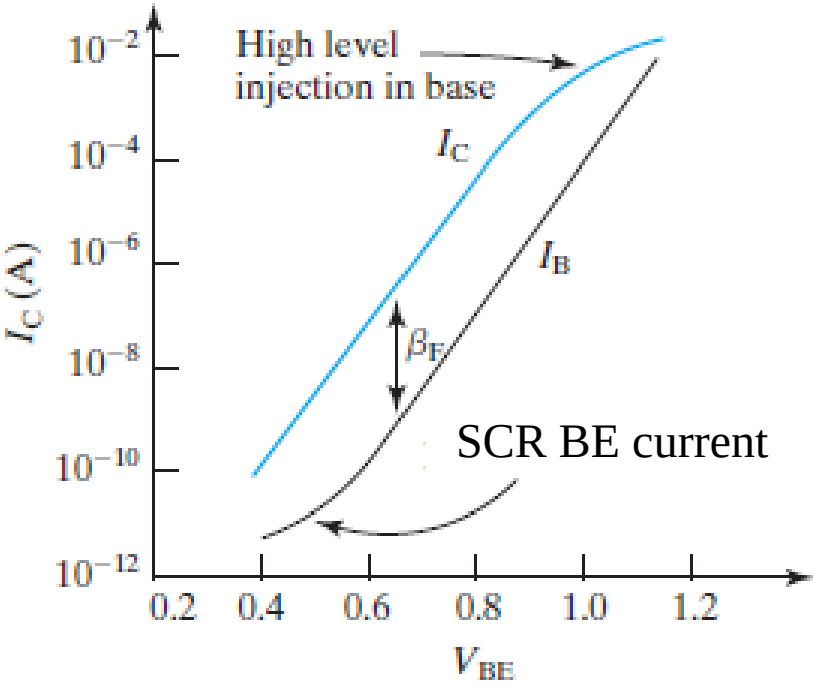
8.4.3 Poly-Silicon Emitter

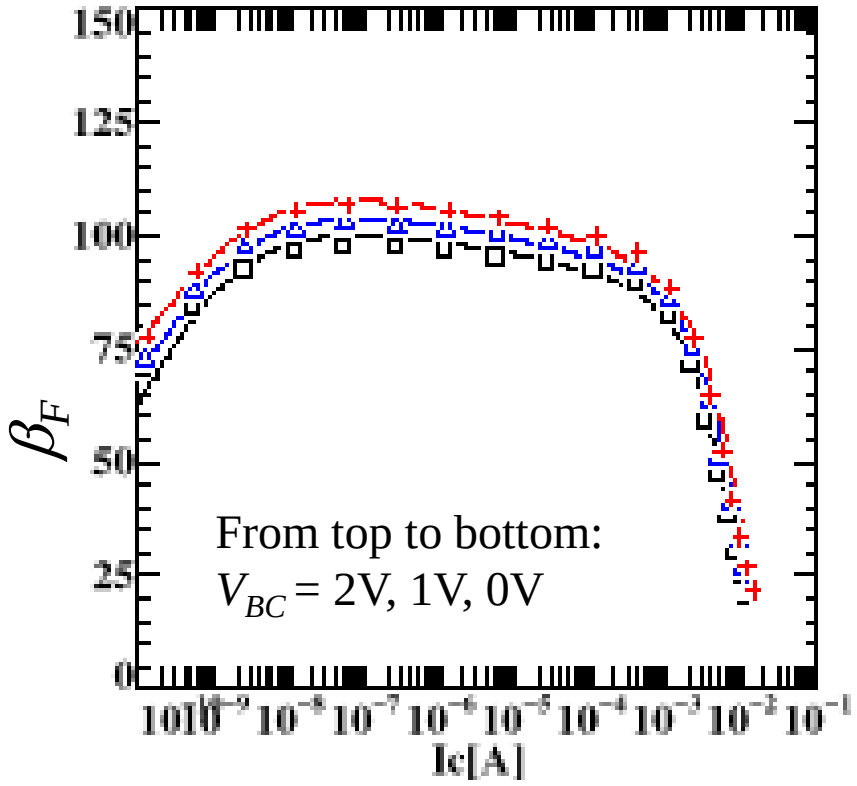
A high-performance BJT typically has a layer of As-doped N^+ poly-silicon film in the emitter.

β_F is larger due to the large W_E, mostly made of the N^+ poly-silicon. (A deep diffused emitter junction tends to cause emitter-collector shorts.)



8.4.4 Gummel Plot and β_F Fall-off at High and Low I_c

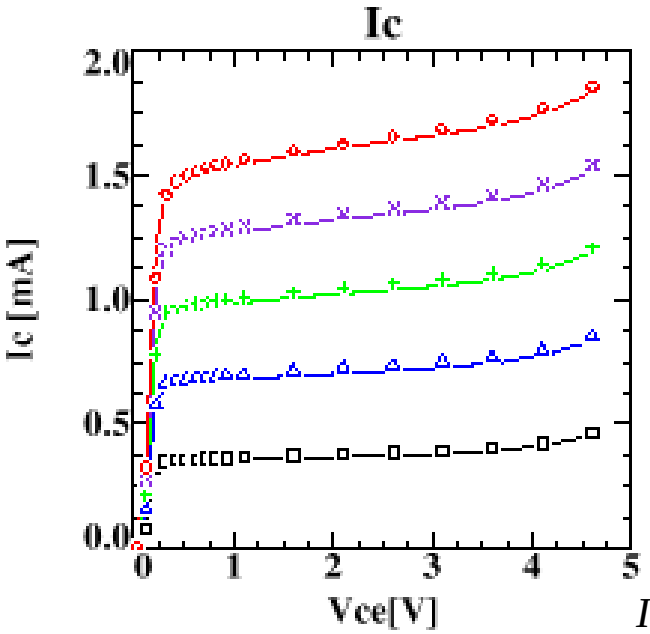




Why does one want to operate BJTs at low I_C and high I_C?
Why is β_F a function of V_{BC} in the right figure?

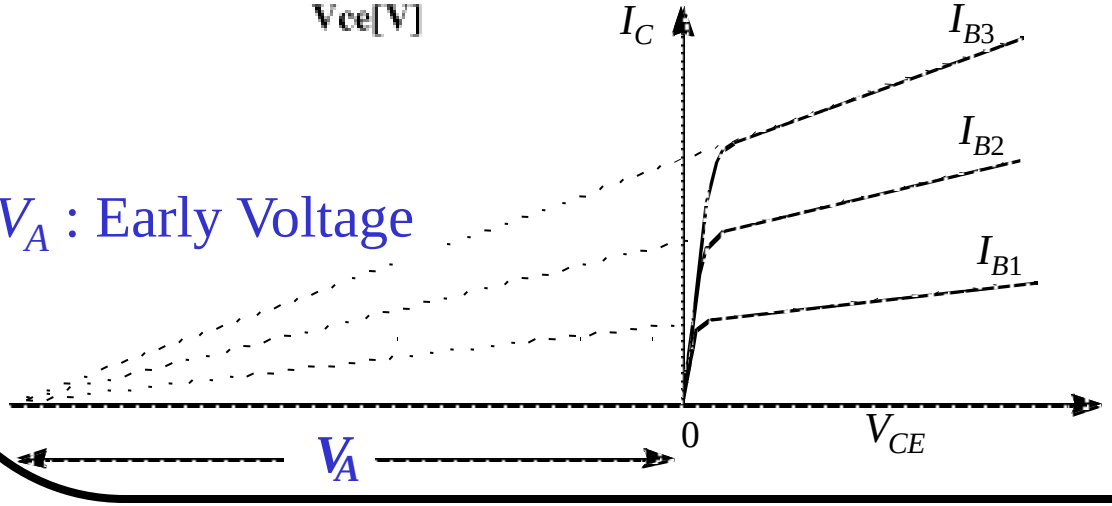
Hint: See Sec. 8.5 and Sec. 8.9.

8.5 Base-Width Modulation by Collector Voltage



Output resistance :

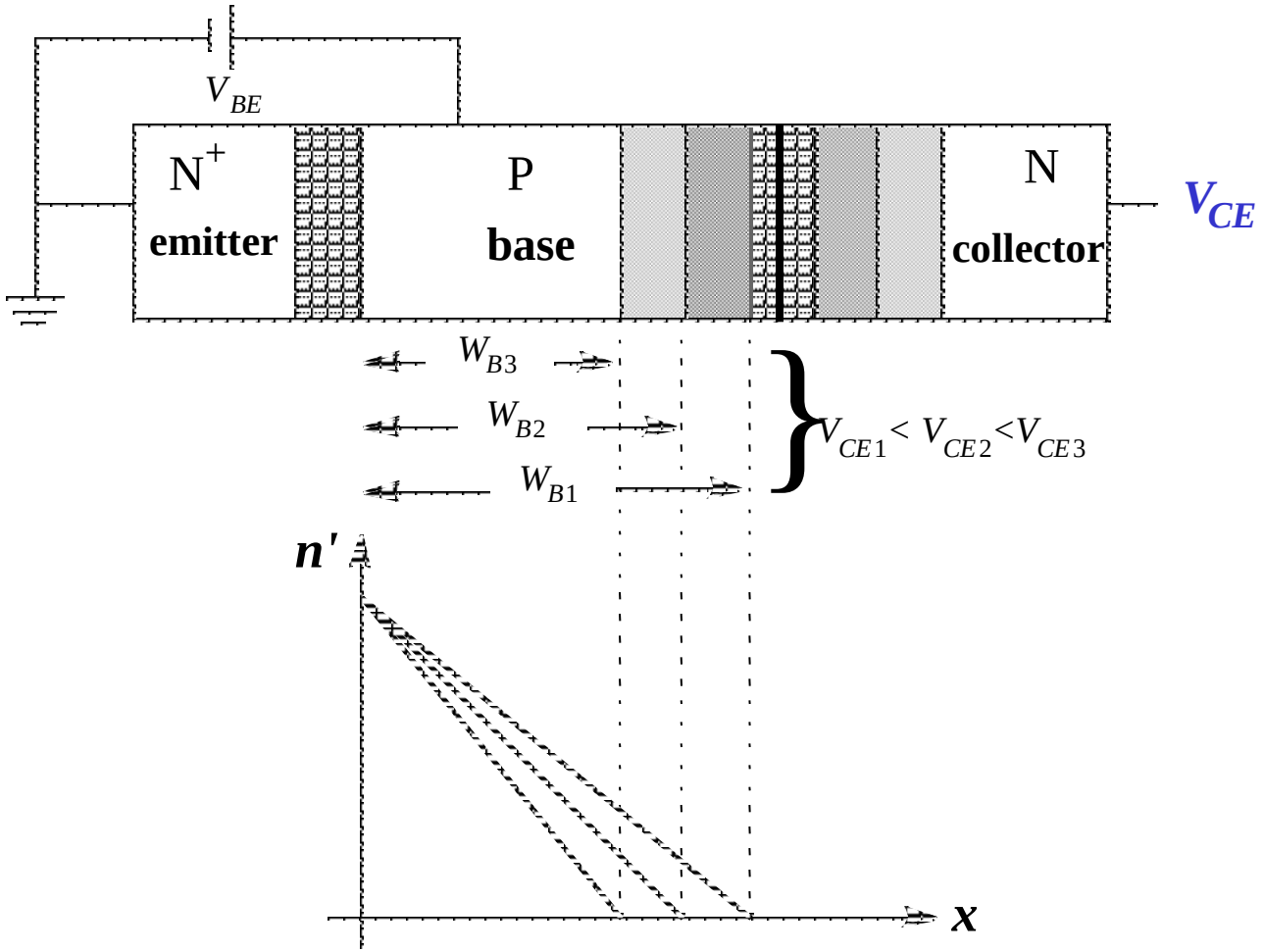
$$r_0 \equiv \left(\frac{\partial I_C}{\partial V_{CE}} \right)^{-1} = \frac{V_A}{I_C}$$



V_A : Early Voltage

Large V_A (large r_o) is desirable for a large voltage gain

8.5 Base-Width Modulation by Collector Voltage

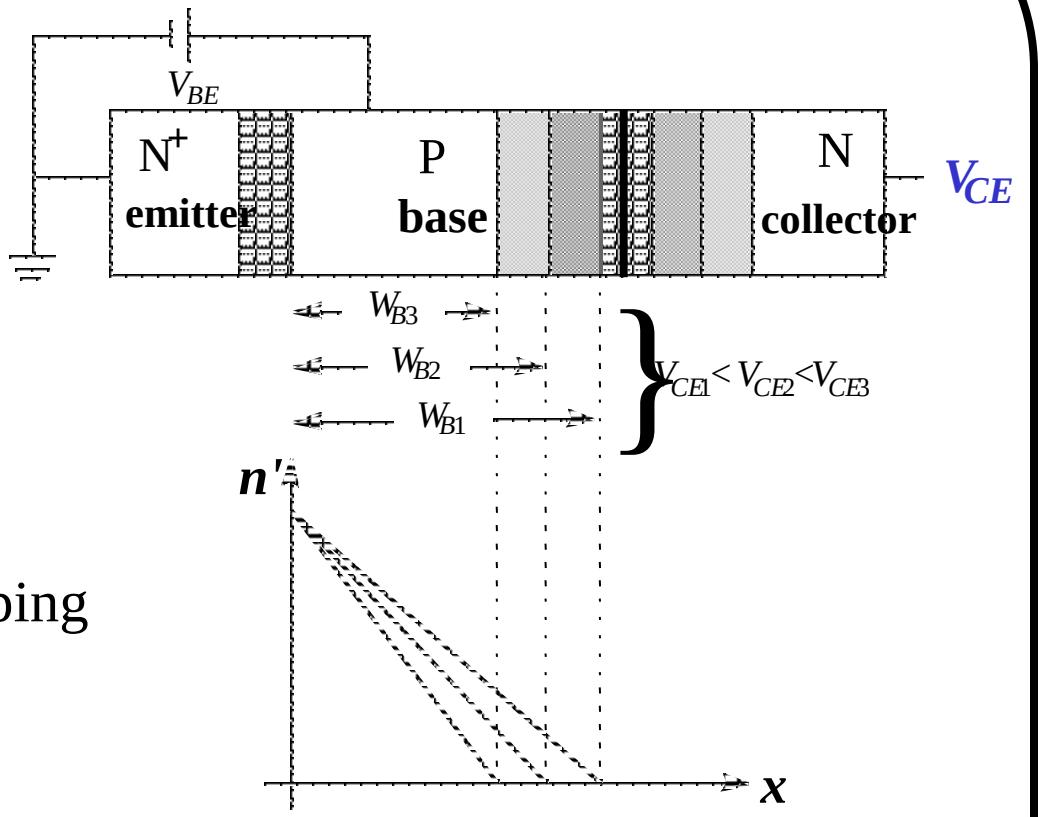


How can we reduce the base-width modulation effect?

8.5 Base-Width Modulation by Collector Voltage

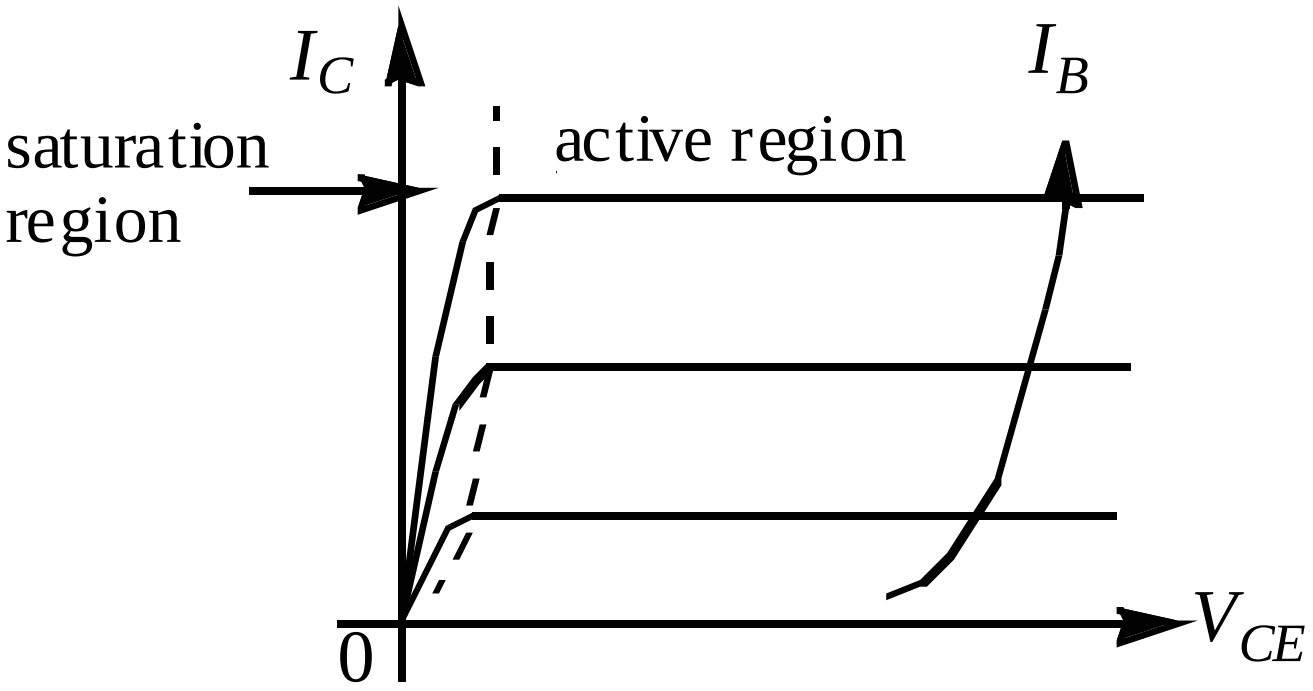
The base-width modulation effect is reduced if we

(A) Increase the base width,
(B) Increase the base doping concentration, N_B, or
(C) Decrease the collector doping concentration, N_C.



Which of the above is the most acceptable action?

8.6 Ebers-Moll Model



The Ebers-Moll model describes both the active and the saturation regions of BJT operation.

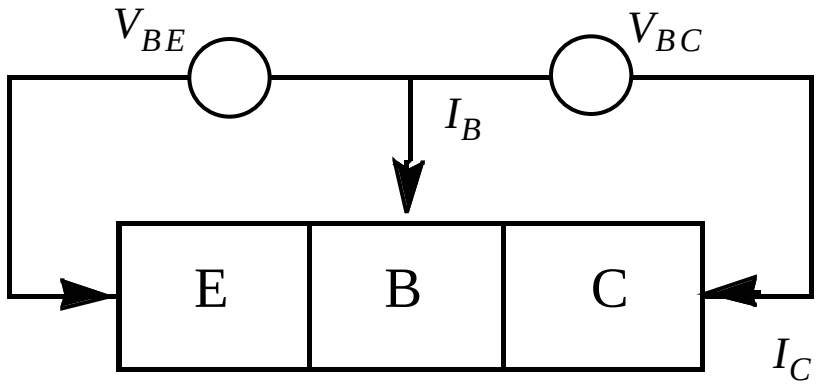
8.6 Ebers-Moll Model

I_C is driven by two two forces, V_{BE} and V_{BC}.

When only V_{BE} is present :

$$I_C = I_S(e^{qV_{BE}/kT} - 1)$$

$$I_B = \frac{I_S}{\beta_F}(e^{qV_{BE}/kT} - 1)$$



Now reverse the roles of emitter and collector.

When only V_{BC} is present :

$$I_E = I_S(e^{qV_{BC}/kT} - 1)$$

$$I_B = \frac{I_S}{\beta_R}(e^{qV_{BC}/kT} - 1)$$

β_R : reverse current gain

β_F : forward current gain

$$I_C = -I_E - I_B = -I_S(1 + \frac{1}{\beta_R})(e^{qV_{BC}/kT} - 1)$$

8.6 Ebers-Moll Model

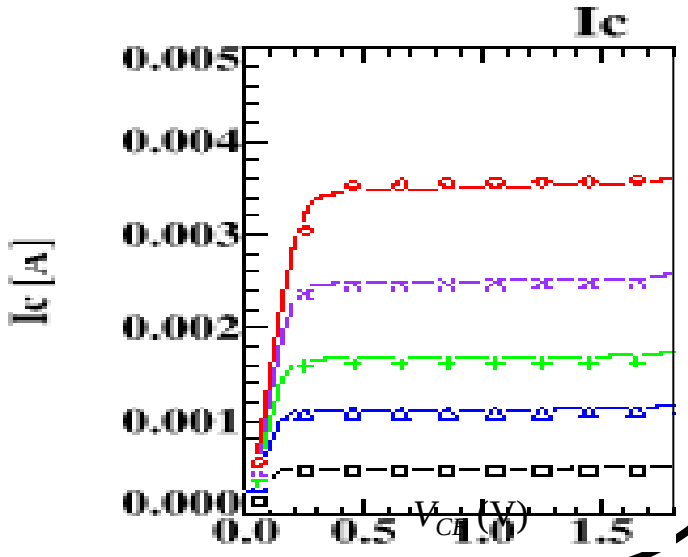
In general, both V_{BE} and V_{BC} are present :

$$I_C = I_S(e^{qV_{BE}/kT} - 1) - I_S(1 + \frac{1}{\beta_R})(e^{qV_{BC}/kT} - 1)$$

$$I_B = \frac{I_S}{\beta_F}(e^{qV_{BE}/kT} - 1) + \frac{I_S}{\beta_F}(e^{qV_{BC}/kT} - 1)$$

In saturation, the BC junction becomes forward-biased, too.

V_{BC} causes a lot of holes to be injec into the collector. This uses up muc of I_B. As a result, I_C drops.



8.7 Transit Time and Charge Storage

When the BE junction is forward-biased, excess holes are stored in the emitter, the base, and even in the depletion layers.

Q_F is all the stored excess hole charge

$$\tau_F \equiv \frac{Q_F}{I_C}$$

τ_F is difficult to be predicted accurately but can be measured.

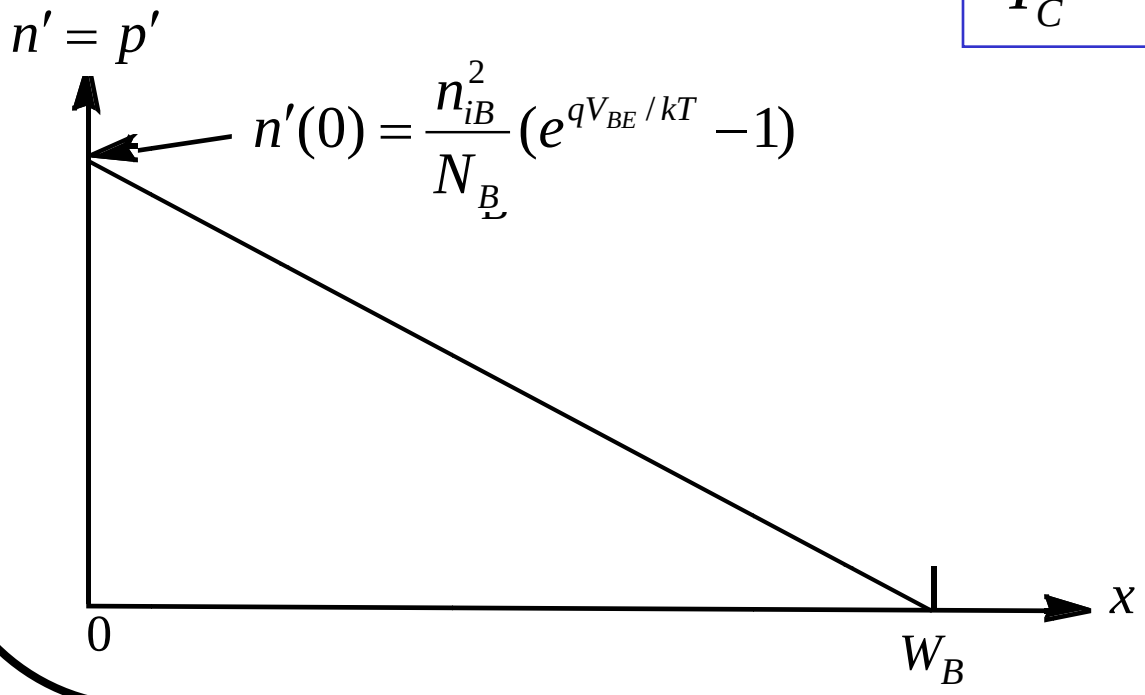
τ_F determines the high-frequency limit of BJT operation.

8.7.1 Base Charge Storage and Base Transit Time

Let's analyze the excess hole charge and transit time in the base only.

$$Q_{FB} = qA_E n'(0) W_B / 2$$

$$\frac{Q_{FB}}{I_C} \equiv \tau_{FB} = \frac{W_B^2}{2D_B}$$



EXAMPLE: Base Transit Time

What is τ_{FB} if W_B = 70 nm and D_B = 10 cm^2/s?

Answer:

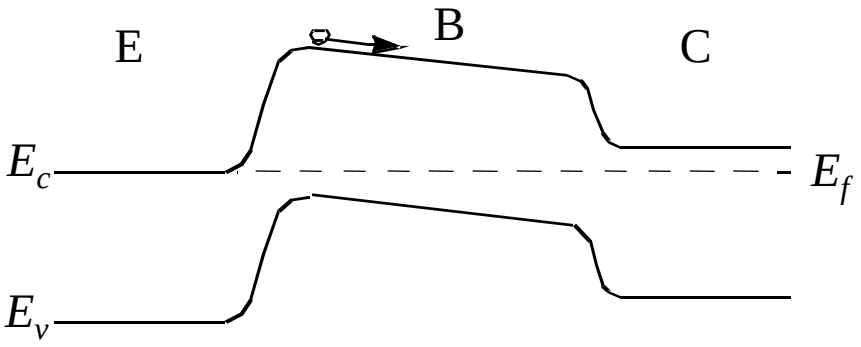
$$\tau_{FB} = \frac{W_B^2}{2D_B} = \frac{(7\times10^{-6}\ \text{cm})^2}{2\times10\ \text{cm}^2/\text{s}} = 2.5\times10^{-12}\,\text{s} = 2.5\,\text{ps}$$

2.5 ps is a very short time. Since light speed is 3×10^8 m/s, light travels only 1.5 mm in 5 ps.

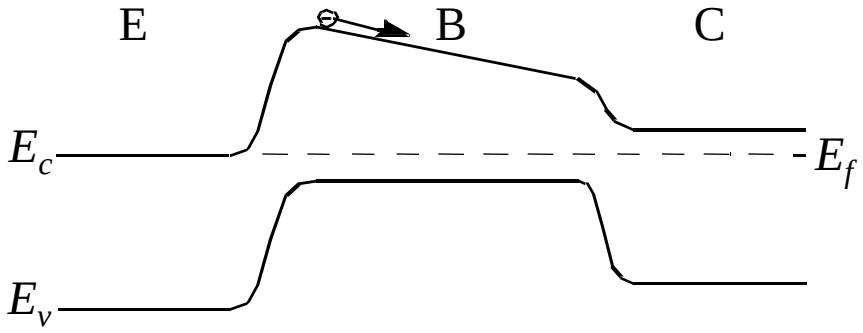
8.7.2 Drift Transistor–Built-in Base Field

The base transit time can be reduced by building into the base a drift field that aids the flow of electrons. Two methods:

- Fixed E_{gB}, N_B decreases from emitter end to collector end.



- Fixed N_B, E_{gB} decreases from emitter end to collector end.

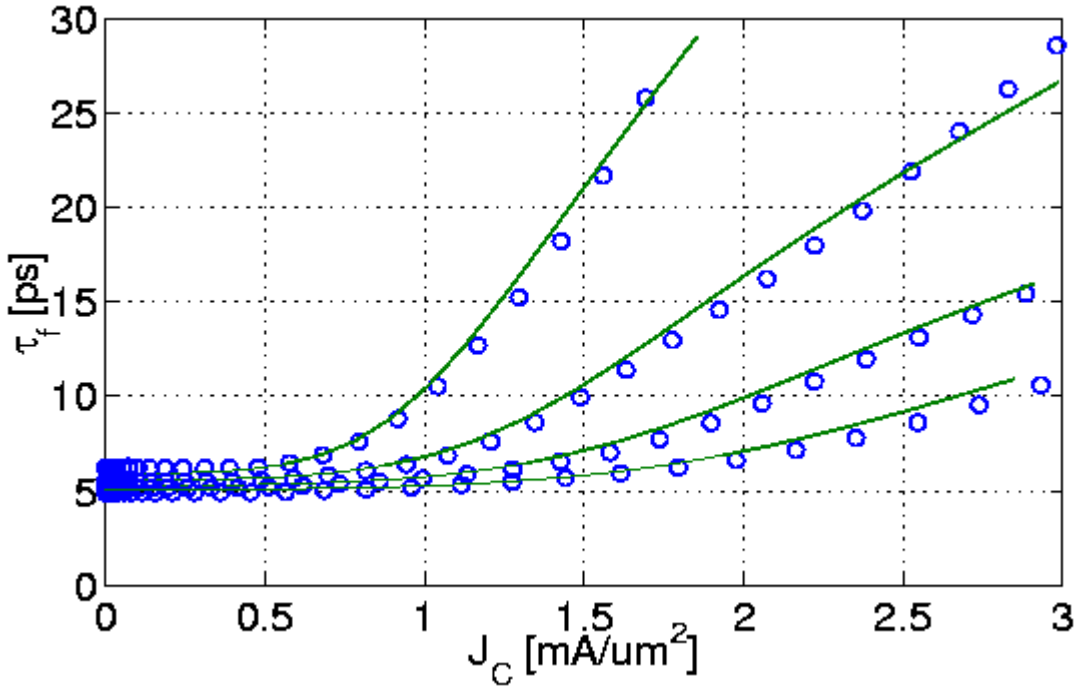


$$\mathcal{E} = \frac{1}{q}\frac{dE_c}{dx}$$

8.7.3 Emitter-to-Collector Transit Time and Kirk Effect

• To reduce the total transit time, emitter and depletion layers must be thin, too.

• Kirk effect or base widening: At high I_C the base widens into the collector. Wider base means larger τ_F .



Top to bottom :
V_{CE} = 0.5V, 0.8V, 1.5V, 3V.

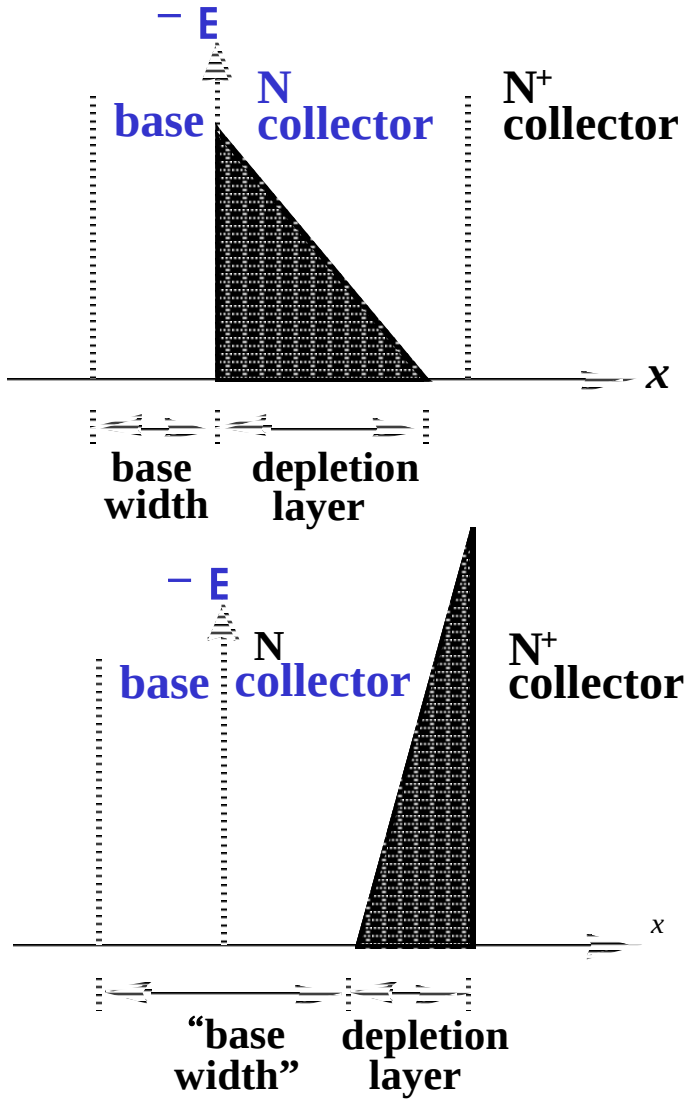
Base Widening at Large I_c

$$I_C = A_E q n v_{sat}$$

$$\rho = qN_C - qn$$

$$= qN_C - \frac{I_C}{A_E v_{sat}}$$

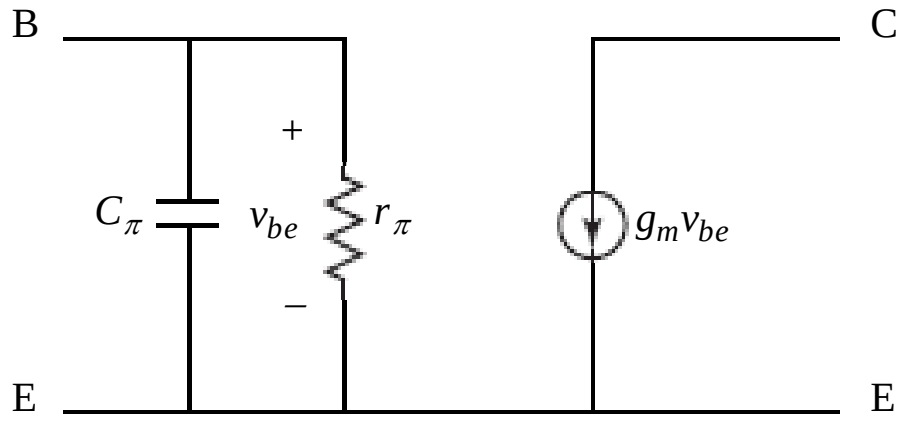
$$\frac{d\mathcal{E}}{dx} = \rho / \varepsilon_s$$



8.8 Small-Signal Model

$$I_C = I_S e^{qV_{BE}/kT}$$

Transconductance:



$$g_m \equiv \frac{dI_C}{dV_{BE}} = \frac{d}{dV_{BE}}(I_S e^{qV_{BE}/kT})$$

$$= \frac{q}{kT} I_S e^{qV_{BE}/kT} = I_C/(kT/q)$$

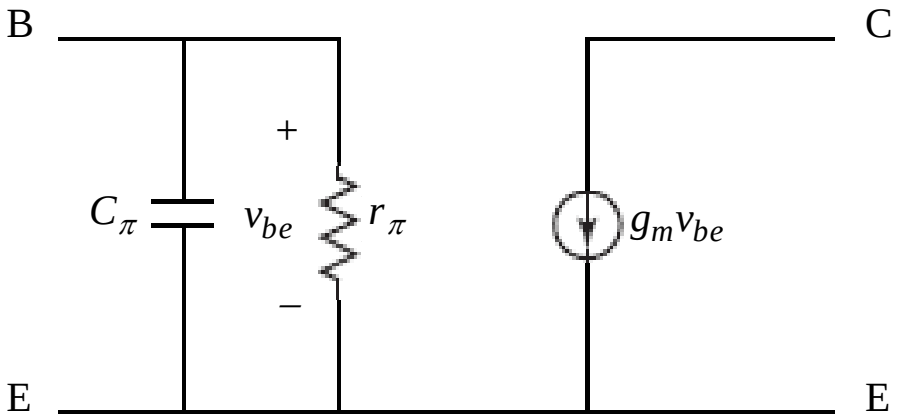
$$\boxed{g_m = I_C/(kT/q)}$$

At 300 K, for example, $g_m = I_C/26\text{mV}$.

8.8 Small-Signal Model

$$\frac{1}{r_\pi} = \frac{dI_B}{dV_{BE}} = \frac{1}{\beta_F}\frac{dI_C}{dV_{BE}} = \frac{g_m}{\beta_F}$$

$$\boxed{r_\pi = \beta_F / g_m}$$



$$C_\pi = \frac{dQ_F}{dV_{BE}} = \frac{d}{dV_{BE}}\tau_F I_C = \tau_F g_m$$

This is the charge-storage capacitance, better known as the ***diffusion capacitance***.

Add the depletion-layer capacitance, C_{dBE} :

$$C_\pi = \tau_F g_m + C_{dBE}$$

EXAMPLE: Small-Signal Model Parameters

A BJT is biased at I_C = 1 mA and V_{CE} = 3 V. β_F=90, τ_F=5 ps, and T = 300 K. Find (a) g_m, (b) r_π, (c) C_π.

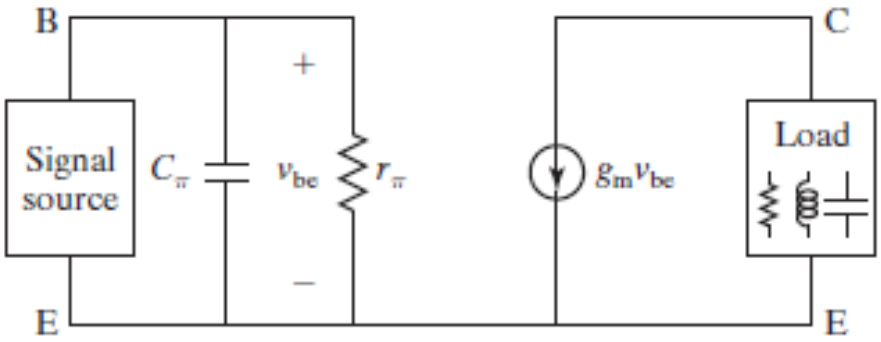
Solution:

(a) $g_m = I_C/(kT/q) = \dfrac{1\,\text{mA}}{26\,\text{mV}} = 39\dfrac{\text{mA}}{\text{V}} = 39\,\text{mS}$ *(milli siemens)*

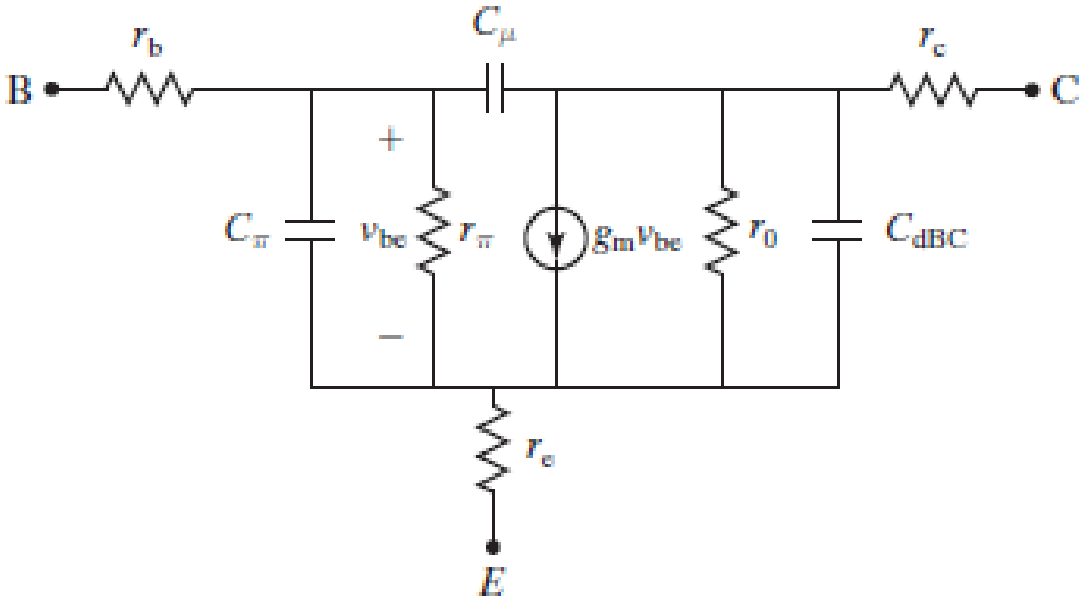
(b) $r_\pi = \beta_F / g_m = \dfrac{90}{39\,\text{mS}} = 2.3\,\text{k}\Omega$

(c) $C_\pi = \tau_F g_m = 5\times10^{-12}\times0.039 \approx 1.9\times10^{-14}\,\text{F} = 19\,\text{fF}$ *(femto farad)*

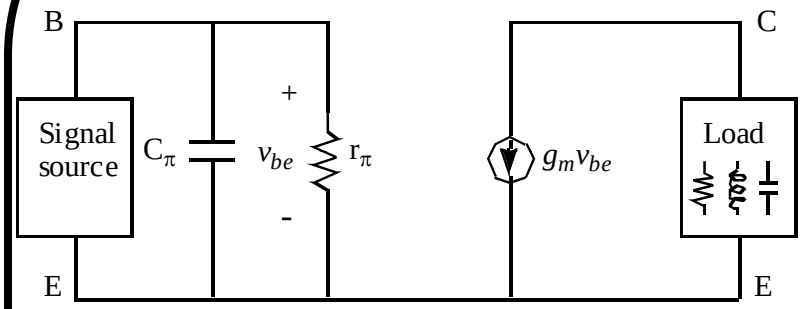
Once the model parameters are determined, one can analyze circuits with arbitrary source and load impedances.



The parameters are routinely determined through comprehensive measurement of the BJT AC and DC characteristics.



8.9 Cutoff Frequency



$$\beta = 1 \text{ at } f_T = \frac{1}{2\pi(\tau_F + C_{dBE}kT/qI_C)}$$

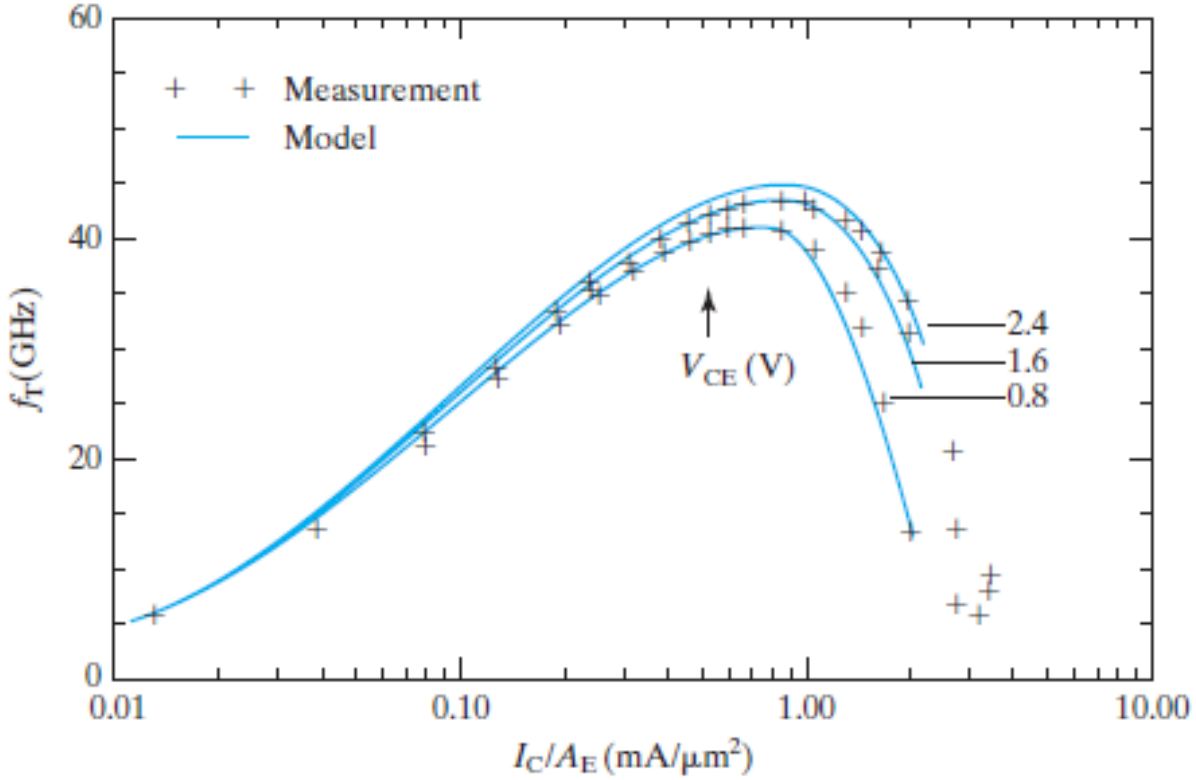
The load is a short circuit. The signal source is a current source, i_b, at frequency, f. At what frequency does the current gain $\beta(\equiv i_c / i_b)$ fall to unity?

$$v_{be} = \frac{i_b}{\text{input admittance}} = \frac{i_b}{1/r_\pi + j\omega C_\pi} \quad , \qquad C_\pi = \tau_F g_m + C_{dBE}$$

$$i_c = g_m v_{be}$$

$$\beta(\omega) = \left|\frac{i_c}{i_b}\right| = \frac{g_m}{\left|1/r_\pi + j\omega C_\pi\right|} = \frac{1}{\left|1/\beta_F + j\omega\tau_F + j\omega C_{dBE}kT/qI_C\right|}$$

8.9 Cutoff Frequency

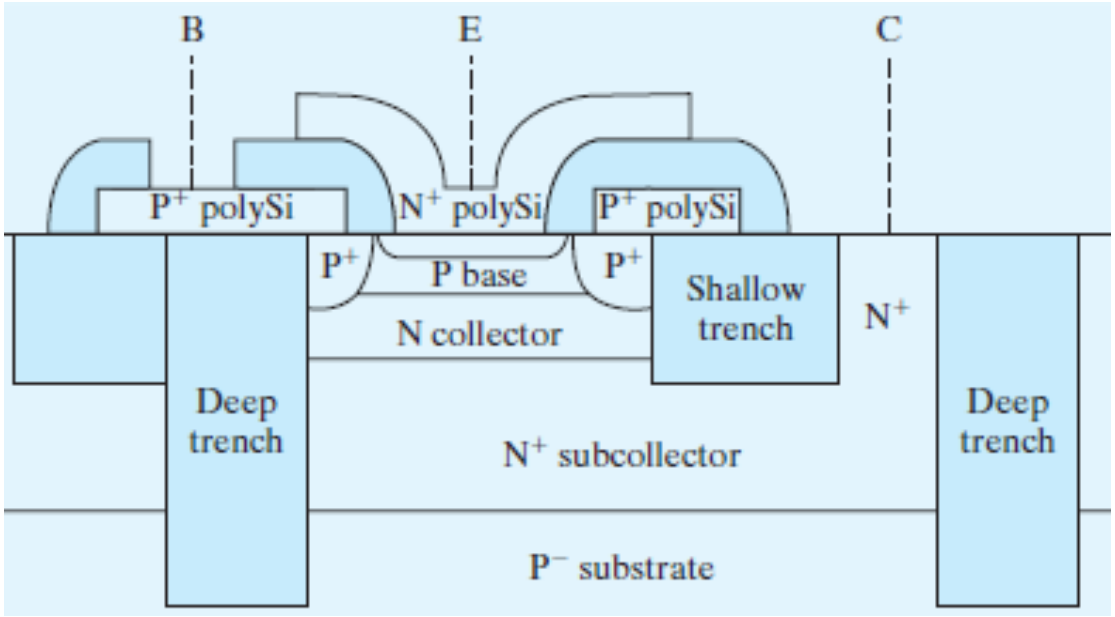


$$f_T = 1/2\pi(\tau_F + C_{dBE}kT/qI_C)$$

f_T is commonly used to compare the speed of transistors.

- Why does f_T increase with increasing I_C?
- Why does f_T fall at high I_C?

BJT Structure for Minimum Parasitics and High Speed



- Poly-Si emitter
- Thin base
- Self-aligned poly-Si base contact
- Narrow emitter opening
- Lightly-doped collector
- Heavily-doped epitaxial subcollector
- Shallow trench and deep trench for electrical **isolation**

8.10 Charge Control Model

•For the DC condition, $I_C(t) = Q_F(t)/\tau_F$

$$I_B = I_C / \beta_F = \frac{Q_F}{\tau_F \beta_F}$$

•In order to sustain an excess hole charge in the transistor, holes must be supplied through I_B to susbtain recombination at the above rate.

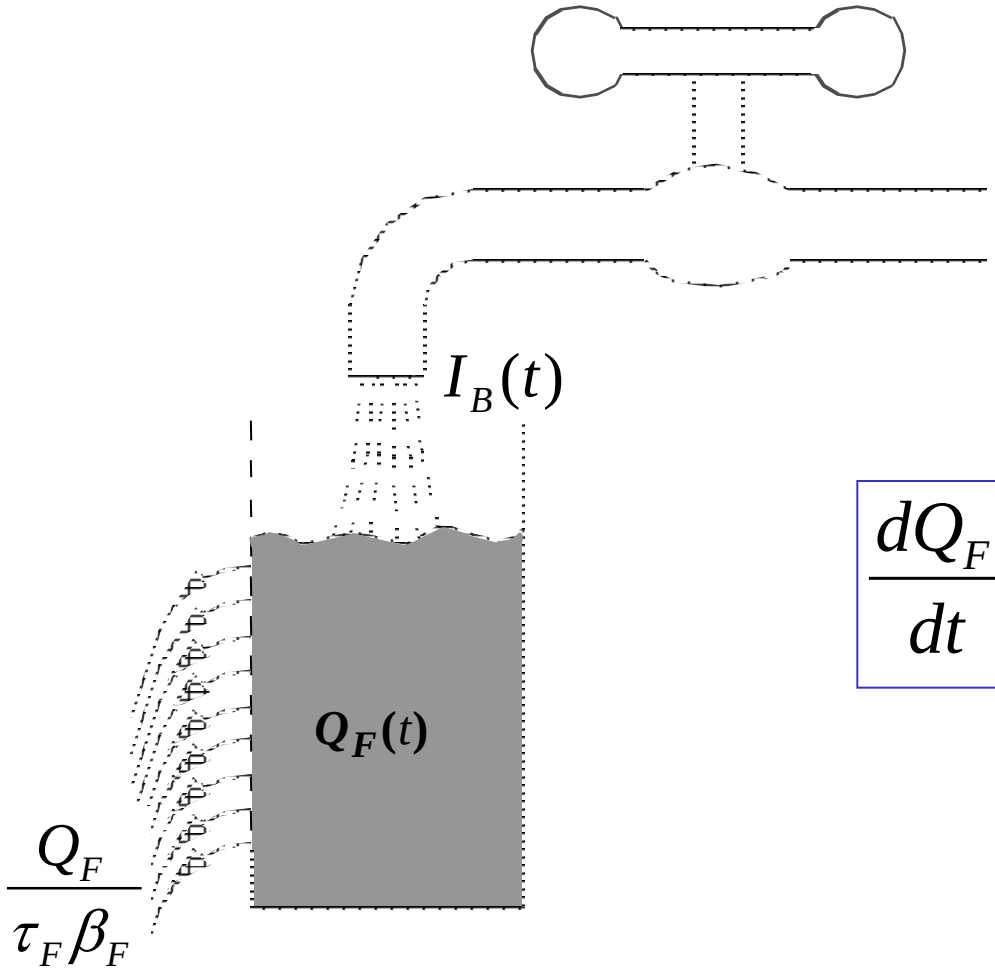
•What if I_B is larger than $Q_F / \tau_F \beta_F$?

$$\frac{dQ_F}{dt} = I_B(t) - \frac{Q_F}{\tau_F \beta_F}$$

Step 1: Solve it for any given $I_B(t)$ *to* find $Q_F(t)$.

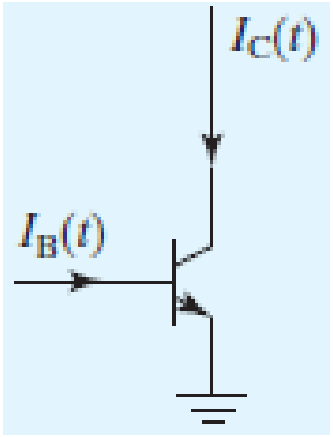
Step 2: Can then find $I_C(t)$ through $I_C(t) = Q_F(t)/\tau_F$.

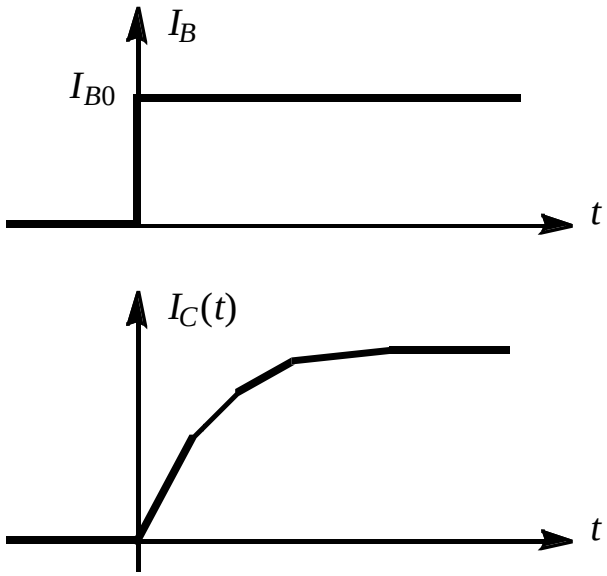
Visualization of $Q_F(t)$



$$\frac{dQ_F}{dt} = I_B(t) - \frac{Q_F}{\tau_F \beta_F}$$

EXAMPLE : Find $I_C(t)$ for a Step $I_B(t)$



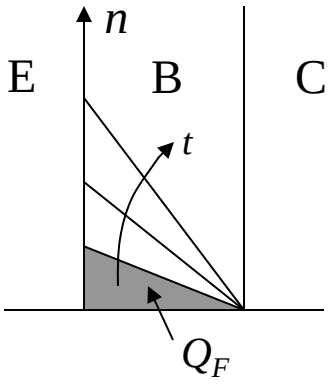


The solution of $\dfrac{dQ_F}{dt} = I_B(t) - \dfrac{Q_F}{\tau_F \beta_F}$ is

$$Q_F = \tau_F \beta_F I_{B0}(1 - e^{-t/\tau_F \beta_F})$$

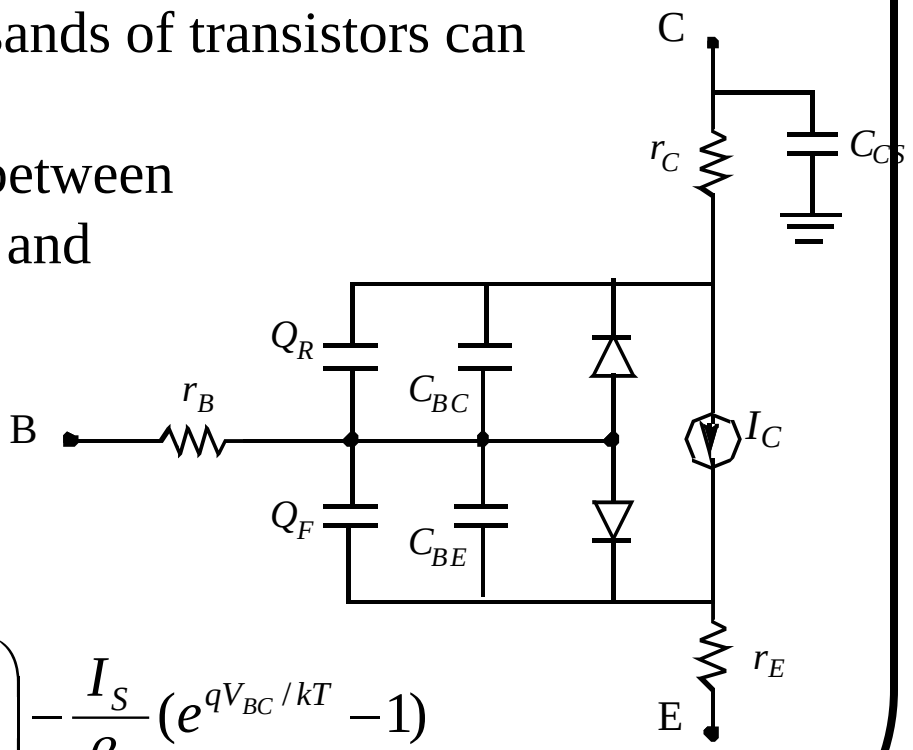
$$I_C(t) = Q_F(t)/\tau_F = \beta_F I_{B0}(1 - e^{-t/\tau_F \beta_F})$$

What is $I_B(\infty)$? $Q_F(0)$? $Q_F(\infty)$?



8.11 Model for Large-Signal Circuit Simulation

- Compact (SPICE) model contains dozens of parameters, mostly determined from measured BJT data.
- Circuits containing tens of thousands of transistors can be simulated.
- Compact model is a "contract" between device/manufacturing engineers and circuit designers.



$$I_C = I_S'(e^{qV_{BE}/kT} - e^{qV_{BC}/kT})\left(1+\frac{V_{CB}}{V_A}\right) - \frac{I_S}{\beta_F}(e^{qV_{BC}/kT} - 1)$$

8.11 Model for Large-Signal Circuit Simulation

A commonly used BJT compact model is the Gummel-Poon model, consisting of

- Ebers-Moll model
- Current-dependent beta
- Early effect
- Transit times
- Kirk effect
- Voltage-dependent capacitances
- Parasitic resistances
- Other effects

8.12 Chapter Summary

- The base-emitter junction is usually forward-biased while the base-collector is reverse-biased. V_{BE} determines the collector current, I_C.

$$I_C = A_E \frac{qn_i^2}{G_B}(e^{qV_{BE}/kT} - 1)$$

$$G_B \equiv \int_0^{W_B} \frac{n_i^2}{n_{iB}^2} \frac{p}{D_B} dx$$

- G_B is the base Gummel number, which represents all the subtleties of BJT design that affect I_C.

8.12 Chapter Summary

- The base (input) current, I_B, is related to I_C by the common-emitter current gain, β_F. This can be related to the common-base current gain, α_F.

$$\beta_F = \frac{I_C}{I_B} \approx \frac{G_E}{G_B} \qquad \alpha_F = \frac{I_C}{I_E} = \frac{\beta_F}{1+\beta_F}$$

- The Gummel plot shows that β_F falls off in the high I_C region due to high-level injection in the base. It also falls off in the low I_C region due to excess base current.
- Base-width modulation by V_{CB} results in a significant slope of the I_C vs. V_{CE} curve in the active region (known as the Early effect).

8.12 Chapter Summary

- Due to the forward bias V_{BE}, a BJT stores a certain amount of excess carrier charge Q_F which is proportional to I_C.

$$Q_F \equiv I_C \tau_F$$

τ_F is the forward transit time. If no excess carriers are stored outside the base, then

$$\tau_F = \tau_{FB} = \frac{W_B^2}{2D_B}$$, the base transit time.

- The charge-control model first calculates $Q_F(t)$ from $I_B(t)$ and then calculates $I_C(t)$.

$$\frac{dQ_F}{dt} = I_B(t) - \frac{Q_F}{\tau_F \beta_F}$$

$$I_C(t) = Q_F(t) / \tau_F$$

8.12 Chapter Summary

The small-signal models employ parameters such as transconductance,

$$g_m \equiv \frac{dI_C}{dV_{BE}} = I_C / \frac{kT}{q}$$

input capacitance,

$$C_\pi = \frac{dQ_F}{dV_{BE}} = \tau_F g_m$$

and input resistance.

$$r_\pi = \frac{dV_{BE}}{dI_B} = \beta_F / g_m$$